



**UNIVERSITY OF NATIONAL AND WORLD ECONOMY**  
*The Spirit Makes The Power*

# **13<sup>TH</sup> INTERNATIONAL CONFERENCE ON APPLICATION OF INFORMATION AND COMMUNICATION TECHNOLOGY AND STATISTICS IN ECONOMY AND EDUCATION ICAICTSEE – 2023**

***December 15 – 16<sup>th</sup>, 2023***

***University of National and World Economy  
Sofia, Bulgaria***

## **CONFERENCE PROCEEDINGS**

**UNDER THE AUSPICES OF:**



***PUBLICATIONS INDEXED BY:***



***CEEOL***



***COOPERATED BY***

***CREATIVITY AND HARMONY SOCIETY***

**COORGANIZING UNIVERSITIES AND INSTITUTIONS**



人とともに 地域とともに  
国立大学法人  
**島根大学**

**Shimane University, Japan**



**沈阳化工大学**  
SHENYANG UNIVERSITY OF CHEMICAL TECHNOLOGY

**Shenyang University of Chemical Technology,  
China**



**Università  
di Genova**

**University of Genoa, Italy**



**Ukrainian Federation of Informatics, Ukraine**



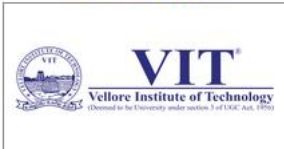
**National Changhua University of  
Education, Taiwan**



**University Danubius of Galati,  
Romania**



**Kharkiv National University of  
Radioelectronics, Ukraine**



**Vellore Institute of Technology,  
India**



**Tomar Politechnic Institute,  
Portugal**



**Volga State University of Technology,  
Russian Federation**



**MB University,  
Serbia**



**Bohdan Khmelnytsky National  
University of Cherkasy, Ukraine**



**Kharkiv State University of Food  
Technology and Trade, Ukraine**



**Korkyt Ata Kyzylorda University,  
Kazakhstan**



**Bashkir State University,  
Russian Federation**



**Kyiv National University of Culture and  
Arts, Ukraine**



**Southern Federal University, Russian  
Federation**



**Gomel State University named after  
Francisk Skorina, Republic of Belarus**



**Artifex University of Bucharest,  
Romania**



**Academia de Studii Economice din  
Moldova**

**PROCEEDINGS OF THE  
13<sup>TH</sup> INTERNATIONAL CONFERENCE ON APPLICATION OF  
INFORMATION AND COMMUNICATION TECHNOLOGY AND  
STATISTICS IN ECONOMY AND EDUCATION  
December 15– 16<sup>th</sup>, 2023, Sofia Bulgaria**

***ICAICTSEE-2023***

**<http://icaictsee.unwe.bg/>**

**EDITORS:**

**PROF. DR. DIMITER G. VELEV**

DEPARTMENT OF INFORMATICS  
UNIVERSITY OF NATIONAL AND WORLD ECONOMY, SOFIA, BULGARIA  
[DGVELEV@UNWE.BG](mailto:DGVELEV@UNWE.BG)

**PROF. DR. PLAMENA ZLATEVA**

DEPARTMENT OF INFORMATICS  
UNIVERSITY OF NATIONAL AND WORLD ECONOMY, SOFIA, BULGARIA

**PROF. DR. SIVAKUMAR R.**

SCHOOL OF ELECTRONICS ENGINEERING  
VELLORE INSTITUTE OF TECHNOLOGY, VELLORE, TAMIL NADU, INDIA

**ISSUED FOR PUBLICATION: OCTOBER 1<sup>ST</sup>, 2024**

**SIZE: JIS-B5**

**QUIRES: 17.75**

**PAGES: 284**

**VOLUME: 50**

**ISSN 2367-7635 (PRINT)**

**ISSN 2367-7643 (ONLINE)**

**PUBLISHING COMPLEX – UNWE, SOFIA, BULGARIA**



## Conference Committees

### *Conference chair*

Dimiter G. <b>Velev</b>	University of National and World Economy	<b>Bulgaria</b>
-------------------------	---	-----------------

### *Virtual Conference Organizer*

Dr. <b>Sivakumar R</b>	School of Electronics Engineering, Vellore Institute of Technology (VIT)	<b>India</b>
------------------------	--	--------------

### **PROGRAM COMMITTEE**

César Correa <b>Arias</b>	Universidad de Guadalajara	<b>Mexico</b>
Saji <b>Baby</b>	GEO Environmental Consultation	<b>Kuwait</b>
Andrei <b>Buiga</b>	Artifex University of Bucharest	<b>Romania</b>
Marian-Cristian <b>Barbu</b>	Artifex University of Bucharest	<b>Romania</b>
Elena <b>Bazhenova</b>	Southern Federal University	<b>Russia</b>
Tetiana <b>Bochulia</b>	Kharkiv State University of Food Technology and Trade	<b>Ukraine</b>
Venelin <b>Boshnakov</b>	University of National and World Economy	<b>Bulgaria</b>
Gaetano <b>Cascini</b>	Politecnico di Milano	<b>Italy</b>
Olena <b>Chaikovska</b>	Kiev National University of Culture and Arts	<b>Ukraine</b>
Zoran <b>Čekerevac</b>	MB University	<b>Serbia</b>
Luisa <b>Chimenz</b>	University of Genoa	<b>Italy</b>
Ghenadie <b>Ciobanu</b>	Artifex University of Bucharest	<b>Romania</b>
Sérgio Manuel Serra da <b>Cruz</b>	Universidade Federal Rural do Rio de Janeiro	<b>Brazil</b>
Tadeusz <b>Czachorski</b>	Institute of Theoretical and Applied Informatics, PAS	<b>Poland</b>
Linda <b>Daniela</b>	University of Latvia	<b>Latvia</b>
Vladimir <b>Dimitrov</b>	Sofia State University	<b>Bulgaria</b>
Marco <b>Ferrante</b>	University of Genoa	<b>Italy</b>
Carlos Arturo T. <b>Gastelu</b>	Universidad Veracruzana	<b>Mexico</b>
Vasil <b>Georgiev</b>	Sofia State University	<b>Bulgaria</b>

Valentin <b>Goev</b>	University of National and World Economy	<b>Bulgaria</b>
Igor <b>Grebennik</b>	Kharkiv National University of Radio Electronics	<b>Ukraine</b>
Chih-Hsuan <b>Huang</b>	Hubei University of Economics	<b>China</b>
Nail <b>Ishmuhametov</b>	Bashkir State University	<b>Russia</b>
Aytekin <b>Isman</b>	Sakarya University	<b>Turkey</b>
Kalinka <b>Kaloyanova</b>	Sofia State University	<b>Bulgaria</b>
Madjid Malikovich <b>Karimov</b>	Tashkent University of Information Technologies	<b>Uzbekistan</b>
Sergii <b>Kavun</b>	Kharkiv University of Technology STEP	<b>Ukraine</b>
Temirbolat <b>Kenshinbay</b>	Korkyt Ata Kyzylorda University	<b>Kazakhstan</b>
Hee Dong <b>Kim</b>	Hankuk University of Foreign Studies	<b>S.Korea</b>
Rosen <b>Kirilov</b>	University of National and World Economy	<b>Bulgaria</b>
Peter <b>Kokol</b>	University of Maribor	<b>Slovenia</b>
Nurbek <b>Kongyrbayev</b>	Korkyt Ata Kyzylorda University	<b>Kazakhstan</b>
Özgen <b>Korkmaz</b>	Amasya University	<b>Turkey</b>
Oleksiy E. <b>Kovalenko</b>	Institute of Mathematical Machines and Systems Problems, NASU	<b>Ukraine</b>
Nurlan <b>Kulmurzayev</b>	Korkyt Ata Kyzylorda University	<b>Kazakhstan</b>
Alexandru-Lucian <b>Manole</b>	Artifex University of Bucharest	<b>Romania</b>
Emanuel-Stefan <b>Marinescu</b>	Danubius University Galati	<b>Romania</b>
Liviu-Mihail <b>Marinescu</b>	Danubius University Galati	<b>Romania</b>
Célio Gonalo Cardoso <b>Marques</b>	Polytechnic Institute of Tomar	<b>Portugal</b>
Pedja M. <b>Milosavljevic</b>	University of Nis	<b>Serbia</b>
A.K. <b>Mishra</b>	Motilal Nehru National Institute of Technology	<b>India</b>
Igor <b>Mladenovic</b>	University of Nis	<b>Serbia</b>
Alexander N. <b>Moiseev</b>	Tomsk State University	<b>Russia</b>
Kaninda <b>Musumbu</b>	Universite Bordeaux	<b>France</b>

<b>Tetsuo Noda</b>	Shimane University	<b>Japan</b>
<b>Florian Nuta</b>	Danubius University Galati	<b>Romania</b>
<b>Serghei Ohrimenco</b>	Laboratory of Information Security, AESM	<b>Moldova</b>
<b>K.R. Pardasani</b>	Maulana Azad National Institute of Technology	<b>India</b>
<b>Igor Petuhov</b>	Volga State University of Technology	<b>Russia</b>
<b>Monica Pocora</b>	Danubius University Galati	<b>Romania</b>
<b>David Nadler Prata</b>	Universidade Federal do Tocantins	<b>Brazil</b>
<b>Olga Pugacheva</b>	Gomel State University named after Francisk Skorina	<b>Belarus</b>
<b>Andy Pusca</b>	Danubius University Galati	<b>Romania</b>
<b>N.Sharmila</b>	Vellore Institute of Technology	<b>India</b>
<b>Yizhe Shi</b>	Shenyang University of Chemical Technology	<b>China</b>
<b>R. Sivakumar</b>	Vellore Institute of Technology	<b>India</b>
<b>Tullio Tanzi</b>	Institute Mines Telecom - Telecom ParisTech	<b>France</b>
<b>S. Vidhya</b>	Vellore Institute of Technology	<b>India</b>
<b>Hsin-Hung Wu</b>	National Changhua University of Education	<b>Taiwan</b>
<b>Nataliia Zachosova</b>	Bohdan Khmelnytsky National University of Cherkasy	<b>Ukraine</b>
<b>Teodora Zapryanova</b>	University of Economics - Varna	<b>Bulgaria</b>
<b>Plamena Zlateva</b>	University of National and World Economy	<b>Bulgaria</b>
<b>Milan Zorman</b>	University of Maribor	<b>Slovenia</b>
<b>Svetoslav Spasov</b>	University of National and World Economy	<b>Bulgaria</b>
<b>Zong Xuejun</b>	Shenyang University of Chemical Technology	<b>China</b>

## **UNWE ORGANIZING COMMITTEE**

<b>Katya Kirilova</b>	University of National and World Economy
<b>Plamen Milev</b>	University of National and World Economy
<b>Venko Andonov</b>	University of National and World Economy
<b>Gabriela Naskova</b>	University of National and World Economy

## **VIT ORGANIZING COMMITTEE**

<b>A Mythili</b>	Vellore Institute of Technology
<b>S Vidhya</b>	Vellore Institute of Technology
<b>N Sharmila</b>	Vellore Institute of Technology

*Dear Colleagues,*

*On behalf of the Conference Committees I have the honor and pleasure to thank you all for your true participation in the work of the **13<sup>th</sup> International Conference on Application of Information and Communication Technology and Statistics in Economy and Education (ICAICTSEE-2023)**, <http://icaictsee.unwe.bg/>, which took place on December 15-16<sup>th</sup>, 2023 at the University of National and World Economy (UNWE), Sofia, Bulgaria.*

***ICAICTSEE** shares experience with the **Recent Trends in Biomedical Technology Conference**, co-organized by the School of Electronics Engineering, Vellore Institute of Technology, Vellore, India.*

*I would like specially to express my gratitude to all of the program committee members for their genuine support without which it would never had happened.*

*Due to the current state of scientific development in all spheres of human activity, the constant knowledge and skill actualization of the academia and researchers in the field of Information and Communication Technology (ICT) is an obligatory necessity, especially when the world is in a global economic and financial crisis. The definition of long-term scientific research tasks in this area has a strategic importance, which is even truer for young researchers, lecturers and Ph.D. students.*

*The conference aims were conducting analysis of the current problems and presenting results of the ICT application in different areas of economy, education and related areas of scientific knowledge; outlining the existing possibilities for the application of modern ICT tools, methods, methodologies and information systems in economy and education; discussing advanced and emerging research trends with a long-term importance in the field of ICT application in economy and education.*

*The conference has established an academic environment that fosters the dialogue and exchange of ideas between different levels of academic and research communities.*

*The conference outcome is 38 published research papers, the explosion of fresh ideas and the establishment of new professional relations.*

*Dimitar G. Velez*  
*ICAICTSEE-2023 Chair*





## CONTENTS

TITLE	Pages
<b>NEW FORMS AND FEATURES OF CYBER CONFRONTATIONS</b> <b>Serghei Ohrimenko<sup>1</sup>, Valeriu Cernei<sup>2</sup></b> <sup>1</sup> Information Security Lab, Academy of Economic Studies of Moldova, Chisinau, Moldova <sup>2</sup> Partner, IT Audit & Advisory, BSD Management SRL, Chisinau, Moldova	17
<b>IMPACT OF AI ON THE KOREAN MEDIA</b> <b>James Kang</b> St. Cyril and St. Methodius University of Veliko Tarnovo, Bulgaria Paisii Hilendarski University of Plovdiv, Bulgaria	23
<b>ABOUT ONLINE DEDICATED RESOURCES FOR INFORMATION ON CYBER THREATS AND CYBER INTELLIGENCE</b> <b>Ivan Hristozov, Veselina Gagamova, Kristina Krumova, Chavdar Kostadinov</b> G. S. Rakovski National Defense College, Sofia, Bulgaria	32
<b>ON VERIFICATION OF CYBER INTELLIGENCE TACTICS, TECHNIQUES AND PROCEDURES</b> <b>Ivan Hristozov, Veselina Gagamova, Chavdar Kostadinov, Kristina Krumova</b> G. S. Rakovski National Defense College, Sofia, Bulgaria	37
<b>INCIDENCE OF FAKE NEWS SHARED ON SOCIAL NETWORKS AND RISK ASSUMED BY UNIVERSITY STUDENTS</b> <b>Carlos A. Torres-Gastelú<sup>1</sup>, Joel Angulo-Armenta<sup>2</sup></b> <sup>1</sup> Universidad Veracruzana, Veracruz, Mexico <sup>2</sup> Instituto Tecnológico de Sonora, Sonora, Mexico	43
<b>ETHICAL BEHAVIOUR ASPECTS IN THE CONTEMPORARY CYBER ENVIRONMENT (ETHICAL BEHAVIOUR AND CYBERSECURITY)</b> <b>Konstantin Poudin</b> University of National and World Economy, Sofia, Bulgaria	52
<b>APPLICATION OF CYBER INTELLIGENCE FOR SECURITY OPERATION CENTER</b> <b>Nedko Tagarev</b> University of National and World Economy, Sofia, Bulgaria	57
<b>GAP ANALYSES IN CYBERSECURITY</b> <b>Nedko Tagarev</b> University of National and World Economy, Sofia, Bulgaria	66

<b>APPLICATION OF DIGITAL TOOLS IN THE EDUCATIONAL PROCESS</b> <b>Katia Kirilova</b> University of National and World Economy, Sofia, Bulgaria	73
<b>THE ROLE OF DATA MANAGEMENT AND OPEN SCIENCE IN SUPPORTING DIGITAL TECHNOLOGIES AT UNIVERSITIES</b> <b>Plamen Milev</b> University of National and World Economy, Sofia, Bulgaria	77
<b>SPECTRAL ANALYSIS THE WAVE SIGNAL OF ELECTRONIC SPHYGMOMANOMETER</b> <b>Lubomir Lahtchev</b> Institute of Robotics, Bulgarian Academy of Sciences, Sofia, Bulgaria	84
<b>CROWDFUNDING AS AN ALTERNATIVE SOURCE OF INVESTMENTS IN THE CONDITIONS OF ECONOMIC DIGITALIZATION: RELEVANCE OF USE IN THE REPUBLIC OF MOLDOVA</b> <b>Irina Dorogaia<sup>1</sup>, Alexandr Scutari<sup>2</sup></b> <sup>1</sup> Academy of Economic Studies of Moldova, Chisinau, Moldova <sup>2</sup> BALCANICA SRL, Chisinau, Republic of Moldova	94
<b>A COMBINATION OF INVESTMENTS IN RENEWABLE ENERGY AND HEAT PUMPS IN THE PRIVATE SECTOR</b> <b>Lesan Timur</b> Academy of Economic Studies of Moldova, Chisinau, Moldova	104
<b>CAPITAL STRUCTURE OF THE ENTERPRISE /METHODOLOGY AND METHODS FOR ANALYSIS IN THE CONTEXT OF DIGITALISATION/</b> <b>Rositsa Ivanova</b> University of National and World Economy, Sofia, Bulgaria	112
<b>SPECIFIC FEATURES OF THE ACCOUNTING OF INSURERS' EXPENSES</b> <b>Maia Natchkova</b> University of National and World Economy, Sofia, Bulgaria	120
<b>ECG SIGNAL PROCESSING AND HEART RATE VARIABILITY ANALYSIS USING LABVIEW</b> <b>Lohitaa J., Sivakumar R.</b> Vellore Institute of Technology, Vellore, India	128
<b>SMART GAIT ANALYSIS FRAMEWORK FOR STRENGTHENING RURAL HEALTHCARE</b> <b>Arshya Loomba, Sivakumar R.</b> Vellore Institute of Technology, Vellore, India	137

<b>SMART BIO-SHIRT: ADVANCED WEARABLE HEALTH MONITORING</b> <b>Shubhankar S., Sivakumar R.</b> Vellore Institute of Technology, Vellore, India	143
<b>ANALYSIS OF THE ULTRASOUND SIGNAL USING HILBERT TRANSFORM</b> <b>Sakthivel J, Sivakumar R</b> Vellore Institute of Technology, Vellore, India	147
<b>WEARABLE TREMOR-ABSORBING DEVICE FOR PARKINSON'S DISEASE PATIENTS</b> <b>Praviene P., Sivakumar R.</b> Vellore Institute of Technology, Vellore, India	153
<b>PIEZOELECTRIC NANOGENERATOR BASED SMART WATCH</b> <b>Guhan C., Sivakumar R.</b> Vellore Institute of Technology, Vellore, India	157
<b>STROKE GAIT ANALYSIS USING IMU SENSORS</b> <b>Gaadha Jeevan, Sivakumar R.</b> Vellore Institute of Technology, Vellore, India	163
<b>IDIOPATHIC TOE WALKING USING MACHINE LEARNING AND GAIT ANALYSIS</b> <b>Maanasa Mahadevan, Pragadeesh Murugan, Sivakumar R.</b> Vellore Institute of Technology, Vellore, India	168
<b>CHARACTERIZATION OF RESPIRATORY DISEASES USING LABVIEW</b> <b>Yuvapriya E., Sivakumar R.</b> Vellore Institute of Technology, Vellore, India	176
<b>WEARABLE TECHNOLOGY IN HEALTHCARE MONITORING</b> <b>Challa Revanth Kumar, Sivakumar R.</b> Vellore Institute of Technology, Vellore, India	183
<b>LABVIEW-BASED EEG SIGNAL ANALYSIS FOR THE IDENTIFICATION AND CHARACTERIZATION OF ABERRANT BRAIN ACTIVITY PATTERNS IN NEUROLOGICAL DISORDERS</b> <b>Saran Aravinda C., Sivakumar R.</b> Vellore Institute of Technology, Vellore, India	188

<b>LABVIEW'S CUTTING-EDGE TECHNOLOGY IN CARDIAC DISEASE MONITORING</b> <b>Chandhini Jeyashanker, Sivakumar R.</b> Vellore Institute of Technology, Vellore, India	193
<b>SMART BANDAGE FOR THE CONTINUOUS MONITORING OF WOUND USING LABVIEW</b> <b>Sakthivel K., Sivakumar R.</b> Vellore Institute of Technology, Vellore, India	199
<b>POSSIBILITIES OF APPLYING THE BIBLIOMETRIC ANALYSIS METHODS OF SCIENTIFIC PUBLICATIONS ON THE RELATIONSHIP BETWEEN CLIMATE CHANGE AND THE NUMBER AND INTENSITY OF FOREST FIRES</b> <b>Tsvetan Tsvetkov, Yassen Tsvetkov</b> University of National and World Economy, Sofia, Bulgaria	206
<b>MODERN TECHNIQUES AND TOOLS FOR GENERATING SQL QUERIES</b> <b>Veska Mihova, Maria Marzovanova</b> University of National and World Economy, Sofia, Bulgaria	215
<b>USING CODE GENERATION IN LLMS FOR AUTOMATIC EXECUTION OF DATA SCIENCE TASKS</b> <b>Venko Andonov</b> University of National and World Economy, Sofia, Bulgaria	223
<b>RAG STRATEGIES FOR A HALLUCINATION-FREE UNIVERSITY ADMISSION AI ASSISTANT</b> <b>Venko Andonov</b> University of National and World Economy, Sofia, Bulgaria	230
<b>FEATURES AND CHALLENGES OF NATURAL LANGUAGE PROCESSING IN THE MODERN DIGITAL WORLD</b> <b>Yavor Tabov</b> University of National and World Economy, Sofia, Bulgaria	237
<b>THE CASH OF THE BIG BULGARIAN COMPANIES: WHAT DO THEIR CASH FLOW STATEMENTS SHOW AND WHAT DO NOT?</b> <b>Fanya Filipova, Teodora Zapryanova, Atanas Atanasov, Rumyna Marinova</b> University of Economics - Varna, Varna, Bulgaria	245
<b>CHALLENGES FOR AUDITORS IN RELATION TO DISCLOSURE OF SUSTAINABILITY INFORMATION IN FINANCIAL STATEMENTS</b> <b>Iavor Bachev</b> University of National and World Economy, Sofia, Bulgaria	257

<b>ANALYSIS ON THE APPLICATIONS OF RETRIEVAL AUGMENTED GENERATION (RAG) ARCHITECTURE</b> <b>Bozhidar Bahov</b> University of National and World Economy, Sofia, Bulgaria	<b>261</b>
<b>UNLOCKING AND EXPLORING THE POWER OF UNSTRUCTURED BIG DATA USING PENTAO BUSINESS ANALYTICS</b> <b>Mariana Kovacheva</b> University of National and World Economy, Sofia, Bulgaria	<b>267</b>
<b>THEORETICAL AND PRACTICAL FINANCIAL ANALYSIS OF A REAL ESTATE INTERMEDIARY COMPANY AND THE ROLE OF SOFTWARE</b> <b>Zlatan Enchev</b> University of National and World Economy, Sofia, Bulgaria	<b>273</b>





# New Forms and Features of Cyber Confrontations

Serghei Ohrimenco<sup>1</sup>, Valeriu Cernei<sup>2</sup>

<sup>1</sup>Information Security Lab, Academy of Economic Studies of Moldova, Chisinau, Moldova

<sup>2</sup>Partner, IT Audit & Advisory, BSD Management SRL, Chisinau, Moldova  
*osa@ase.md, valeriu.cernei@bsd.md*

**Abstract.** The evolution of the cyberspace domain and its accompanying management tools has ushered in an era characterized by the emergence of new forms of confrontation: cyber blockades and cyber sanctions. This development has paved the ways for enhanced capabilities in economic management, the expansion of diplomatic and trade relationships, and ultimately, the safeguarding of national interests at an unprecedented level. This paper presents the results of an analysis of these new categories, elucidating their interconnectedness, and highlighting the evolution and continued development of these two domains, particularly within cyberspace. The authors draw attention to the new dimensions of confrontation within the cyberspace realm, encompassing explicit and hidden threats to individuals, society, and the state.

**Keywords:** Cyberspace domain, Cybersecurity, Cyberconflict, Cybersanction, Cyber threats

## 1. Introduction

The interactions between users, information systems, and their components can traverse several distinct stages: cooperation, integration, competition, and confrontation. It is noteworthy that these stages span across various domains, including political, military, economic, technological, social, and ideological. Interaction and confrontation manifest themselves at multiple levels, encompassing the global, strategic, tactical, and individual spheres. The significance of this topic is greatly magnified in the context of each nation's quest to safeguard its national interests concerning cybersecurity [1]. A crucial observation arises from [2, 3], emphasizing the unique characteristics of the cyberspace domain when compared to other domains.

Firstly, cyberspace is a product of human creation. Secondly, military and civilian (diplomatic, trade, etc.) capabilities in other domains are managed through and by means of the cyber domain. Thirdly, distinguishing between military and civilian aspects of cyber operations is often challenging and occasionally impossible due to their intricate intertwinement. Fourthly, attributing operations conducted within cyberspace is an exceedingly complex endeavor (Fig.1).

Not too long ago, humanity recognized only two domains, namely land and sea. However, with the relentless advancement of science and technology, new domains such as aerospace, information, and most recently, cyberspace have emerged. It is imperative to fundamentally reevaluate the system for managing cyber risks in these new conditions [4-8].

## 2. Literature Overview

It is important to acknowledge that the first and the most comprehensive work on the subject of cyber blockades can be found in the works of [9-10], and [11]. A. Russel, in

particular, deserves commendation for spearheading the establishment and exploration of issues pertaining to the introduction and assessment of cyber blockades.

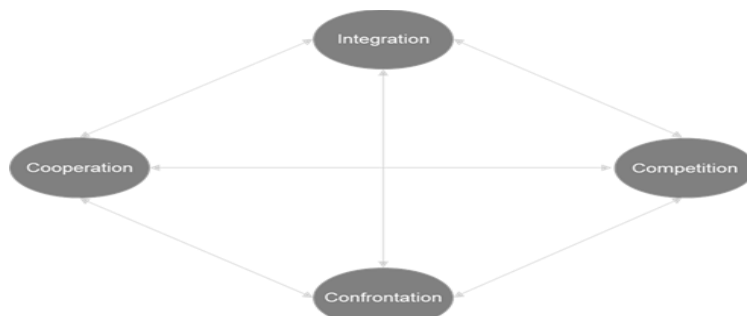


Fig. 1. Users-Systems interaction stages

The literature utilized in the preparation of this material can be categorized into several groups. Firstly, there are sources that analyze and elucidate the issues surrounding the genesis and subsequent development of domains, including cyberspace [5]. In [12], conflicts within cyberspace are dissected, with a particular emphasis on topics ranging from "Cyber Threats to Cyber Risks" to empirical analyses of cyber conflicts. Another noteworthy source is Nicholas Michael Sambaluk's book [1], which delves into facets of 21st-century technology. It elucidates how the United States and the Soviet Union sought to explore and exploit the space domain to advance national security goals, including intelligence data collection, communications, and guidance.

Secondly, there is an examination of the mechanisms underpinning political and economic confrontation – sanctions (unilateral, multilateral, and "smart" sanctions) [13]. In the book [14], sanctions are portrayed as an increasingly popular tool of foreign policy, not only at the multilateral level (within the UN) but also regionally (notably within the EU) and unilaterally. The book scrutinizes the contemporary practices of various actors and delves into the legality (or lack thereof) of their actions. In the subsequent book [15], the utilization of targeted sanctions as a central instrument to address challenges to international peace and security is discussed, which has become a defining feature of UN Security Council practice. [16], delves into the fundamentals of the concept of economic statecraft, as well as alternative concepts such as Foreign Economic Policy, International Economic Policy, Economic Diplomacy, Economic Leverage, Economic Sanctions, Economic Warfare, Economic Coercion, and more.

Thirdly, there are important theoretical developments of a politico-economic nature [17-20]. Among them, specific chapters in the book [21] stand out. These include Chapter 22, "Economics of Conflict: An Overview" (authored by Michelle R. Garfinkel and Stergios Skaperdas), and Chapter 27, "The Political Economy of Economic Sanctions" (written by William H. Kaempfer and Anton D. Lowenberg). The next important source [22] acknowledges that there can be no simple equation that reveals whether or not sanctions will be successful.

### 3. Research Methodology

The methodological approach adopted in this publication is grounded in a systemic perspective, wherein the research object is viewed as a system comprised of interconnected elements and a complex interplay of entities with their respective attributes and relationships.

The research concept was formulated, key objectives were defined and scrutinized, and statistical data was collected from monographic sources and companies' reports. This gathered information was subsequently subjected to an analysis employing game theory, econometric analysis, and modeling techniques.

Documents from the World Economic Forum [23-25], notably "The Global Risks Report 2023," underscore that the upcoming decade (2023-2030) will be characterized by a series of crises. A new term, "polycrisis," has been coined, denoting a scenario, where various risks intersect, and their interdependencies become acutely palpable. "Old" risks have evolved, including inflation, cost-of-living crises, trade disputes, social unrest, geopolitical tensions, and more. In tandem with these traditional risks, new challenges have arisen, encompassing high levels of government debt, diminished investments, deglobalization, and other factors. All of this underscores the necessity to reevaluate the mechanisms for assessing the impact of sanctions and conflicts in the cyberspace domain.

#### 4. Analysis

As a result of the comprehensive analysis conducted, several key terms related to cyber blockades and its components interaction have been delineated and are presented further.

**Cyberspace:** This term encapsulates a physical network that can be used and manipulated to apply punitive measures against targets by obstructing their access to the data flow critical for both, security and prosperity.

**Subjects:** These actors utilize cyber blockades as instruments of international relations and sanctions due to their effectiveness and cost-efficiency in controlling the targets' access to modern networks. They can be executed in a manner that enhances the perpetrator's anonymity or deniability, thereby reducing the risk of retaliation. Additionally, various alternative courses of action can be employed to achieve similar results, particularly for non-state actors.

**Cyber Blockade:** A state arising from an attack on cyber infrastructure or systems that obstructs user access to cyberspace, and so impeding data transmission beyond geographical borders. Cyber blockades are recognized as legitimate instruments of international statecraft, and in accordance with other forms of blockades, may be considered acts of war (though the target state ultimately decides whether to classify them as acts of war and potentially escalate the situation). Cyber blockades are directed at entire states and usually seek to induce disruptions within critical infrastructure elements. Their primary objective is to prevent the transmission of data across geographical borders through the manipulation, control, or domination of cyberspace and associated technologies, causing political, economic, social, or psychological harm upon the adversary.

**Effectiveness of Cyber Blockades:** The success of a cyber blockade consists on its ability to prevent information transmission. The duration of the blockade is a secondary consideration, relevant only in terms of achieving the desired outcome. For instance, a blockade lasting some seconds may hold relatively little significance, while a blockade occurring in a critical moment (such as on an election day) or one persisting for several weeks or months can be highly effective, contingent on the objectives. Analogous to maritime blockades, maintaining a cyber blockade for a pre-defined period is not obligatory. What assumes paramount importance is the ultimate effectiveness of the blockade in realizing the stated goals.

To assess the trajectories of cyber warfare, the work presented by a consortium of authors from the RAND Corporation [26] offers valuable insights. Future war trends are delineated across several dimensions: geopolitical trends, military trends, space and nuclear trends, cyber trends, and deterrence trends. The analysis of information points to a complete

alignment of targets across all trends that the United States is anticipated to counter. These encompass Russia, China, North Korea, and Iran, in addition to non-state actors. This implies a complex array of questions, including the responses the United States has contemplated in reaction to cyber compromises of state information systems, whether past responses have substantially influenced the adversary's behavior, and how analogous incidents should be addressed in the future. Answers to these inquiries should be sought in forthcoming research endeavors.

## 5. Discussion

The new segment of the discussion focuses on the organization and execution of sanctions within the cyberspace domain, an area yet to be fully explored. Specifically, the new political and economic instrument, as well as a new means of exerting influence, has emerged: cyber sanctions. With the development of IT components that have shaped the digital landscape, cyberspace has evolved into a competitive arena among leading information-oriented nations and foremost producers of computer hardware and software. A constellation of challenges concerning the digital environment, spanning individuals, society, and states, has arisen [27-30].

Cyber sanctions are defined as economic and financial measures aimed at affecting behavioral change in target entities through malicious activities in cyberspace and/or intrusions [31-33]. An analysis of the theory and practice of interactions among various entities within the cyber sphere underscores that the concept of cyber sanctions represents a relatively new area of research within the global cyberspace domain.

Cyber sanctions, as a relatively new instrument of Governments, are likely to see increased use in the future. Governments and international organizations may refine their strategies for imposing cyber sanctions, potentially developing more standardized protocols and frameworks for their application. This could include defining clear criteria for when cyber sanctions are warranted and specifying the range of potential responses.

Moreover, as cyberspace becomes increasingly interconnected with critical infrastructure, the consequences of cyber sanctions could become even more profound. Governments already seek to impose sanctions that directly impact a target nation's critical infrastructure, such as its energy grid or financial systems. This raises complex ethical and humanitarian questions about the collateral effects of such sanctions on civilian populations.

Another important topic is the one related to future evolution of cyber blockades. As technology continues to advance, the nature and effectiveness of cyber blockades are likely to evolve. The integration of artificial intelligence, quantum computing, and autonomous systems into cyber operations may open up new frontiers in cyber warfare. These advancements could potentially lead to more sophisticated and stealthy cyber blockades, making them even harder to detect and mitigate.

The source attribution problem in cyberspace is expected to persist. Accurately identifying the source of a cyber blockade remains a complex challenge, and as state and non-state actors become more adept at concealing their origins, the difficulty of attributing cyberattacks will continue to pose diplomatic and strategic challenges.

Another direction of discussions is the global cooperation and norms in the cyberspace. The escalating use of cyber blockades, cyber sanctions, and cyberattacks in international conflicts lowers the pressing need for global cooperation and the development of norms in the cyberspace. The international community will likely intensify efforts to establish rules of behavior in cyberspace, similar to existing norms in conventional warfare. Building consensus on these matters will be challenging but vital for maintaining stability in the cyberspace.

And finally, the role of non-state actors would be crucial. Non-state actors, including hacktivist groups and cybercriminal organizations, are becoming increasingly influential in cyberspace. These groups can carry out disruptive actions independently or on behalf of state actors, blurring the lines between state-sponsored and non-state cyber activities. In the future, their roles in cyber blockades and cyber sanctions may grow, making it essential for governments and international organizations to develop strategies for dealing with these entities effectively. This includes enhancing cybersecurity measures to defend against attacks from both state and non-state actors.

## 6. Conclusion

The comparison of the impact of cyberblockades and sanctions suggests a fundamental transformation, transforming the "tool of war prevention" into a weapon of geo-economic warfare, accompanied by its attendant consequences. The achievement of declared political objectives now predominantly relies on the utilization of economic instruments, at times severe, and in some cases, even harsh measures.

In other words, the effectiveness of sanctions and blockades can be traced through the assessment of a complex set of indicators characterizing the contradictions of capital, which have intensified within leading global powers and between them.

Sanctions increases the contradiction between transnational capital and national labor. The reduction in real incomes among the most vulnerable societal groups threatens stable consumption and augments the demand for government expenditures. Simultaneously, channels of financing labor reproduction through borrowing in global financial markets constrict due to a "denial of any credit, credit guarantees, or other financial assistance."

The intricate interplay between cyber blockades, sanctions, and global capital underscores the evolving landscape of international relations and conflicts, with cyberspace occupying a pivotal role in reshaping the dynamics of modern warfare and diplomacy. This paradigm shift requires ongoing research and analysis to comprehend the full spectrum of implications and devise effective strategies for an increasingly interconnected world.

In conclusion, the evolution of cyber blockades and cyber sanctions is an ongoing and dynamic process. As technology continues to advance and cyberspace becomes increasingly integral to all aspects of modern life, the impact of these measures will likely intensify. Policymakers, cybersecurity experts, and international stakeholders must remain vigilant and adaptive in their efforts to navigate this complex and ever-changing landscape. This involves not only responding to current challenges but also proactively shaping the future of cybersecurity and international relations in the digital age.

## References

1. Sambaluk N.M. (2019). Conflict In the 21st -Century: The Impact of Cyber Warfare, Social Media, And Technology. ABC-CLIO, LLC. ISBN: 978-1-4408-6000-3
2. Brantly A. F. (2016). The Decision to Attack Military and Intelligence Cyber Decision- Making. University of Georgia Press.
3. Richard A. Clarke, Robert K. Knake (2019). The Fifth Domain: Defending Our Country, Our Companies, and Ourselves in the Age of Cyber Threats. Penguin Press.
4. Pogrebna G., Skilton M. (2019). Navigating New Cyber Risks: How Businesses Can Plan, Build and Manage Safe Spaces in the Digital Age. Palgrave Macmillan.
5. Friis K., Ringsmose J. (2016). Conflict in Cyber Space: Theoretical, Strategic and Legal Perspectives. Routledge.
6. Saffady W. (2020). Managing Information Risks: Threats, Vulnerabilities, and Responses. Rowman & Littlefield.

7. Andersen T. J. (2016). *The Routledge Companion to Strategic Risk Management*. Routledge.
8. Oh K-B., Ho B., Slade B. (2022). *Cybersecurity Risk Management; An Enterprise Risk Management Approach*. Nova Science Publ. ISBN 978-1-68507-505-7, <https://doi.org/10.52305/TNSD3712>
9. Russell A. L. (2014). *Cyber blockades*. Georgetown University Press.
10. Russell A. L. (2015). Strategic anti-access/area denial in cyberspace. 2015 7th International Conference on Cyber Conflict: Architectures in Cyberspace. doi:10.1109/cycon.2015.7158475
11. Russell A. L. (n.d.). The Physical Layer. Strategic A2/AD in Cyberspace, 26–39. doi:10.1017/9781316817001.003
12. Whyte Ch., Thrall A.T., Brian M. Mazanec B. M. (2020). *Information Warfare in the Age of Cyber Conflict*. Routledge.
13. Bogdanova I. (2022). *Unilateral Sanctions in International Law and the Enforcement of Human Rights: The Impact of the Principle of Common Concern of Humankind*. World Trade Institute Advanced Studies..
14. Happold M., Eden P. (2016) *Economic Sanctions and International Law: Studies in Int. Law*. Hart Publ. 7
15. Biersteker T. J., Eckert S. E., Tourinho M. (2016). *Targeted Sanctions: The Impacts and Effectiveness of United Nations Action*. Cambridge University Press.
16. Baldwin D. A., Kapstein E. B. (2020). *Economic statecraft*. Princeton University Press.
17. Kirkham K. (2022). *The Political Economy of Sanctions: Resilience and Transformation in Russia and Iran*. Palgrave Macmillan.
18. Karatzogianni A. (2006). *The Politics of Cyberconflict*. Routledge.
19. Jasper S. (2012). *Conflict and Cooperation in the Global Commons: A Comprehensive Approach for International Security*. Georgetown University Press.
20. Laurent S. Y. (2021) *Conflicts, Crimes and Regulations in Cyberspace*. ISTE Ltd, John Wiley & Sons
21. Sandler T., Hartley K. (2007). *Handbook of Defense Economics: Defense in a Globalized World*. vol.2, North Holland.
22. Jaeger M. (2018) *Coercive Sanctions and International Conflicts: A Sociological Theory*. Routledge
23. The Global Risks Report 2023. 18th Edition. [https://www3.weforum.org/docs/WEF\\_Global\\_Risks\\_Report\\_2023.pdf](https://www3.weforum.org/docs/WEF_Global_Risks_Report_2023.pdf)
24. Digital Safety Risk Assessment in Action: A Framework and Bank of Case Studies. May 2023. [https://www3.weforum.org/docs/WEF\\_Global\\_Coalition\\_Digital\\_Safety\\_Risk\\_Assessments\\_2023.pdf](https://www3.weforum.org/docs/WEF_Global_Coalition_Digital_Safety_Risk_Assessments_2023.pdf)
25. Chief Risk Officers Outlook. July 2023. [https://www3.weforum.org/docs/WEF\\_Chief\\_RiskOfficers\\_Outlook\\_2023.pdf](https://www3.weforum.org/docs/WEF_Chief_RiskOfficers_Outlook_2023.pdf)
26. Raphael S. Cohen R. S., Chandler N., Efron S., Frederick B., Han E., Klein K., Morgan F.E., Rhoades A. L., Shatz H. J. (2020). *The Future of Warfare in 2030*. RAND Corp. ISBN: 978-1-9774-0295-0
27. Ohrimenko S., Cernei V. (2021). Shadow Digital Technologies Threats to National Security. Int. Sc. Conf. on Economic and Social Development "Economics, Management, Finance and Banking". Svishtov, 28-30 Sept. 2022, pp. 344-350. <https://www.zbw.eu/econis-archiv/handle/11159/12318>
28. Ohrimenko S., Borta G., Cernei V. (2021) Estimation of the Key Segments of the Cyber Crime Economics. 2021 IEEE International Conference on Problems of Info communications. Science and Technology PIC S&T '2021. Oct. 5-7, 2021. Kharkiv, Ukraine. DOI: 10.1109/PICST54195.2021
29. Lehto M., Neittaanmäki P. (2018). *Cyber Security: Power and Technology*. Springer Int. Publishing
30. Maurer T. (2018). *Cyber Mercenaries: The State, Hackers, and Power*. Cambridge University Press.
31. Miadzvetskaya Y. (2020). Challenges of the Cyber Sanctions Regime under the Common Foreign and Security Policy (CFSP). In: *Security and Law Legal and Ethical Aspects of Public Security, Cyber Security and Critical Infrastructure Security*. Intersentia, pp. 277-298. DOI: 10.1017/9781780688909.012
32. Bossler A. M. (2019) Perceived Formal and Informal Sanctions on the Willingness to Commit Cyber Attacks Against Domestic and Foreign Targets, *Journal of Crime and Justice*, 42:5, 599-615, DOI: 10.1080/0735648X.2019.1692423
33. Walentek D., Broere J., Cinelli M., Dekker M., Haslbeck J. (2021): Success of Economic Sanctions Threats: Coercion, Information and Commitment, *International Interactions*, DOI: 10.1080/03050629.2021.1860034



# Impact of AI on the Korean Media

**James Kang**

St. Cyril and St. Methodius University of Veliko Tarnovo, Bulgaria

Paisii Hilendarski University of Plovdiv, Bulgaria

*j.kang@ts.uni-vt.bg*

**Abstract.** Korean media emerged in the late 19th century following the introduction of Western newspaper publishing technologies. Despite enduring numerous political upheavals, including Japanese colonial rule, the Korean War, military dictatorships, and democratization movements, the media in Korea continued to evolve. However, the rapid advancements in artificial intelligence (AI) in the 21st century have fundamentally altered the traditional media landscape. This paper aims to analyze the impact of AI on the development of Korean media and explore ways in which both entities can coexist harmoniously.

**Keywords:** Korean Media History, Artificial Intelligence (AI), Democratization, Expert Systems, Natural Language Processing (NLP), AI in Journalism

## 1. Korean media: from the enlightenment to democracy

The history of Korean media is closely linked with modern Korean history. Its origins date back to the Enlightenment Period in Korea, with the publication of Hansung Sunbo in 1883, Korea's first newspaper. Initially, the media served to deliver government messages, but this role evolved under Japanese colonial rule (1910-1945), where censorship became severe.

During this period, despite heavy repression, newspapers such as Donga Ilbo and Chosun Ilbo supported the independence movement and fostered national consciousness. For instance, during the March 1st Movement of 1919, Korean media secretly distributed the Declaration of Independence and raised international awareness of Japan's brutal repression. Similarly, during the 1936 Berlin Olympics, Donga Ilbo defied censorship by erasing the Japanese flag from a photograph of Korean marathoner Sohn Kee-chung, which led to government backlash and arrests.

After liberation in 1945, Korean media briefly experienced freedom, but this was soon suppressed under military regimes. The 1959 Kyunghyang Shinmun Closure Incident and the Park Chung-hee regime's strict media control in the 1970s, including the 1974 Donga Ilbo White Space Incident, demonstrated the continued repression of the press. Despite these challenges, the media remained a tool for resistance.

The 1987 June Democratic Uprising marked a turning point, with the media playing a crucial role in spreading the pro-democracy movement, which led to the adoption of a direct presidential election system. In more recent history, the 2016-2017 Candlelight Protests against President Park Geun-hye showed the media's ongoing importance, as outlets like JTBC exposed corruption, sparking her impeachment.

From colonial times through military dictatorship and into democratization, Korean media has been both a target of repression and a key force for change [1].

Korean media has played a crucial social role throughout modern history, oscillating between repression and resistance. Despite strict censorship during the Japanese colonial period and military dictatorships, the media became a key instrument for the independence and

democratization movements. Notable examples include the March 1st Movement of 1919, the June Democratic Uprising of 1987, and the 2016-2017 Candlelight Protests, where the media helped raise public awareness and drive social change. This historical experience demonstrates that Korean media goes beyond merely delivering information; it plays a vital role in contributing to the realization of social justice.

### **Organizations and agencies addressing fake news and misinformation in the Korean media**

In Korea, there are several organizations preventing and controlling the spread of fake news and misinformation, ensuring accurate and reliable information for the public [2].

Organization or Agency	Role	Key Functions
Korea Press Foundation (KPF)	Enhancing media quality through research and education	Operates FactCheckNet, collaborates with media outlets to verify news accuracy and combat the spread of fake news
SNU FactCheck	Verifying news stories and monitoring misinformation	Partners with media organizations to ensure transparency, covers politics, economics, and social issues
Korea Communications Standards Commission (KCSC)	Monitoring broadcasting and telecommunications content	Reviews broadcast programs, imposes sanctions on false/harmful content, issues correction orders to broadcasters
Press Arbitration Commission	Mediating media disputes and providing remedies for inaccurate reporting	Handles defamation and fake news cases, orders retractions or corrections when necessary
Newstapa	Investigative journalism focused on uncovering and correcting misinformation	Conducts in-depth investigations on social issues, emphasizes accuracy and integrity in journalism
Table 1: Organizations and agencies preventing and controlling the spread of fake news and misinformation		

This table summarizes the organizations and their roles in addressing fake news and misinformation in Korean media. The organizations listed play critical roles in maintaining the integrity of Korean media by addressing misinformation and ensuring accurate reporting. Korea Press Foundation (KPF) focuses on improving media quality through research and operates FactCheckNet to combat fake news. SNU FactCheck partners with media outlets to verify stories, ensuring transparency across various fields like politics and economics. KCSC monitors broadcasting content, enforcing standards by reviewing and penalizing harmful or inaccurate information. The Press Arbitration Commission resolves media disputes, ensuring that corrections are made in cases of defamation or misinformation. Lastly, Newstapa specializes in investigative journalism, working to expose and correct social issues with a focus on accuracy and integrity. These organizations collectively contribute to a media environment where reliability and truth are prioritized.

## **2. The emergence and development of AI in Korea**

The development of AI in Korea reflects both global advancements and Korea's unique socio-economic environment. Over the decades, Korea has grown from a country in the early

stages of AI exploration to a nation poised to become a global leader in AI innovation. This evolution has been marked by several key milestones, from initial research efforts in the 1980s to the rise of a robust AI industry in recent years.

The first wave of AI research in Korea began in the 1980s, coinciding with a global surge in interest in artificial intelligence. During this period, Korean universities and research institutes laid the foundation for AI development. One of the pioneering institutions was the Korea Electronics and Telecommunications Research Institute (ETRI), which initiated AI studies focusing on expert systems and natural language processing. These early efforts were heavily influenced by technological developments in Japan and the United States, with Korea primarily engaged in the conceptual stages of AI research at this time [3].

By the 1990s, AI research in Korea transitioned from theoretical exploration to practical applications. One notable development during this period was the introduction of expert systems. These systems were designed to emulate the decision-making abilities of human experts in specific fields, and they found commercial applications in industries such as healthcare, law, and manufacturing. For instance, POSCO, a major steel manufacturer, successfully implemented expert systems to optimize its production processes. This system allowed the company to adjust complex variables automatically, improving the quality and efficiency of steel production. The use of AI in this capacity demonstrated the potential for AI to revolutionize traditional industries by enhancing operational performance.

The early 2000s marked a significant turning point in Korea's AI development, driven by the country's emergence as a global IT powerhouse. With one of the highest internet penetration rates in the world, Korea created an ideal environment for AI research and innovation. The transition to an information society enabled rapid advancements in data processing capabilities, which are critical to AI development. In 2005, the Korean government officially recognized AI as a national strategic technology, significantly increasing its investment in AI research and development (R&D). This period saw major advancements in areas such as natural language processing, speech recognition, and robotics, with a focus on developing practical AI applications that could benefit various sectors of society.

One of the most notable achievements of this era was the development of natural language processing (NLP) technology for the Korean language. Korean, with its complex particles and verb conjugations, posed significant challenges for NLP researchers. However, by the mid-2000s, significant progress had been made. The Exobrain Project, spearheaded by ETRI, successfully developed NLP technologies that contributed to advancements in speech recognition, translation systems, and early AI-powered voice assistants for the Korean language. This breakthrough was particularly important for Korea, as it enabled the creation of AI systems tailored specifically to the linguistic needs of the Korean population.

The mid-2010s witnessed the rapid acceleration of AI development, driven largely by the global success of deep learning. The landmark moment for AI in Korea occurred in 2016, when AlphaGo, an AI developed by Google's DeepMind, defeated Korean Go champion Lee Sedol. This event shocked the nation and ignited a wave of public and governmental interest in artificial intelligence. In the aftermath, both the government and private companies significantly ramped up their investments in AI research. The AlphaGo match served as a wake-up call, pushing Korea to recognize the transformative potential of AI and prompting a nationwide push toward becoming a leader in this field.

In response to this growing interest, the Korean government unveiled its National AI Strategy in 2017, which set ambitious goals for establishing Korea as a global AI powerhouse by 2030. The strategy included substantial investments in AI infrastructure, such as the construction of data centers and the development of AI talent through educational programs. The government allocated over 1 trillion won to AI research and development, a clear

indication of its commitment to advancing the field. Additionally, major corporations like Samsung Electronics and LG Electronics established dedicated AI research institutes, integrating AI into consumer electronics and smartphones. For example, Samsung introduced Bixby, an AI-powered virtual assistant, as part of its broader efforts to incorporate AI into smart devices.

In parallel with these corporate efforts, the late 2010s also saw a boom in AI startups across Korea. Companies like Naver and Kakao established their own AI research labs, focusing on areas such as search algorithms and digital assistants. Meanwhile, startups like Riid revolutionized the education sector by developing AI-driven learning platforms. The COVID-19 pandemic in 2020 further accelerated AI adoption in fields such as healthcare, finance, and logistics. For example, AI-powered diagnostic systems became crucial in identifying and treating COVID-19 patients, while AI-based chatbots facilitated contactless services across multiple industries [4].

### **The government's AI policy and investment**

In 2017, the Korean government unveiled its “National AI Strategy,” setting ambitious goals to position Korea as a global AI powerhouse by 2030. The strategy included substantial investments in AI infrastructure and talent development, with the Ministry of Science and ICT earmarking over 1 trillion won for AI research and development. Key initiatives included the construction of data centers and the launch of programs to cultivate AI expertise.

In 2018, major corporations like Samsung Electronics and LG Electronics established dedicated AI research institutes, focusing on integrating AI into consumer electronics and smartphones. Samsung introduced Bixby, an AI-powered virtual assistant, as part of its efforts to integrate AI into smart devices [5].

In conclusion, the emergence and development of artificial intelligence in Korea has been a dynamic process, shaped by both global trends and local initiatives. From its early stages of conceptual research in the 1980s to its current role as a leader in AI-driven innovation, Korea has consistently embraced the transformative potential of AI. With strong government support, active corporate investment, and a thriving startup ecosystem, Korea is well-positioned to continue its rise as a global leader in artificial intelligence by 2030.

### **3. The integration of AI in Korean media: Progress and current status**

The integration of artificial intelligence (AI) into the Korean media landscape has revolutionized the way news is produced, distributed, and consumed. Over the past decade, AI has grown from a tool for simple automation to a sophisticated technology reshaping the industry. This transformation has taken place in stages, beginning with early AI adoption for news automation and evolving into widespread applications in content curation, data journalism, and personalized news experiences.

The initial integration of AI into Korean media began in the early 2010s, marking the start of news automation. A significant milestone occurred in 2016 when Yonhap News introduced the AI Newsbot Project. This project aimed to automate the production of articles, particularly in data-heavy areas such as stock market trends and economic indicators. The use of AI in this context allowed news organizations to quickly generate accurate reports, enhancing the speed at which information was delivered to the public.

AI-driven automation proved especially useful in areas requiring constant updates, where the ability to process large volumes of data in real-time became a game changer. This

early phase of AI adoption focused on increasing the efficiency of news production while also maintaining accuracy in reporting [6].

By 2017, the use of AI in Korean media had expanded significantly with the rise of AI-powered newsbots. These systems were not only automating content creation but also improving the overall user experience by curating personalized news. One prominent example was Naver's AiRS system, which analyzed users' news consumption habits to recommend tailored content. Similarly, Yonhap News continued to refine its Newsbot technology, which curated and personalized news content for individual readers. These AI-driven systems transformed the way news was delivered, making it more relevant and engaging for readers. By automating routine tasks, such as compiling reports or categorizing content, AI allowed journalists to focus on more substantive work, ultimately enhancing the quality of journalism.

Since 2019, the role of AI in Korean media has continued to expand, extending beyond news automation to more complex applications such as data analysis, content curation, and real-time reporting. AI is now employed to automate repetitive tasks, generate articles instantaneously, and curate vast amounts of data to provide deeper insights into current events. The rise of data journalism is one of the most notable developments during this period. AI, coupled with big data analytics, has enabled media organizations to visualize complex data sets, making them more accessible to the public. This approach has enriched reporting, particularly in areas like politics and finance, where data-driven insights are critical. For instance, SBS has utilized AI to analyze election data, improving the accuracy and depth of its political coverage [7].

However, the integration of AI in media has not been without challenges. While AI offers numerous benefits, such as increased productivity and enhanced personalization, it also raises concerns about job displacement within the media industry. As AI systems take over tasks traditionally performed by human journalists, there is growing anxiety about the future of employment in the sector. Additionally, the reliance on AI raises ethical concerns regarding transparency and accountability. Ensuring that AI algorithms remain unbiased and that AI-generated content maintains journalistic integrity is a critical challenge that media organizations must address.

Despite these obstacles, the future of AI in Korean media is promising. The continued investment in AI technologies by both private companies and the government signals a commitment to harnessing the full potential of AI to innovate the industry. As AI technology continues to evolve, it is likely to play an even more central role in shaping the media landscape, offering new opportunities for improving the speed, accuracy, and personalization of news.

In conclusion, the integration of AI into Korean media has significantly transformed the industry, enhancing the efficiency and scope of news production while also raising important ethical and practical challenges. From the early days of news automation to the current era of data journalism and personalized news experiences, AI has become an indispensable tool in the modern media environment. As the technology continues to develop, the Korean media industry must balance the benefits of AI innovation with the need to preserve journalistic standards and protect employment within the sector.

### **Regulations and control of AI in Korea**

As AI technology rapidly advances, Korea has developed a comprehensive framework of regulations and guidelines to ensure that AI is developed and deployed ethically, safely, and responsibly. These measures aim to foster innovation while addressing potential risks and societal impacts [8].

Area	Regulation or Guideline	Key Details
AI Ethics Guidelines	Artificial Intelligence Ethics Standards (2019)	<ul style="list-style-type: none"> <li>-Ensures AI respects human dignity and rights.</li> <li>- Promotes fairness, transparency, accountability, and safety.</li> </ul>
Data Protection and Privacy	Personal Information Protection Act (PIPA), Information and Communications Network Act, Credit Information Act	<ul style="list-style-type: none"> <li>- Regulates the collection, processing, storage, and sharing of data.</li> <li>- Ensures personal information is protected.</li> </ul>
AI Technology Regulation and Safety	AI Safety Management System(Under development)	<ul style="list-style-type: none"> <li>- Establishes technical standards and guidelines for safe AI use.</li> <li>- Minimizes risks like unintended behavior or errors.</li> </ul>
AI Legislation	Artificial Intelligence Framework Act (Draft, 2021)	<ul style="list-style-type: none"> <li>- Mandates social responsibility in AI development.</li> <li>- Enhances transparency in data and algorithms.</li> <li>- Clarifies liability for AI-related incidents.</li> </ul>
International Cooperation and Standards	Collaboration with OECD, G20, EU	<ul style="list-style-type: none"> <li>- Engages in global discussions to set international AI standards.</li> <li>- Promotes ethical AI through international cooperation.</li> </ul>
Table 2: Regulations and Guidelines of AI		

The above table highlights Korea's comprehensive approach to regulating AI, emphasizing ethical development, data protection, safety, and global cooperation. The AI Ethics Guidelines focus on human dignity, fairness, and transparency, ensuring responsible AI use. Strong data protection laws like PIPA safeguard personal information, while the AI Safety Management System aims to establish technical standards to prevent risks from AI errors. The AI Framework Act further mandates social responsibility and clarifies legal accountability in AI development. Korea's active international cooperation with organizations like the OECD and G20 promotes ethical AI standards on a global scale, demonstrating the country's commitment to both innovation and responsible governance.

#### 4. The convergence of AI and media: Challenges and prospects

The integration of artificial intelligence (AI) into the media industry offers promising opportunities but also presents significant challenges that must be carefully navigated. As AI continues to influence Korean media, issues related to objectivity, the evolving role of journalists, and the risk of misinformation come to the forefront. These challenges raise



concerns about the future of journalism and the need for a balanced approach that preserves the core principles of the media while embracing technological innovation [9].

One of the primary challenges posed by AI in media is its potential to compromise objectivity and ethics. AI models are trained on historical data, which often contains embedded biases. As a result, AI-generated content may unintentionally reflect these biases, leading to the overrepresentation of certain perspectives and the underrepresentation of others. This can undermine the fairness and neutrality that are essential to credible journalism. If AI is used to automate news production or content recommendations, there is a risk that these biases will be perpetuated, distorting the information that reaches the public. Therefore, it is crucial to develop AI systems that prioritize fairness and transparency, ensuring that they do not compromise the objectivity of the news.

Another challenge concerns the changing role of journalists in an AI-driven media landscape. As AI systems become increasingly adept at automating routine tasks, such as reporting on basic events or generating data-driven stories, the demand for traditional journalistic roles may diminish. AI can efficiently handle repetitive tasks that require speed and precision, which could reduce the need for human involvement in these areas. Consequently, the role of journalists must evolve to focus more on creative, investigative, and analytical work. Journalists will need to differentiate themselves from AI by providing deeper insights, offering nuanced perspectives, and engaging in storytelling that requires human intuition and empathy. The media industry must adapt to this shift by encouraging journalists to specialize in areas where AI cannot replicate human expertise, such as investigative journalism and critical analysis.

A third major concern is the issue of misinformation and fake news. While AI can greatly enhance the speed and efficiency of news production, it also increases the risk of disseminating false information. AI-generated content relies on the data it is trained on, and if this data contains inaccuracies, there is a high likelihood that the AI will produce misleading or false news. The rapid spread of such misinformation can have serious consequences, particularly in an era where news circulates quickly through digital platforms. Therefore, it is essential to implement rigorous fact-checking procedures and minimize errors in AI systems to ensure that the information provided to the public is accurate and reliable.

Looking to the future, Korean media must navigate the convergence of AI and journalism with a clear focus on balancing innovation with ethical responsibility. While AI presents numerous opportunities for improving the efficiency of news production and personalizing content for audiences, these advancements must not come at the expense of journalistic integrity. The media's core mission of providing truthful, unbiased reporting and fulfilling its social responsibility should remain intact, even as AI becomes more deeply integrated into the industry.

To achieve this balance, several key strategies must be employed. First, media organizations must prioritize the preservation of journalism's fundamental role as a guardian of public discourse. Although AI can assist in producing and curating news, human judgment and ethical considerations are essential to maintaining the public's trust in the media. Journalists must continue to play a critical role in ensuring that AI-generated content aligns with the ethical standards of the profession.

Secondly, the regulation and control of AI in media are crucial to managing the risks associated with this technology. While Korea has already established AI ethics guidelines, stricter rules and ethical standards specifically tailored to the media industry are needed. Media organizations must recognize the potential risks AI poses, such as bias, misinformation, and reduced human oversight, and proactively establish systems to mitigate these risks.

Finally, the role of journalists in the age of AI must be redefined. Rather than competing with AI in areas where it excels, such as data analysis and routine reporting, journalists should focus on cultivating critical thinking skills, investigative abilities, and creative storytelling. By leveraging AI as a tool while maintaining a human-centered approach to journalism, journalists can continue to provide value that goes beyond what AI can offer. At the same time, readers and audiences must also develop critical thinking skills to assess the reliability of the information they consume, especially in an era where AI-generated content is becoming more prevalent [10].

In conclusion, the convergence of AI and media presents both challenges and opportunities. Korean media must carefully navigate this evolving landscape by embracing the benefits of AI while safeguarding the ethical principles that are the foundation of journalism. With thoughtful regulation, a renewed focus on human judgment, and an evolving role for journalists, the media can harness the power of AI to enhance its capabilities without compromising its core mission of serving the public good.

## 5. Conclusion

The convergence of artificial intelligence and Korean media marks a transformative era for both industries, offering immense potential to reshape how news is produced, distributed, and consumed. AI's ability to enhance efficiency, improve content personalization, and enable data-driven journalism presents significant opportunities for innovation. However, this rapid advancement also brings forth critical challenges, particularly regarding ethics, objectivity, job displacement, and the risk of misinformation.

For Korean media to thrive in the AI-driven landscape, it is essential to establish a balanced approach that embraces technological advancements while safeguarding the core values of journalism—truthfulness, fairness, and social responsibility. Thoughtful regulation and governance of AI, combined with a renewed focus on the unique human elements of journalism such as investigative reporting and critical thinking, will ensure that AI enhances rather than undermines the integrity of the media.

As Korea continues to position itself as a global leader in AI, the success of its media industry will depend on its ability to integrate AI responsibly, maintaining the trust of the public while leveraging AI's potential to push the boundaries of modern journalism. By fostering innovation alongside ethical oversight, Korean media can navigate this dynamic intersection of technology and journalism, securing a future where both AI and human journalists coexist and contribute to a more informed and engaged society.

## References

1. Choi, J. (2011). A History of Korean Journalism. Seoul: Hanul Publishing Co. March 1st Movement, Donga Ilbo's Role: pp. 112-125, Japanese Colonial Rule and Media Censorship: pp. 88-110.
2. Kim, H. J. (2019). Media and Democracy in Korea: The Battle Against Fake News. Seoul: Media Studies Press. Korea Press Foundation and FactCheckNet: pp. 102-118, Role of SNU FactCheck in Verifying Misinformation: pp. 119-135.
3. Lee, S. J. (2020). Artificial Intelligence in South Korea: History, Development, and Future Prospects. Seoul: Korean Academic Press. Early AI Research in the 1980s (ETRI, NLP): pp. 45-62, POSCO and Expert Systems in the 1990s: pp. 98-112, AlphaGo's Impact and National AI Strategy: pp. 150-170.
4. Kim, H. S. (2021). The Rise of AI Startups in South Korea: Innovation and Challenges. Seoul: Tech Innovation Press. Naver and Kakao's AI Research Labs: pp. 90-110, Riid's Impact on the Education Sector: pp. 120-135.

5. Park, J. H. (2019). AI in Korea: From Conceptual Research to Global Leadership. Seoul National University Press. Samsung and LG's AI Research Institutes and Bixby Introduction: pp. 135-150, Corporate Contributions to AI Development in Korea: pp. 151-170.
6. Choi, Y. J. (2021). AI and the Future of Journalism in Korea. Seoul: MediaTech Press. Yonhap News AI Newsbot Project and Early AI Adoption (2016): pp. 102-118, Naver's AiRS System and Personalization in News: pp. 135-150, Ilbo's Role: pp. 112-125, Japanese Colonial Rule and Media Censorship: pp. 88-110.
7. Kim, H. Y. (2021). The Impact of AI on Media and Journalism in South Korea. Seoul National University Press. Advanced Applications of AI in Korean Media Since 2019: pp. 95-115, Data-Driven Insights in Political and Financial Reporting: pp. 130-145.
8. Lee, J. H. (2021). Data Protection and AI Regulation in Korea. Seoul National University Press. Personal Information Protection Act (PIPA) and Data Privacy in AI: pp. 50-72, AI Framework Act and Legal Accountability in AI Development: pp. 130-145.
9. Choi, Y. S. (2020). The Future of Media in the Age of AI: Balancing Innovation and Integrity. Seoul National University Press. Misinformation, Fake News, and AI's Role: pp. 135-155, Regulation and Control of AI in Media: pp. 160-180.
10. Korea Press Foundation Report (2019), Journalism Ethics in the AI Era: Korea's Response Directions. Relevant Pages: pp. 45-60 - Proposes solutions for ethical problems and regulatory approaches related to the adoption of AI in the media.

### **Website references**

Korea Communications Commission (KCC): <https://www.kcc.go.kr>; Provides information on the regulation and development of Korean media, including reports on media history and policies.

Yonhap News Agency: <https://www.yna.co.kr>; Korea's leading news agency, offering archives and historical reports on significant events in Korean media history.

Korea Press Foundation: <https://www.kpf.or.kr>; Offers resources and publications on Korean media, including research on media trends, ethics, and historical developments.

Electronics and Telecommunications Research Institute (ETRI): <https://www.etri.re.kr>; Contains detailed information on AI research and technological advancements in Korea.

Ministry of Science and ICT: <https://www.msit.go.kr>; Government policies and initiatives regarding AI development, including reports on the National AI Strategy and R&D investments.

Samsung Electronics: <https://www.samsung.com>; Provides insights into Samsung's AI research and development, including the Bixby virtual assistant.

Korea Institute of Science and Technology Information (KISTI): <https://www.kisti.re.kr>; Features reports on the development of AI technology in Korea and strategic responses from the government.

Naver AI Research: <https://www.naverlabs.com>; Offers details on Naver's AI initiatives, research labs, and advancements in artificial intelligence technology.

JTBC News: <https://news.jtbc.joins.com>; Provides reports and articles on major political scandals, including the coverage of the Park Geun-hye-choi Soon-sil case.

Riiid Labs: <https://www.riiid.co>; Focuses on AI-driven educational platforms and their influence on the education sector in Korea.

## About Online Dedicated Resources for Information on Cyber Threats and Cyber Intelligence

**Ivan Hristozov, Veselina Gagamova, Kristina Krumova, Chavdar Kostadinov**

G. S. Rakovski National Defense College, Sofia, Bulgaria

*ihristoz@rndc.bg, v.gagamova@rndc.bg, k.krumova@rndc.bg, ch.kostadinova@rndc.bg*

**Abstract.** This paper aims to explore the potential and capabilities of information systems in the context of cyber intelligence. By examining the processes of data collection, processing, and dissemination, this analysis seeks to identify the opportunities that modern information systems present for enhancing cyber intelligence operations. Some of their features and possibilities are also presented.

**Keywords:** Cyber intelligence, Information systems, Cyber threats

### 1. Introduction

The advent of sophisticated information systems has revolutionized the field of cyber intelligence. These systems play a crucial role in gathering, processing, and distributing information pertinent to cyber threats and security. The purpose of this analysis is to evaluate the effectiveness and opportunities provided by information systems in these critical areas.

Successful cyber intelligence and cyber threat intelligence involves processing, storing, and sharing a large amount of information. From a technology-evolutionary and cyber-defense perspective, cybersecurity information transfer and dissemination systems are of great importance for realizing cybersecurity objectives. Naturally, the Internet is emerging as a leading platform for knowledge transfer as it offers a number of forms of exchange: web portals, electronic bulletin boards, forums, video conferencing, screen sharing, electronic meeting systems, workflow systems, etc.

### 2. Cyber Intelligence Sharing

The advent of sophisticated information systems has revolutionized the field of cyber intelligence. These systems play a crucial role in gathering, processing, and distributing information pertinent to cyber threats and security. The purpose of this analysis is to evaluate the effectiveness and opportunities provided by information systems in these critical areas.

There are a number of organizations that provide information related to cyber intelligence and emerging cyber security threats. Network security organizations include SANS, Mitre, FIRST, SecurityNewsWire, ISC2 and CIS. Much information can be found in blogs and podcasts, government organizations, or private companies - Cisco, FireEye, etc.

The U.S. Cybersecurity and Infrastructure Security Agency (CISA) offers a free service called Automated Indicator Sharing (AIS) [2], which are machine-readable indicators of cyber threats and protective measures between public and private sector organizations.

The U.S. National Institute of Standards and Technology (NIST) has created a guidebook with recommendations for best practices to follow in a cyber incident - Special Publication 800-61, revision 2, "Computer Security Incident Handling Guide" [3].

MITRE has created and maintains a catalog of known cyberthreats under the name Common Vulnerabilities and Exposures (CVE) [1]. The catalogue establishes a standard and common and widely accessible terminology glossary, publishes information on disclosed vulnerabilities and provides a means for information sharing.

A similar service is offered by a separate unit of Cisco, the Cisco Talos Threat Intelligence Group [4]. Talos is one of the world's largest threat investigation teams and maintains a blog and podcast that enable the exchange of information on vulnerabilities, indicators of compromise, and threat mitigation techniques.

As a direct application of the principle of collecting, categorizing and sharing information is the Vocabulary for Event Recording and Incident Sharing (VERIS) system[5]. VERIS is a system of metrics built to allow the description of security incidents in a structured and repeatable way. A distinguishing feature of this product is that it allows the anonymous sharing of information. The VERIS database has free and open access to a collection of publicly disclosed incidents in the so-called VERIS format (predefined and redefined as possible information fields).

The VERIS database has established itself as the cyber community's hub for lessons-learned that is aimed at facilitating decision-making before, during and after a cyber incident occurs.

The Malware Information Sharing Platform (MISP) [6] is an open source platform for sharing indicators of compromise for newly discovered threats. MISP is supported by the European Union and is used by over 6000 organizations worldwide. MISP enables automated sharing of IOCs between humans and machines using STIX and other export formats.

As we have seen, there are many sources of threat intelligence information, each of which may have its own data format. Accessing and using multiple sources of threat intelligence can be time consuming. To help cybersecurity personnel make the best use of threat intelligence, Threat Intelligence Platforms (TIPs) have evolved.

A threat intelligence platform centralizes the collection of threat data from multiple sources and data formats.

It can be said that there are three main types of threat intelligence data:

- Indicators of compromise.
- Tools, techniques and procedures.
- Reputational information about Internet destinations or domains.

The volume of threat intelligence data can be enormous, so a threat intelligence platform is designed to aggregate the data in one place and - most importantly - present it in an understandable and usable format. Organizations can contribute to threat intelligence by sharing their intrusion data over the Internet, typically through automation. Many threat intelligence services use subscriber data to improve their products and stay current with the ever-changing landscape of emerging threats.

Threat intelligence services enable the exchange of threat information such as vulnerabilities, indicators of compromise (IOCs) and mitigation techniques. This information is shared not only with personnel but also with security systems. When threats occur, threat intelligence services create and distribute firewall rules and IOCs to devices that have subscribed to the service.

There are several open standards for exchanging Cyber Threat Intelligence (CTI) information in an automated, consistent, and machine-readable format.

The main standards are:

- Structured Threat Information Expression (STIX) [7] – this is a set of specifications for exchanging cyber threat information between organizations. The Cyber Observable Expression (CybOX) standard is included in STIX;

- Trusted Automated Exchange of Indicator Information (TAXII) [8] – this is the application layer protocol specification that enables CTI communication over HTTPS. TAXII is designed to support STIX.

- Cyber Observable Expression (CybOX) – this is a standardized scheme for specifying, capturing, characterizing, and transmitting network operation events and properties that supports many cybersecurity functions.

These open standards provide the specifications that support the automated exchange of cyber threat intelligence in a standardized format.

Automated information systems for the collection, processing and dissemination of cyber-related intelligence.

Network security analysts must quickly and accurately assess the significance of any security event and answer the following critical questions: Who is associated with this event? Does the user have access to other sensitive resources? Does this event present a potential compliance issue? Does the user have access to intellectual property or sensitive information? Is the user authorized to access this resource?

To answer these questions, security analysts use systems and platforms such as:

- 1). Security Incident Event Management (SIEM)
- 2). Security orchestration, automation, and response (SOAR).
- 3). Integrated security platforms.

Security incident event management is a technology used in enterprise organizations to provide real-time reporting and long-term analysis of security events. Network devices including firewall, IPS, ESA, WSA, routers, switches, servers, and hosts are configured to send log events to the SIEM software. The SIEM software correlates the millions of events using machine learning and special analysis software to identify traffic that needs to be examined. The SIEM processes all data generated by firewalls, network devices, intrusion detection systems and other devices. SIEM systems are used to collect and filter data, detect and classify threats, and analyze and investigate threats. SIEM systems can also manage resources to implement preventative measures and address future threats. SIEM file logs are used by organizations for which obtaining real-time information and subsequent processing of generated information is important [9].

Well-known proprietary SIEM systems are SolarWinds Security Event Manager [10] and Splunk Enterprise Security [11] SIEM systems, and an open source SIEM solution is Security Onion with Elasticsearch, Logstash, and Kibana (ELK) [12].

SOAR platforms are similar to SIEMs in that they aggregate, correlate, and analyze alerts. However, SOAR technology goes a step further by integrating threat intelligence and automating incident investigation and response workflows based on recommendations and methodologies developed by the security team. SIEM and SOAR often link together as they have capabilities that complement each other.

SOAR Security Platforms:

- Collect alarm data from each component of the system;
- Provide tools that allow cases to be investigated, assessed and investigated;
- Emphasize integration as a means to automate complex incident response workflows that enable faster response and adaptive protection strategies;
- Include pre-defined manuals that enable automated response to specific threats. Manuals can be triggered automatically based on predefined rules or can be triggered by security personnel.

SOAR solutions:

- Provide case management tools that allow cybersecurity personnel to research and

investigate incidents, often by integrating threat information into the network security platform.

- Uses artificial intelligence to detect incidents and aid in incident analysis and response. Once data is collected, the challenge lies in processing and analyzing it to extract meaningful insights. Information systems equipped with artificial intelligence (AI) and machine learning (ML) capabilities can sift through enormous datasets to identify patterns, anomalies, and potential threats. It is important to explore the processing techniques and the role of AI and ML in augmenting the analytical capabilities of cyber intelligence systems.

- Automates complex incident response procedures and investigations, which are potentially labor-intensive tasks.

- Offers dashboards and reports to document incident response to improve key performance indicators and can significantly improve network security for organizations.

Among the top ten SOAR solutions for 2024 are [13]: Cyberbit's SOC 3D, Cyware; IBM's Resilient; Logsign's SOAR; Palo Alto Networks' Cortex XSOAR; Rapid7's InsightConnect solution; Siemplify's SOAR solution; Splunk's Splunk Phantom; Swimlane's SOAR platform and ThreatConnect's SOAR solution.

The integrated security platforms go beyond SIEM and SOAR to bring together multiple security technologies, processes and people into a unified team whose components build on each other rather than interfere with each other. These integrated platforms promise to address the complexities of network security monitoring by integrating multiple functions and data sources into a single platform that will significantly improve alert accuracy while offering robust protection.

Leading integrated security platforms include Cisco SecureX, Fortinet Security Fabric, and Palo Alto Networks Cortex XDR, FireEye, Helix and others.

The dissemination of processed information is vital for timely decision-making and response. Information systems provide platforms for the secure and efficient distribution of intelligence reports to relevant stakeholders. There is a need to examine the various dissemination channels, including automated alert systems and collaborative platforms, and their impact on the effectiveness of cyber intelligence operations.

### **3. Findings**

The underlined development of the approach of incorporation and exchange of knowledge between stakeholders in the field of cybersecurity is understandable and expected, which is conditioned on the one hand by the speed of development and improvement of cyber-attacks and on the other - by the lack of a unified and binding legal framework, anchored by appropriate sanctions for deviation from the set rule.

Logically and naturally, the most active are the representatives of the expert circles of IT equipment manufacturers and, of course, the representatives of science. This is how the two different approaches to gathering and systematizing knowledge - through research and based on experience - take shape.

### **4. Conclusion**

Information systems offer substantial opportunities for improving the collection, processing, and dissemination of cyber intelligence information. By leveraging advanced technologies, these systems enhance the accuracy, speed, and reliability of cyber intelligence efforts. Future research should focus on addressing the challenges associated with these

systems, such as data privacy and integration issues, to fully realize their potential in the cyber intelligence domain.

The use of platforms to collect, process and disseminate information related to cyber threat intelligence facilitates and improves security management. In the context of the entire cyber community, this allows threat intelligence and information sharing to be expanded and improved. These projects are an important part of any organization's cyber resilience, enabling the sharing of important threat warnings and incident information at an early stage.

## **Acknowledgements**

This publication was financed by the Ministry of Education and Science in implementation of the National Scientific Program – Security and Defence that is funded by Ministry of Education and Science of the Republic of Bulgaria in implementation of National Strategy for the Development of Scientific Research 2017-2030 and was adopted by Decision of the Council of Ministers No. 731 of October 21, 2021.

## **References**

1. Common Vulnerabilities and Exposures — CVE ®, The Standard for Information Security Vulnerability Names, <https://cve.mitre.org/>;
2. <https://www.cisa.gov/topics/cyber-threats-and-advisories/information-sharing/automated-indicator-sharing-ais>;
3. Special Publication 800-61, revision 2, “Computer Security Incident Handling Guide”, <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-61r2.pdf>;
4. <https://talosintelligence.com/>;
5. VERIS The Vocabulary for Event Recording and Incident Sharing, <https://verisframework.org/>;
6. MISP, <https://www.misp-project.org/>;
7. About STIX, <https://stixproject.github.io/about/>;
8. Introduction to TAXII, <https://oasis-open.github.io/cti-documentation/taxii/intro.html>;
9. CCNA Cybersecurity Operations, Online course, <https://lms.netacad.com/course/view.php?id=1222692>;
10. SolarWinds Security Event Manager, <https://www.solarwinds.com/security-event-manager>;
11. Splunk Enterprise Security, [https://www.splunk.com/en\\_us/products/enterprise-security.html](https://www.splunk.com/en_us/products/enterprise-security.html)
12. Security Onion Documentation, <https://docs.securityonion.net/en/2.4/>;
13. The Best SOAR Solutions to Consider in 2024, <https://solutionsreview.com/security-information-event-management/the-best-soar-solutions-to-consider/>.



# On Verification of Cyber Intelligence Tactics, Techniques and Procedures

**Ivan Hristozov, Veselina Gagamova, Chavdar Kostadinov, Kristina Krumova**

G. S. Rakovski National Defense College, Sofia, Bulgaria

*ihristoz@rndc.bg, v.gagamova@rndc.bg, ch.kostadinova@rndc.bg, k.krumova@rndc.bg*

**Abstract.** In this report, the authors present a research related to a cyber-intelligence procedure and the techniques and tactics it employs. Sample results of their application are shown and analyzed.

**Keywords:** Cyber intelligence, Procedure, Techniques, Tactics

## 1. Introduction

In this paper, a cyber intelligence scenario is proposed using the Active Scan technique (Fig. 1) and an Active Scan procedure (Fig. 2). The Active Scan procedure itself involves the application of a number of sub-techniques: gathering information about the target, determining the IP addresses of the active devices, scanning the open ports of the active devices, detecting vulnerabilities, and finally applying a means of breaching the target IP network or system. It is necessary to obtain certain data about the target-victim and break it.

Figure 2 shows some of the results of applying cyber intelligence sub-techniques. The goal of this work is to prove that they are achievable using commonly available methods and tools. What is verification? Simply put, verification is checking the veracity of the data reported. The term is derived from two Latin words *verus* "true" and *facere* "do", which literally translated means "to make something true" [8].

A basic principle of verification is that data must be experimentally confirmed and must not contradict generally accepted facts [8]. Verification is used to check that a system, program or apparatus device actually has the expected properties and functionalities.

In the present work, an experiment is performed to prove that the procedure presented in Fig. 2 can achieve the results shown by using known methods and tools. The applied empirical verification method can be successfully applied similarly to the verification of techniques and procedures in other information domains.

The application of the sub-techniques is carried out using tools and means, and an attempt is made to review the most popular ones to date.

Tools and instruments for the implementation of the Active Scan technique procedure are. Means for discovering information about the target's IP network, means of discovering information about the IP addresses of active hosts on the target network, tools to detect open ports on active devices, vulnerability detection tools and means of exploiting vulnerabilities.

## 2. Means for Discovering Information about the Target's IP Network

One of the tools is Whois Lookup. This is a protocol that is used to discover information about the holders of resources on the Internet, such as a domain name, a server or an IP address.

It is not a hacking tool, but rather a means to retrieve information from databases about the holders of things on the Internet.

There are many sites on the Internet that offer this service, one of which is <http://whois.domaintools.com>. There, if you enter the domain name of the target, a lot of information about the target's web page will appear. One can see the e-mail address that can be used for the domain name information.

Typically, one can see the address of the company that registered that domain name, the IP address, the localization of the IP address, who the name server is and which company is providing it, etc.

If you run a ping command to the domain name, you will see that the response is from the same IP address as shown by the Whois lookup.

The IP address can also be learned from the `nslookup` command.

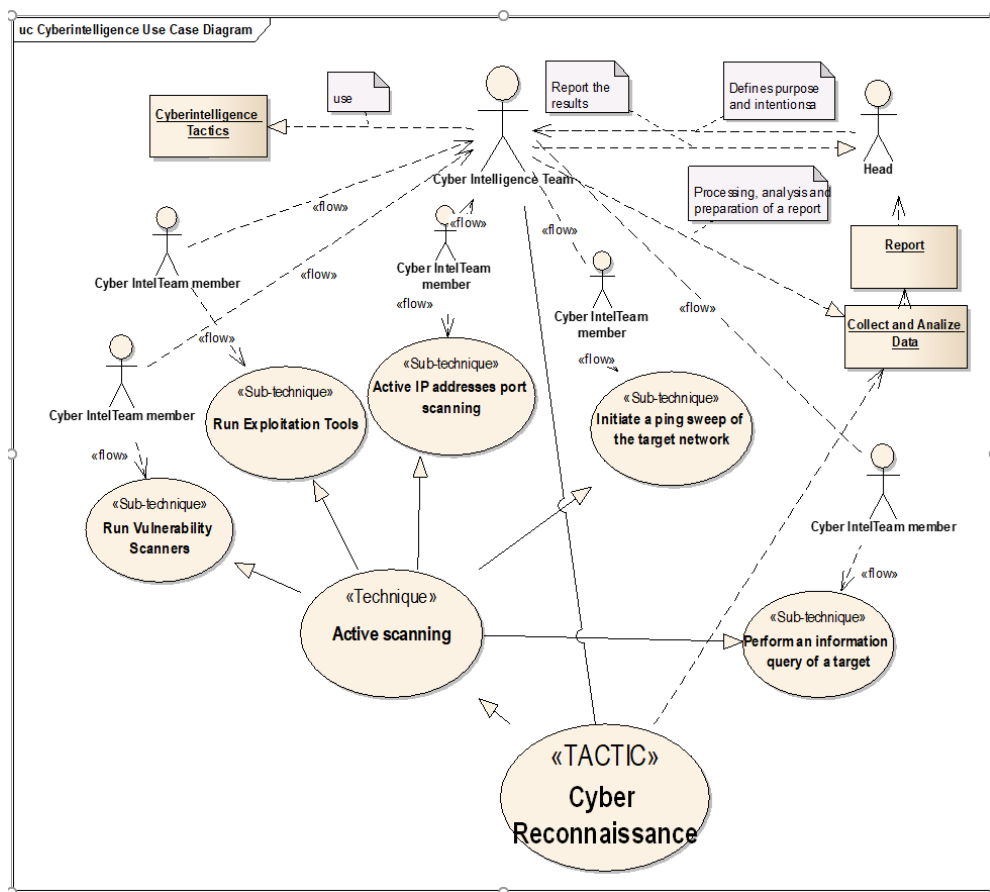


Fig. 1. Cyber intelligence scenario using Active Scanning technique

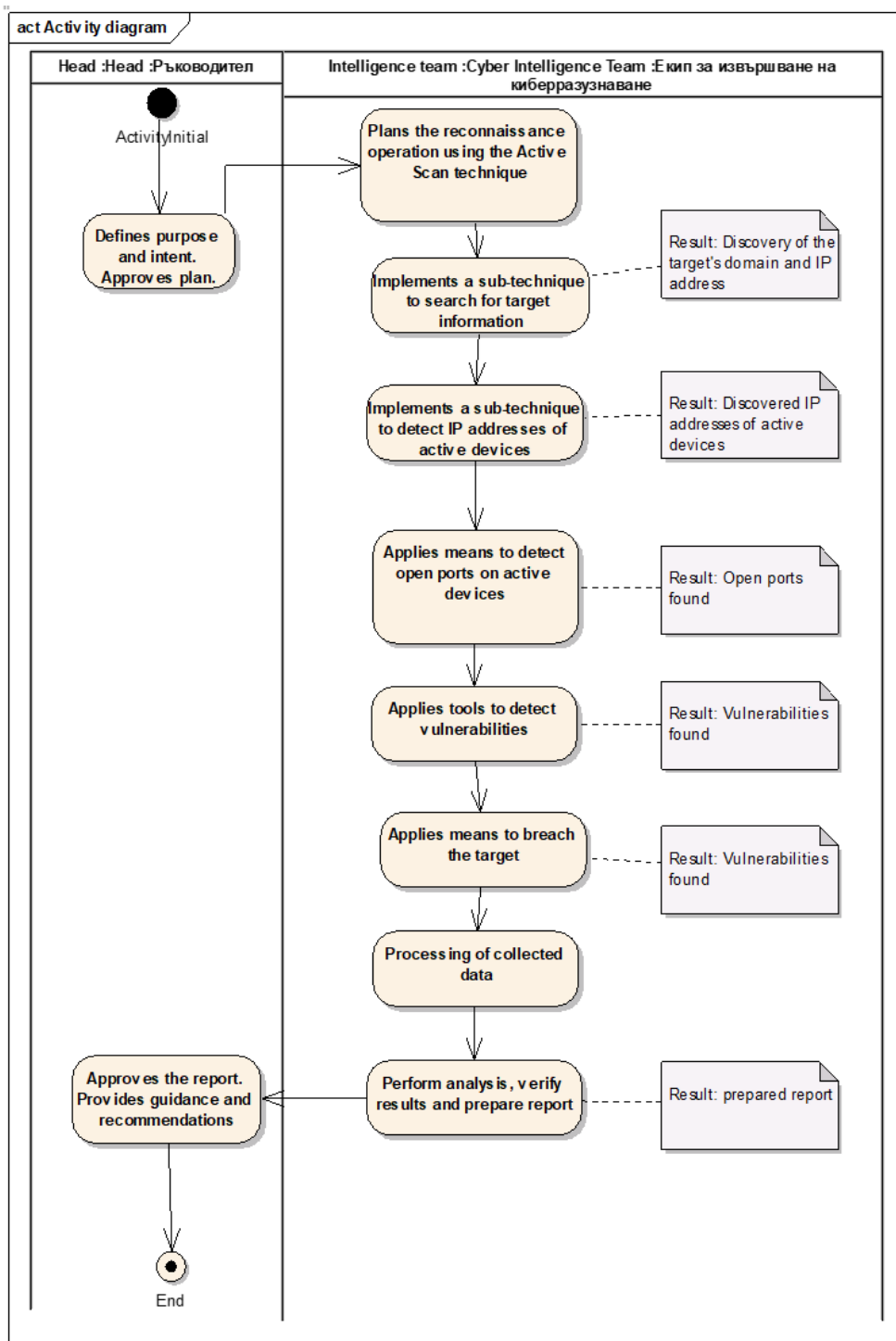


Fig. 2. Procedure for applying the Active Scan technique.

### **3. Means of Discovering Information About the IP Addresses of Active Hosts on the Target Network**

IP address scanning is a basic functionality offered by network monitoring tools. With these tools, security teams can map a network, its topology, and find unauthorized devices hidden in corporate networks.

To test whether a host is reachable over the IP network, network administrators use the ping command. This is a tool (command) that uses the Internet Control Message Protocol (ICMP).

The ping sweep tool (also known as ICMP sweep) is a basic network scanning technique used to determine the IP address ranges of active and inactive hosts on a network [2]. While a single ping indicates whether a single host exists and is reachable on a network, the ping sweep facility sends ICMP ECHO requests to many hosts simultaneously. If there is an active host behind an address, it should return an ICMP ECHO response (unless explicitly disabled). Ping sweep is also used for other purposes, such as improving network security.

Some of the known Ping Sweep tools are [3]:

- SolarWinds IP Address Manager (IPAM);
- SolarWinds Engineer's Toolset (ETS);
- ManageEngine OpManager;
- Advanced IP Scanner.

There are a number of other tools that perform the same as ping sweep, for example fping, gping, and nmap for UNIX systems and SolarWinds' Pinger software or Ping Sweep for Windows systems.

Other tools for network scanning are [4]:

- SolarWinds IP Address Manager;
- Network Performance Monitor;
- User Device Tracker;
- Engineer's Toolset;
- Angry IP Scanner;
- Network Monitor;
- ManageEngine Advanced IP Scanner;
- MyLanViewer;
- Advanced IP Scanner;
- IP Address Tracker.

### **4. Tools to Detect Open Ports on Active Devices**

The tools for detecting open ports on active devices are network scanners. A network scanner is a tool or application used to detect and categorize all devices on networks by their IP addresses, MAC addresses, manufacturers, ports, etc. to aid in network diagnostics, testing for intrusion capabilities, and forensic investigations of cybersecurity incidents.

A port scanner (also known as a network scanner) is a software application used to scan networks, hosts, or IP addresses looking for open, closed, or filtered ports. Port scanners are among the top cybersecurity tools used by researchers, security teams, and ethical hackers to diagnose network problems, audit networks, conduct penetration tests, and perform vulnerability scanning tasks.

All port-scanners work in a very elementary way:

- Send a request to a remote or local TCP or UDP port;
- Waiting for a response;
- Record the response from the host as "open", "filtered" or "closed".

Five most popular port scanning tools can be listed [5]:

- Nmap;
- Unicornscan;
- Angry IP Scan;
- Netcat;
- Zenmap.

## **5. Vulnerability Detection Tools**

Vulnerability scanners are valuable tools that look for and report what known vulnerabilities are present in an organization's information infrastructure. Using a vulnerability scanner is a simple but critical security practice that any organization can benefit from. These scans can give an organization insight into what security threats they may be facing by providing a view of potential security weaknesses present in their environment [1].

A number of organizations use multiple vulnerability scanners to ensure that they get full coverage of each asset, creating a complete picture.

Many different scanners have been developed over the years, with the most well-known, listed alphabetically, being the following [7]:

- Acunetix
- beSECURE
- Burp Suite
- GFI Languard
- Fortra VM
- Nessus
- Nexpose by Rapid7
- Nmap
- OpenVAS
- Qualys Cloud Platform
- Qualys Web Application Scanner
- SAINT's Security Suite
- Teneble.sc and Teneble.io
- Tripwire IP360

The choice of a vulnerability scanner can be driven by some of the most critical considerations, such as deployment, features, ease of use, customer support, and integrations.

## **6. Means of Exploiting Vulnerabilities**

Once hackers discover vulnerabilities in systems and platforms, they can find ways to exploit these vulnerabilities. This is known as exploiting a vulnerability and allows hackers to take advantage of vulnerable systems. Vulnerability exploitation can occur at different levels. Tools, including the hundreds found in Kali Linux, are needed to accomplish this. These tools

are varied and can be applicable in a wide range. In general, the areas where these tools can be applied can be defined as follows [6]:

- Web vulnerability exploitation tools – BurpSuite, OWASP ZAP, Commix, w3af, Jexboss;
- Operating system exploitation tools – Metasploit Framework, Mimikatz, Nmap, John the Ripper, Hashcat;
- Database tools – Sqlmap, DBever
- SQL ninja, BSQL Hacker, Safe3, SQL Injector
- Mobile application tools – Frida, MobSF, Runtime Mobile Security (RMS).

## **7. Conclusion**

Presenting the cyber intelligence scenario using the Active Scanning technique as a procedure allows for the systematic description and application of a number of sub-techniques and their expected outcomes.

The tools and instruments considered successfully verify the described procedure of the Active Scan technique. Analysis of the procedure leads to the conclusion that competence building in users is necessary to identify and determine when and which tool is appropriate to use.

The proposed verification approach is interdisciplinary in nature and can be applied to other domains, as the verification of other techniques and procedures can be presented in a similar way.

## **Acknowledgements**

This publication was financed by the Ministry of Education and Science in implementation of the National Scientific Program – Security and Defence that is funded by Ministry of Education and Science of the Republic of Bulgaria in implementation of National Strategy for the Development of Scientific Research 2017-2030 and was adopted by Decision of the Council of Ministers No. 731 of October 21, 2021.

## **References**

1. <https://www.coresecurity.com/profile/pablo-zurro>
2. [https://www.cybertraining365.com/cybertraining/Topics/Ping\\_sweep\\_\(ICMP\\_sweep\)](https://www.cybertraining365.com/cybertraining/Topics/Ping_sweep_(ICMP_sweep))
3. Top 4 Ping Sweep Tools to Ping a Range of IP Addresses, <https://www.tek-tools.com/network/best-ping-sweep-tools>
4. Top 10 Tools to Find IP Addresses and Scan a Network, <https://www.tek-tools.com/network/network-ip-scan-guide-and-tools>
5. Esteban Borges, Top 5 Most Popular Port Scanners in CyberSecurity, <https://securitytrails.com/blog/best-port-scanners>
6. Lester Obbayi, Top 18 tools for vulnerability exploitation in Kali Linux, <https://www.infosecinstitute.com/resources/penetration-testing/top-18-tools-for-vulnerability-exploitation-in-kali-linux/>
7. Pablo Zurro, Top 14 Vulnerability Scanners for Cybersecurity Professionals, <https://www.coresecurity.com/blog/top-14-vulnerability-scanners-cybersecurity-professionals>
8. <https://zvonobot.ru/blog/verifikatsiy-a-cto-eto-prostyimi-slovami-printsipy-i-primery>

## **Incidence of Fake News Shared on Social Networks and Risk Assumed by University Students**

**Carlos A. Torres-Gastelú<sup>1</sup>, Joel Angulo-Armenta<sup>2</sup>**

<sup>1</sup> Universidad Veracruzana, Veracruz, Mexico

<sup>2</sup> Instituto Tecnológico de Sonora, Sonora, Mexico

*ctorres@uv.mx, joel.angulo@potros.itson.edu.mx*

**Abstract.** The objective of this paper is to analyze the perception of university students towards fake news by assessing the relationship between the incidence of fake news shared by themselves within their social networks and the risk assumed by students when sharing news they do not fully trust. Quantitative study through the application of a survey using a Likert scale applied to 313 students from different careers of the Faculties of Administration and Pedagogy of the Universidad Veracruzana in Mexico. Processing was carried out with descriptive and inferential statistics. The results point out the correlation between the risk assumed and the cases of false news shared indicating the existence of cases in university students with little predisposition to inquire into the reliability of the information, and who in turn, risk sharing it anyway. The variables of sex, age, semester, and educational program do not influence the risk that students assume when sharing news. However, students who claim not to share news they distrust also share fake news they do believe in on their social networks.

**Keywords:** Fake News, Social Networks, Students, University, Mexico.

### **1. Introduction**

Although fake news is not limited to online environments, it is within these that it spreads more quickly and easily. The characteristics of the Internet, such as instantaneity, interactivity, viralization and globalization, make the medium a predisposed environment for the mass dissemination of any information [1].

For several authors [2, 3, 4], this is potentialized due to the technological tools that social networks provide to users for content editing, along with the ability to publish at no cost and without many restriction rules. In the same vein, one of the studies reviewed [5] state that social networks allow users to disseminate large amounts of information immediately.

Also, there are studies that [6, 7] consider that the ability of anyone to publish whatever they want makes these platforms an ideal ecosystem for the propagation of misleading information for political, ideological or lucrative purposes.

Several authors have pointed to science as the area that has been most affected by the phenomenon of fake news, as the erroneous information that is disseminated in some cases is usually related to the environment, technology, health, and other areas [8]. Such misinformation can have a great impact on society and influence, for example, the creation or practice of harmful or misleading habits. Other authors [9, 10], point to politics, terrorism, natural disasters, financial and ideological information, and again science, as the topics most impacted by the phenomenon.

At the same time, other studies [11, 12] show that young people's preference for social networks for information lies in the fact that on these platforms there are analysts, experts and opinion leaders who lend credibility to the news. In this regard, a study was detected in which

they found that young people have social networks as their main source of information, specifically, they tend to follow journalists' profiles to stay informed [13].

Despite the displacement of information channels, traditional media continue to maintain greater public trust over social networks, although it is true that social networks serve as a means of positioning journalists and institutions that wish to have other ways of reaching the public [14]. Because news is no longer disseminated exclusively by the press, radio and television, other media are overshadowing them, replacing professional journalism with content created and shared by users of social networks [15].

Although the decentralization of information sources with the advent of the Internet caused the loss of some of the rigor in the process of communicating events and knowledge, it is true that, for some authors, the loss of credibility in the traditional and official media opens the way to the rise of social networks as new channels of information [16].

So, in this sense the objective of this paper is to analyze the perception of university students towards fake news by assessing the relationship between the incidence of fake news shared by themselves within their social networks and the risk assumed by students when sharing news they do not fully trust.

## 2. Methodology

The type of study was quantitative, non-experimental, cross-sectional and descriptive in nature because the data collection was carried out at a single point in time and the variables were not manipulated. A questionnaire with optional response options and Likert scale was applied. The sampling was non-probabilistic by convenience, where the inclusion criterion was to be a student assigned to the Universidad Veracruzana in Mexico, studying in any semester and educational program, in the School of Pedagogy and Administration.

The sample size was 313 students between 18 and 30 years of age as shown in Table 1. The semesters taken by the surveyed students range from second to tenth semester, as shown in Table 2, for the careers of International Logistics and Customs (25.2%), Administration (42.8%), Tourism Administration (0.3%), Information Technologies in Organizations (3.5%), Pedagogy (27.2%) and Administrative Computer Systems (1%). Female students accounted for 67.7% (212), male students for 29.7% (93) and other sexual orientations for 2.6% (8).

Table 1. Distribution of the sample age variable

<i>Age</i>	<i>Frequencies</i>	<i>Percentage</i>
18 – 19 years old	118	37.7%
20 – 22 years old	166	53%
23 – 25 years old	25	8%
26 years and older	4	1.3%

Table 2. Distribution of the sample schoolperiod (semester)

<i>Semester</i>	<i>Frequencies</i>	<i>Percentage</i>
1ro, 2do y 3ro	132	42.2%
4to, 5to y 6to	79	25.2%
7mo, 8vo, 9no y 10mo	102	32.6%

A questionnaire was designed based on instruments developed in previous studies. The items considered appropriate were integrated into a single one, and others were discarded, modified and added to ensure the internal validity of the instrument, consisting of 89 items.



The instrument consisted of three sections; the first collects the demographic data of the population (Age, Sex, Career, Semester) being nominal and ordinal in nature. The second collects information on accessibility to social networks and the frequency with which news is shared within them. The third part explores several dimensions, one of which is addressed in this text: Incidence and Assumed Risk regarding the Sharing of False News, consisting of the 7 items shown in Table 3. The responses were based on a Likert-type scale with five possible answers for the statements: Strongly Disagree (TD), Disagree (D), Indifferent (I), Agree (A) and Strongly Agree (TA).

Table 3. Items of the Incidence and Assumed Risk dimension with respect to Sharing of Fake News

<i>Description of dimension items</i>
I shared a news item that turned out to be false on the social network...Facebook (advocacy).
I have shared a news item that turned out to be false on the social network...Twitter (advocacy).
I shared a news item that turned out to be false on the social network...WhatsApp (incidence).
I have shared a news item that turned out to be false on the social network...TikTok (advocacy).
I shared a news item that turned out to be fake on the social network...Instagram (incidence).
I shared a news item that turned out to be false on the social network...YouTube (incidence).
If I doubt the authenticity of a piece of news I... I share it anyway (risk assumed).

The authors of the instrument report that it is a reliable questionnaire since it obtained a Cronbach's alpha of 0.906. On the other hand, a value of 0.797 was obtained in the exploratory factor analysis of the dimensions, through the Kaiser-Meyer-Olkin sample adequacy index (KMO), that is, that there is a correct adequacy between the items.

The hypotheses addressed in this document are two. The first is that there are no differences between the variables of sex, age, semester (school period) and degree regarding the risk assumed when sharing news by students. The second hypothesis is that there is no correlation between the incidence of fake news shared and the risk assumed when sharing news by students.

In response to the objective of the study, using SPSS 25 software, the Kruskal-Wallis statistical test was carried out to determine whether there are differences between the groups with respect to the risk they assume before sharing news. It is worth mentioning that the normality test was performed on each of the variables, resulting in non-parametric data. To test the correlation of the variables, Spearman's statistical test was applied for nonparametric variables.

### 3. Results

For the school period (semester) variable, 78% of students in the first, second and third semesters stated that they disagreed with the statement of sharing news even if they doubted the authenticity of a news item; 15.9% were indifferent and 6.1% agreed. Meanwhile, for the fourth, fifth and sixth semesters, 70.9% did not identify with the statement compared to 16.5% who did and 12.7% were indifferent. And for students in semesters equal to or higher than the

seventh semester, 77% disagreed, 13% were indifferent and 10% agreed. To verify the hypothesis about the variable school period (semester), the value of 0.674 was obtained for 2 degrees of freedom in the Kruskal-Wallis statistical test, since the significance value of 0.714 is greater than 0.05, it is concluded that there are no differences between the school periods (semesters) that the participants were studying with respect to the risk assumed when sharing news by students (Table 4).

Table 4. Item: If I doubt the authenticity of a news item, I... I share it anyway (Semester)

	<i>Strongly disagree</i>	<i>Disagree</i>	<i>Indifferent</i>	<i>Agreed</i>	<i>Totally agree</i>
1st, 2nd y 3rd	60.6%	17.4%	15.9%	3.8%	2.3%
4th, 5th y 6th	65.8%	5%	12.7%	8.9%	7.6%
7th, 8th, 9th y 10th	67.6%	12.7%	8.8%	5%	5.9%
Total of the sample	64.2%	12.8%	12.8%	5.4%	4.8%
<i>Kruskal-Wallis of H</i>	.674		<i>Asymptotics sig.</i>		.714

For the age variable, 79.7% of students between 18 and 19 years of age stated that they disagreed with the statement of sharing news even if they doubted its authenticity; 14.4% adopted an indifferent position and 5.9% agreed. Meanwhile, for those between 20 and 22 years of age, 74.7% did not identify with the statement compared to 13.2% who did and 12% had an indifferent position. As for students between 23 and 25 years of age, 88% deny sharing news if they doubt its authenticity, while 8% do and 4% remain indifferent. And for students over 26 years of age, 25% disagreed, 50% were indifferent and 25% agreed. To verify the hypothesis about the age variable, the value of 6.542 was obtained for 3 degrees of freedom in the Kruskal-Wallis statistical test, since the significance value of 0.88 is greater than 0.05, it is concluded that there are no differences between the ages of the participants with respect to the risk assumed when sharing news by students (Table 5).

Table 5. Item: If I doubt the authenticity of a news item, I... I share it anyway (Age)

	<i>Strongly disagree</i>	<i>Disagree</i>	<i>Indifferent</i>	<i>Agreed</i>	<i>Totally agree</i>
18 – 19 years old	66.1%	13.6%	14.4%	5.1%	0.8%
20 – 22 years old	62%	12.7%	12%	6%	7.3%
23 – 25 years old	76%	12%	4%	4%	4%
26 years and older	25%	0%	50%	0%	25%
Total of the sample	64.2%	12.8%	12.8%	5.4%	4.8%
<i>Kruskal-Wallis of H</i>	6.542		<i>Asymptotics sig.</i>		.088

For the gender variable, 77.4% of women stated that they disagreed with the statement about sharing news even if they doubted the authenticity of a news item; 11.3% were indifferent and 11.3% agreed. Meanwhile, 75.3% of men did not identify with the statement

compared to 8.7% who did and 16.1% were indifferent. And for students with another sexual orientation, 87.5% disagreed and 12.5% were indifferent. To verify the hypothesis about the sex variable, the value of 0.602 was obtained for 2 degrees of freedom in the Kruskal-Wallis statistical test, since the significance value of 0.740 is greater than 0.05, it is concluded that there are no differences between the sex of the participants with respect to the risk assumed when sharing news by students (Table 6).

Table 6. Item: If I doubt the authenticity of a news item, I... I share it anyway (Gender)

	<i>Strongly disagree</i>	<i>Disagree</i>	<i>Indifferent</i>	<i>Agreed</i>	<i>Totally agree</i>
Woman	63.7%	13.7%	11.3%	7.1%	4.2%
Man	64.5%	10.8%	16%	2.2%	6.5%
Another	75%	12.5%	12.5%	0%	0%
Total sample	64.2%	12.8%	12.8%	5.4%	4.8%
<i>Kruskal-Wallis of H</i>	0.602		<i>Asymptotics sig.</i>		.740

For the educational program variable, 84% of the students enrolled in the International Logistics and Customs degree program stated that they disagreed with the statement of sharing news even if they doubted the authenticity of a news item; 12.7% were indifferent and 2.5% agreed. Meanwhile, for related management careers (Administration and Tourism Management), 76.9% did not identify with the statement compared to 13.4% who did and 9.7% had an indifferent position. As for the students of the related careers of Information Technologies in Organizations and Administrative Computer Systems, 78.6% denied sharing news if they doubted the authenticity of the information, In contrast, 7.1% do and 14.3% are indifferent. And for Pedagogy students, 69.4% disagreed, 17.6% were indifferent and 13% agreed. To verify the hypothesis about the educational program variable, the value of 6.764 was obtained for 3 degrees of freedom in the Kruskal-Wallis statistical test, since the significance value of 0.239 is greater than 0.05, it is concluded that there are no differences between the sex of the participants with respect to the risk assumed when sharing news by students (Table 7).

Table 7. Item: If I doubt the authenticity of a news item, I... I share it anyway

Educational Program	<i>Strongly disagree</i>	<i>Disagree</i>	<i>Indifferent</i>	<i>Agreed</i>	<i>Totally agree</i>
International Logistics and Customs	72.1%	12.7%	12.7%	0%	2.5%
Administration and Tourist Administration	65.9%	11.1%	9.7%	8.1%	5.2%
Information Technology in Organizations	50%	28.6%	14.3%	0%	7.1%
Pedagogy	56.5%	12.9%	17.6%	7.1%	5.9%
Total sample	64.2%	12.8%	12.8%	5.4%	4.8%
<i>Kruskal-Wallis of H</i>	6.764		<i>Asymptotics sig.</i>		.239

Regarding students who reported having shared a news item that turned out to be false within the Facebook social network, Table 9 shows that 67.4% denied sharing a false news item if they doubted its authenticity, compared to 13.9% who said they did so and 18.7% who were indifferent. Meanwhile Twitter, 55.1% of the incidences of fake news shared corresponded to students who disagreed with sharing news if they doubt its veracity, while 13.9% did agree and 22.45% maintained an indifferent stance. For the Instagram social network 38.1% of students who claim to share news even if they doubt its authenticity incurred in sharing fake news on the platform, meanwhile, 40.5% of students in disagreement and 21.4% indifferent made the same mistake. Regarding the WhatsApp social network, 56.6% of students who shared fake news also denied sharing news if they are not sure of its authenticity, compared to 21.2% who do and 22.5% who remain indifferent. Meanwhile, on the YouTube platform, students who do not share news unless they fully trust the information were 46.5% responsible for sharing fake news, while 32.5% were so in the case of students who always share. Finally, on the TikTok platform, 55.6% of students who were sure they had shared fake news denied sharing if they did not trust it, compared to 25.4% who shared anyway and 19% who adopted an indifferent stance.

Table 9. False News Sharing Incidents: *I have shared a new item that turned out to be a fake on the social network...*

<i>Social Network</i>	<i>I fully agree</i>	<i>Agree</i>	<i>Indifferent</i>	<i>Disagree</i>	<i>Strongly disagree</i>
Facebook	16%	23.3%	9.6%	16.6%	34.5%
Twitter	6.4%	9.3%	14.4%	22.7%	47.2%
Instagram	5.4%	8%	16.9%	25.9%	43.8%
WhatsApp	8.9%	16.6%	13.1%	20.8%	40.6%
YouTube	6%	7.7%	17.6%	21.4%	47.3%
TikTok	8.9%	11.2%	15%	18.3%	46.6%

Figure 1 shows the perception of false news shared by students by mistake. Facebook, with 39.3% of students claiming to have shared a fake news story at some point, is the social network with the highest number of incidences, with 9.6% unsure of having done so. Then, WhatsApp with 25.5% students sure of having shared a fake news and 13.1% unsure of having done so occupies the second position with the highest number of cases. It is followed by TikTok with 20.1% agreeing and 15% indifferent to the statement. Then, Twitter with 15.7% sure of having done it and 14.4% indifferent. Fifth place goes to YouTube with 13.8% saying they have done it and 17.6% don't know. Finally, Instagram with 13.4% of students who shared a false news item and 16.9% who do not know, is the social network with the lowest number of perceived incidences.

As for students who take the risk of sharing a news item they do not trust, Table 10 shows that their number is higher for the group of the same students who claim to have shared a news item that turned out to be false, this for all social networks. Specifically, 13.8% of the students who shared a fake news on Facebook have the habit of sharing news even if they doubt its authenticity, while 67.5% of them do not have that habit. For Twitter, 22.45% of students who shared a fake news usually take the risk versus 55.1% who do not. Regarding Instagram, 38.1% of the incidences came from students who share news regardless of its credibility, and 40.5% who do not. Referring to the social network WhatsApp, 21.25% of fake news were shared by students who assume the risk, while 56.25% came from students who do not share news but fully trust it. For YouTube, 32.6% of cases were caused by students who claim to share news even if they doubt its authenticity, and 46.5% from students who do not.

Finally, on the social network TikTok, 7.4% of students who claim to have shared a fake news story also claimed to share news regardless of its credibility, compared to 87% who deny doing so.

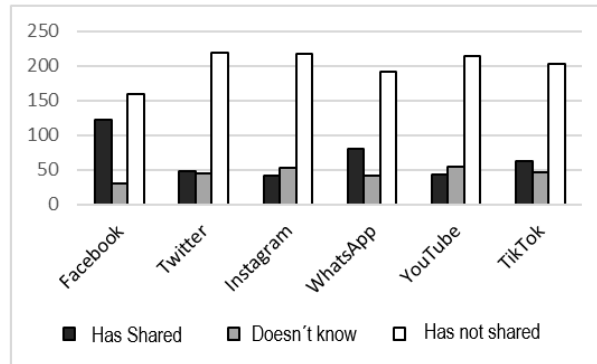


Fig. 1. Incidence of Fake News Shared by social network

Table 10 shows the incidences of fake news shared about the risks taken by students. The letters in the columns correspond to the values in percentages for each of the social networks: (A) Facebook, (B) Twitter or X, (C) Instagram, (D) Whatsapp, (E) Youtube and (F) TikTok .

Table 10. Incidence of False News Sharing on the risk taken by students

	<i>I have shared a news item that turned out to be false on the social network... (incidence)</i>					
	A	B	C	D	E	F
<i>If I doubt the authenticity of a news I... I share it anyway (assume es risk)</i>	13.8	22.5	38.1	21.2	32.6	7.4
Indifferent	18.7	22.5	21.4	22.5	20.9	5.6
<i>If I doubt the authenticity of a news I... I do not share it (does not assume the risk)</i>	67.5	55	40.5	56.3	46.5	87
	<i>I did not share a news item that turned out to be false on the social network... (advocacy)</i>					
	A	B	C	D	E	F
<i>If I doubt the Authenticity of a news I... I share it anyway. (assume es risk)</i>	7.5	5.5	5	5.7	5.2	5.4
Indifferent	8.8	9.1	8.3	6.8	7.4	6.9
<i>If I doubt the authenticity of a new item I... I do not share (no assumes the risk)</i>	83.7	85.4	86.7	87.5	87.4	87.7

Table 11 shows that there is a correlation between the variables of both items, since the significance for none of the social networks exceeds the value of 0.05, so the correlation hypothesis is accepted. Specifically, for the social network Facebook there is a low direct relationship with a coefficient of 0.170. While the direct relationship is moderate for the other

social networks: Twitter with 0.350, Instagram with 0.404, WhatsApp with 0.374, YouTube with 0.415 and TikTok with 0.372.

Table 11. Correlation of Incidence of False News Sharing on Risk Assumption by students

<i>Web</i>	<i>Sig</i>	<i>Coefficient</i>	<i>Web</i>	<i>Sig</i>	<i>Coefficient</i>
Facebook	0.003	0.170	Twitter	0.000	0.350
Instagram	0.000	0.404	WhatsApp	0.000	0.374
YouTube	0.000	0.415	TikTok	0.000	0.372

#### 4. Conclusion

There is evidence of the existence of cases in university students with little predisposition to inquire into the reliability of the information, and who in turn, risk sharing it anyway. Furthermore, the correlation of the variables reveals the impact this has on the incidence of false news being shared. According to the previous results it can be affirmed that the factors of gender, age, semester and educational program do not influence the risk taken by students when sharing news propitiating the proliferation of the Fake News phenomenon.

However, students who claim not to share news they distrust also come to share false news that they did believe, this being the group with the highest number of incidences. This translates as a lack of digital competence A1C2 [6], as they are not able to distinguish reliable information from unreliable sources. Although there are cases of false news shared for all categories and groups, the percentage that denies having done so is higher, then, it can be inferred that most students do have this competence and are able to accurately judge the reliability of the information. The latter may be due to the fact that they simply avoid sharing news they do not trust and do not know how to corroborate their information, which is also a reflection of the digital citizenship present in young people, since they take a cautious stance before the risk of sharing a fallacy.

This evidence shows that there is still a long and arduous road ahead for Mexican public universities in terms of training digital competencies that will allow them to perform adequately as digital citizens. Hence, it is suggested to develop an institutional educational strategy for university students to strengthen their knowledge, skills and attitudes as digital citizens to perform safely, legally and ethically [12] when surfing the Internet focused on the identification of reliable sources, the development of skills to identify possible fake news and the generation of a critical sense to avoid the proliferation of possible news of dubious origin and reporting information that is not verified.

#### References

1. Rodríguez, C. (2019). No diga fake news, di desinformación: una revisión sobre el fenómeno de las noticias falsas y sus implicaciones. *Comunicación Social*, 40, 65-74.
2. Fagundes, V. O.; Massarani, L.; Castelfranchi, Y.; Mendes, I. M.; Carvalho, V. B. D.; Malcher, M. A.; Miranda, F. C.; Lopes, S. C.: Jovens e sua percepção sobre fake news na ciência. *Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas*, vol. 16, no. 1, (2021).
3. Varona-Aramburu, D.; Sánchez-Muñoz, G.: Las redes sociales como fuentes de información periodística: motivos para la desconfianza entre los periodistas españoles. *El Profesional de la Información*, vol. 25, no. 5, pp. 795-802 (2016).
4. Shabahang, R.; Aruguete, M. S.; Shim, H.: Online News Addiction: Future Anxiety, Fear of Missing Out on News, and Interpersonal Trust Contribute to Excessive Online News Consumption. *Online Journal of Communication and Media Technologies*, vol. 11, no. 2, pp. e202105 (2021).
5. Alonso, L., y Saraiva, I. (2020). Búsqueda y evaluación de información: dos competencias necesarias en el contexto de las fake news. *Palabra Clave [La Plata]*, 9(2), e090.

6. González Fajardo, J. A.: El Dr. Google y las Falsas Noticias. *Angiología*, vol. 71, no. 3, pp. 91-94 (2019).
7. Mendiguren, T.; Pérez-Dasilva, J.; Meso-Ayerdi, K.: Actitud ante las Fake News: Estudio del caso de los estudiantes de la Universidad del País Vasco. *Revista de Comunicación*, vol. 19, no. 1, pp. 171-184 (2020).
8. Sousa, A. C. L., y Feitosa, E. M. A. (2021). Abordagem de fake news no ensino de química: concepções e práticas de professores. *Ensino Em Perspectivas*, 2(3), 1-12.
9. Pineda, H.; Jima-González, A.; Paradela-López, M.; García, C.; Parra, E.; Loaiza, K.; Areiza, S.; Ospina, T.; Agudelo, A.; Giraldo, S.; Acevedo, H.; Holguín, C.; Ortega, J. P.; Montoya, A.: ¿Preparados para las Fake News? Un estudio exploratorio de la comunidad universitaria del Tecnológico de Antioquia. *Revista En-contexto*, vol. 8, no. 12, (2020).
10. Tandoc, E. C., Lim, Z. W., y Ling, R. (2017). Defining “Fake News”. *Digital Journalism*, 6(2), 137-153.
11. Vázquez-Barrio, T.; Torrecillas-Lacave, T.; Suárez-Álvarez, R.: Credibilidad de los contenidos informativos en tiempos de fake news: Comunidad de Madrid. *Cuadernos.info*, vol. 49, pp. 192-214 (2021).
12. Catalina-García, B., Sousa, J. P., y Cristina Silva Sousa, L. C. S. (2019). Consumo de noticias y percepción de fake news entre estudiantes de Comunicación de Brasil, España y Portugal. *Revista de Comunicación*, 18(2), 93-115.
13. Cea, N.; Vicente, A. M. D.: La información de actualidad en las redes sociales: análisis de uso y percepción de credibilidad. *Posverdad en las redes*, vol. 8, pp. 103-129 (2021).
14. Curiel, E. H. (2015). La credibilidad de las redes sociales en el ámbito periodístico. *Transinformacao*, 27(2), 165-171.
15. Blázquez, M.: El problema de las noticias falsas: detección y contramedidas. *Verdad y falsedad de la información*, pp. 13-43 (2019).
16. Rúas, J., y Gualda, E. (2019). Teorías de la conspiración, credibilidad y confianza en la información. *Communication y Society*, 32(1), 179-195.

# Ethical Behaviour Aspects in the Contemporary Cyber Environment (Ethical Behaviour and Cybersecurity)

**Konstantin Poudin**

University of National and World Economy, Sofia, Bulgaria  
*kpoudin@unwe.bg*

**Abstract.** This paper draws attention on the ethical behaviour in the contemporary cyber environment. It does not address all aspects of ethical behaviour. The focus is on the ethical behaviour of employees in a given organization in cybersecurity context. The principles of ethical behaviour that employees must be guided by in order to perform their duties correctly and professionally, thereby contributing to ensuring the cybersecurity of the employer organization, are analysed.

**Keywords:** Ethical behaviour, Cyber environment, Cybersecurity

## 1. Introduction

Each person has a sense of what is good and what is bad, a sense of right and wrong. This sense can be stronger or weaker depending on the upbringing, the environment in which this person grew up or currently lives, on his/her life experience, on his/her personality.

Ethical behaviour is a behaviour that follows the good and aims to achieve good as a final result.

Ethical behaviour in the contemporary environment implies compliance with principles and rules that contribute to its stability and development. It is the basis of honest and constructive relations between people in this environment. In this way, it can also be seen as a way to guarantee security, which in the broadest sense implies the absence of evil, understood as risks, dangers or threats, and the presence of more good, understood as favourable conditions for implementing the interests of numerous and various consumers of security – individuals, groups of people, enterprises, states etc.

This research paper, without claiming to be exhaustive, presents ethical behaviour in the context of cybersecurity at the organizational level.

The report presents results received from the author's research activity, carried out within task 2.1.1. of the National Science Program Defence and Security, funded by the Ministry of Education and Science.

## 2. Humans and Contemporary Cyber Environment

People spend more and more of their daily life in cyberspace, carrying out various activities there. They work in the cyberspace, they inform themselves, educate, entertain or simply communicate.

In the four-layer model of cyberspace, David Clark places humans at the top level. He points out that they participate in the cyber-experience by communicating, working with information, making decisions and carry out plans, and also transforming cyberspace, using its constituent services and capabilities. Three other layers that Clark places immediately below humans are the information that is stored, transmitted, and transformed in cyberspace,



the logical building blocks that make up the services and support the platform nature of cyberspace, and the physical foundations that support the logical elements. [1]

The authors of the Digital 2023 Global Review Report point out that the average time spent on the Internet by users aged 16 to 64 was 6 h 37m in 2022. In 2013 the average time users aged 16 to 64 spent online was 6h 09m, which means that it has been steadily increasing over the years. In the pandemic conditions in 2020 and 2021, the average time users spent on the Internet was the highest - 6h 55m (2020) and 6h 57m (2021). The users in South Africa spend the most time online – 9h 38m. They are followed by the users in Brazil – 9h 36m, and the users in the Philippines – 9h 14m. The Japanese spend the least average time on the Internet – 3h 45m, the Danes – 4h 58m and the Germans – 5h 12m. [3]

Although people quickly get used to their life in cyberspace and try to take full advantage of the opportunities it provides, various security issues come to the fore. The digital world, in addition to offering many advantages, also hides real threats. It provides opportunities for malicious persons to make quick, large and secure profits from criminal activity directed against various organizations (banks, companies, etc.) or specific individuals.

Digitization leads to a rethinking of temporal and spatial dimensions by facilitating connectivity between people. Virtually at any time and regardless of distance, contact can be made with any person, who has the appropriate device and available coverage. This also changes the idea of personal space, which can very often be the object of interest and encroachment by bad-wishers.

In the context of the above, the ethical behaviour issues in cyberspace also come to the fore. The ethical behaviour is based on values, principles and resulting conscious actions that are relevant to ensuring cyber security in all its dimensions.

### **3. Essence of Ethical Behaviour**

The ethical behaviour in the contemporary cyber environment is not fundamentally different from ethical behaviour in the traditional environment. The ethical principles are something that have been established over centuries and are valid for each sphere, of course, relevant specifics may exist.

The basic principle underlying ethical behaviour and on which the entire ethical system is developed is that people should strive for good and avoid evil.

The ethical principles mostly refer to the understanding of good and evil, right and wrong. They can be found in the holy books of the religions, in the laws of countries, in the internal documents of companies, etc. These ethical principles shape people's thinking and the way they perceive themselves and others, the world around them and everything that happens in it. In this way, ethical principles also determine human behaviour.

The ethical principles should contribute to the harmonious development of the individual, to the regulation of interpersonal relationships, so that they are not characterized by tension, contradictions and conflicts, but are peaceful and constructive. Ethical principles are a prerequisite for creating and maintaining order in society. All this should contribute to improving people's life and ultimately making them happy.

The ethical behaviour within an organization, demonstrated by its employees, which in some sense may be equated with professional behaviour, is established in the process of training and/or developed in the workplace. The professional behaviour is mainly associated with: 1/ application of knowledge, skills and competences and 2/ adherence to certain behaviour by the persons who practice a given profession. This behaviour is expected by the member of the society. It is determined by the specifics of each activity, as well as the related attitudes, beliefs, values and principles.

Often the principles of ethical behaviour are given in ethical codes generally valid for persons belonging to a certain professional group. Some organizations introduce ethical codes for all their employees, regardless of the field of activity.

There are basic principles of ethical behaviour that may be valid for all spheres, such as:

- Compliance with legal requirements;
- Compliance with existing general rules in the organization;
- Compliance with the existing standards regarding the performance of official duties;
- Loyalty to the organization;
- Honesty in relations with management, employees, clients and other persons;
- Objectivity in the performance of official duties;
- Good faith in the performance of official duties;
- Responsibility in the performance of official duties;
- Dedication to duties performance;
- Respect to clients;
- Respect to partners;
- Respect to managers;
- Respect to the other employees.

These principles could be grouped as principles concerning the ethical behaviour of the specific employee towards the organization and fulfilment of professional duties, towards managers and other employees, towards clients and partners, towards the society in general and its members.

Logically, unethical behaviour implies behaviour that is not in accordance with the above ethical principles. It can be expressed in non-observance of the laws and rules in the organization, disrespect towards management, colleagues, partners and clients, carelessness regarding work, disloyal attitude towards an organization, including taking conscious actions against its interests and ruining its reputation.

#### **4. Cybersecurity and Ethical Behaviour in Contemporary Cyber Environment**

Different forms of unethical behaviour may exist in the contemporary cyber environment. The reason for a such behaviour are also different. Unethical behaviour may be demonstrated by employees in an organization, as well as by people who are outside it.

Very often persons, who are outside of the organization, take malicious actions against it. E.g. these actions may be against information security of the organization which ensures the confidentiality, availability and integrity of information. [4] The attackers could use malicious software to retrieve, destroy and modify information. The motivation may be mostly related to getting material benefits, getting competitive advantages, and/or cause losses. Sometimes the attackers cause damages just to feel satisfied with the achieved goal – the security breach.

If employees take malicious actions against the organization, where they work, they are so called “insider threats”. Unethical behaviour of the employees could be provoked by bad and unfair treatment of employee/s followed by desire for revenge. It could be powered by selfishness, emotional immaturity, envy, lack of sufficient life experience, poor upbringing, etc. Sometimes the employees, who takes malicious actions, are driven by moral motives. E.g. his/her personal moral profile does not correspond to the ethical standards in the organization. It is possible for an organization to have ethical standards, but it does not respect them. It only

declares them, demonstrating in practice different, even unethical, behaviour. Sometimes an employee could have understandings that are radically different from the generally accepted understandings in the society. They can have their own moral, which could cause problems.

As stated above, the values and principles of ethical behaviour are included in ethical codes generally valid for persons belonging to a certain professional group or valid for employees in a certain organization. E.g. Malaysia has adopted Code of ethics for information security professionals. It sets the values and principles of ethical behaviour for individuals whose professional responsibilities are in the field of information security. However, having in mind that all employees in a given organization with their actions contribute to ensuring security and cybersecurity in particular, these values and principles may be valid not only for the information security professionals but also for all other employees.

The Malaysian Code of ethics for information security professionals sets 3 basic values, such as: *integrity*, *objectivity* and *professional competence and due care*.

The first value *integrity* means:

- Perform duties in accordance with existing laws and exercising the highest moral principles;
- Refrain from activities that would constitute a conflict of interest;
- Act in the best interests of stakeholders consistent with public interest;
- Act honourably, justly, responsibly, and legally in every aspect of your profession.

The second value *objectivity* implies:

- Perform all duties in fair manner and without prejudice;
- Exercise independent professional judgment, in order to provide unbiased analysis and advice;
- When an opinion is provided, note it as opinion rather than fact.

The third value *professional competence and due care* includes:

- Perform services diligently and professionally;
- Act with diligence and promptness in rendering service;
- Render only those services which you are fully competent and qualified;
- Ensure that work performed meets the highest professional standards. Where restrictions exist, ensure that your work is both correct and complete within those limits. If, in your professional opinion, resources are inadequate to achieve an acceptable result, so inform clients and principals;
- Be supportive of colleagues, and encourage their professional development. Recognize and acknowledge the contributions of others, and respect the decisions of principals and co-workers;
- Keep stakeholders informed regarding the progress of your work;
- Refrain from conduct which would damage the reputation of the profession, or the practice of colleagues, clients, and employers;
- Report violations of the rules to the appropriate governing body in a timely manner. [2]

The Code includes 7 guiding principles for ethical behaviour, which are related to the three basic values. These 7 principles are as follows:

- Act all times in accordance with existing laws, association values and exercising highest moral principles;
- Protect and maintain appropriate level of confidentiality, integrity and availability of sensitive information in any course of professional activities;

- Perform the service with fairness, courtesy and good faith towards clients, colleagues and others, give credit where it is due and accept, as well as give, honest and fair professional comments;
- Do not engage in any crimes or improper practices;
- Perform all professional activities and duties in accordance with the highest ethical principles;
- Avoid professional association with those whose practices or reputation might harm the profession;
- Provide service with competence, honesty and correctly about limitations, experience and education [2].

## 5. Conclusion

The ethical behaviour, based on fundamental values and principles that deny evil and affirm good as a way of life and ultimate goal, should guarantee peace, understanding, stability, development and prosperity of society. It is a prerequisite for the stability, normal functioning and development of each organization.

The ethical behaviour in the contemporary cyber environment does not differ significantly from ethical behaviour in a traditional, non-virtual, environment. It presupposes, above all, the demonstration of honesty and respect to all persons being in the same environment, loyalty to the organization, responsibility to the duties, compliance with existing laws and restrictions regarding cyberspace and its components, and also compliance with the rules that a particular organization has adopted and applies in the process of its activity.

If, in a general sense, cybersecurity means the absence of potential dangers and specific threats to cyberspace and in particular to its components and the processes that take place in it, then strict ethical behaviour should aim at creating a favourable environment and eliminating any harmful factors.

## Acknowledgement

The scientific publication is in fulfillment of the objectives of the National Scientific Program "Security and Defense", financed by the Ministry of Education and Science of the Republic of Bulgaria in implementation of the Decision of the Council of Ministers No. 731 of October 21, 2021 and according to Agreement No. D01-74 /19.05.2022.

## References

1. Clark, D. (2010). Characterizing Cyberspace: Past, Present and Future [Online]. Available form: <https://ecir.mit.edu/sites/default/files/documents/%5BClark%5D%20Characterizing%20Cyberspace-%20Past%2C%20Present%20and%20Future.pdf> [Accessed: 12 October 2023].
2. Code of Ethics for Information Security Professionals. CyberSecurity Malaysia [Online]. Available form: [https://www.cybersecurity.my/data/content\\_files/11/764.pdf](https://www.cybersecurity.my/data/content_files/11/764.pdf) [Accessed: 12 June 2024].
3. Digital 2023 Global Review Report [Online] 2023. Available from: <https://wearesocial.com/wp-content/uploads/2023/03/Digital-2023-Global-Overview-Report.pdf> [Accessed: 17 February 2024].
4. Pavlov, G. (2021). Information security. Vulnerability, Attacks and Countermeasures. Tenth chapter in Corporate security (Authors: Dimitrov D., Ivanov T., Tsvetkov Ts., Pavlov G., Pudín K., Dimitrov N., Tsenkov, Y., Bogomilova E., Gechkova T., Dimitrov, A., Ivanov N.), pp. 251-265, Publishing complex-UNWE, Sofia (in Bulgarian), [Pavlov, G. (2021). Deseta glava. Informatsionna sigurnost. Uyzvimost, Ataki i Protivodeistvie v Dimitrov, D., Ivanov, T., Tsvetkov, Tsv., Pavlov, G., Poudin, K., Dimitrov, N., Tsenkov, Yuri, Bogomilova, Ek., Gechkova, T., Dimitrov, At., Ivanov, N. Korporativna sigurnost. Sofiya: IK-UNSS, s. 251-265].

# Application of Cyber Intelligence for Security Operation Center

**Nedko Tagarev**

University of National and World Economy, Sofia, Bulgaria  
*ntagarev@unwe.org*

**Abstract.** Ensuring practical cyber intelligence for a Security Operation Center (SOC) involves collecting, analyzing, and utilizing information about cyber threats to enhance the organization's security posture. This encompasses the gathering and analysis of data on potential and active cyber threats, including Indicators of Compromise (IoCs)" which refer to signs that a security breach may have occurred, while "Tactics, Techniques, and Procedures (TTPs)" refer to the methods and behaviours used by an attacker to carry out a cyber attack and utilized by threat actors. Additionally, it entails the use of advanced tools to monitor networks, systems, and applications for suspicious activities, such as the utilization of Security Information and Event Management (SIEM) systems, Intrusion Detection Systems (IDS), and Intrusion Prevention Systems (IPS). Furthermore, it involves developing and implementing plans to respond to security incidents, including identifying, containing, eradicating, and recovering from cyber-attacks. Regularly scanning systems for vulnerabilities and applying patches or mitigations to reduce the attack surface are also crucial components of cyber intelligence in the SOC. Lastly, monitoring and analyzing user and entity behaviour to detect anomalies that could indicate insider threats or compromised accounts is another significant aspect of effectively integrating cyber intelligence into SOC operations.

**Keywords:** GAP analyses, Cybersecurity

## 1. Introduction

Understanding the distinction between *cybersecurity* and *cyber intelligence* is essential in the face of escalating cyber threats. Both cybersecurity and cyber intelligence play critical roles in safeguarding against cyber-attacks, making it imperative to grasp the significance of each [1].

*Cyber intelligence* entails the comprehensive comprehension of the risks and threats confronted by your organization to enable thorough preparedness. A profound understanding of your adversaries is pivotal for effective defence against cyber attacks. Cyber intelligence gives insights into their origins, motives, methodologies, ongoing actions, and potential affiliations [2].

*Cyber intelligence is about understanding the cyber risks and threats to your organization.* This involves learning to gather information about cyber threats relevant to your organization, analyze it, and use that insight to respond effectively. This information may consist of the techniques and tactics employed by adversaries and the tools and infrastructure they use. Cyber intelligence also covers data regarding attackers' behaviour, including the sources of their attacks, how they execute them, and what they seek to accomplish through them.

*Cyber intelligence assists* in comprehending attackers, their motives, actions, capabilities, and tactics. It goes beyond simply mining data and involves the ability to analyze

real-time events. Cyber intelligence also aims to minimize the impact of attacks on companies by offering proactive guidance on defending against threats.

*In essence, cyber intelligence is a component of cybersecurity.* Both terms address the necessity of safeguarding computer systems and their users from malicious actors, although cybersecurity encompasses various aspects, with cyber intelligence being a smaller subset.

On the other hand, *cybersecurity involves* protecting computer systems from various threats such as viruses, spam, malware, ransomware, cybercrime, and cyberterrorism. It safeguards an organization's critical infrastructure from unauthorized access or malware. On the other hand, *cyber intelligence focuses* on understanding these threats to enhance defences and monitor for malicious activities.

***While both are essential, their purposes differ: cybersecurity aims to protect systems, while cyberintelligence concentrates on comprehending the nature of threats.***

## **2. Critical difference between cybersecurity and threat intelligence**

*Cyber intelligence (Threat intelligence)* is an ongoing process of collecting, processing, analyzing, and sharing information about potential threats that pose risks to applications and systems. This information is collected in real time from any number of sources. Threat intelligence is typically aggregated in a single database, providing security professionals with a centralized source of information on vulnerabilities and active threats that bad actors exploit.

*Cybersecurity goes beyond threat intelligence by taking proactive measures to combat attacks instead of observing them passively.* A proactive approach is based on data collected by *threat intelligence*. In this way, cybersecurity is responsible for the protection of vital networks, applications, devices, and data from unauthorized access or cyberattacks. Cybersecurity measures aim to stay ahead of new attack vectors to prevent intrusions. *Security professionals also develop responses to attacks to mitigate any damage as quickly as possible. Threat intelligence is a subset of cybersecurity.* Cybersecurity uses threat intelligence to detect and stop cyberthreats.

Understanding the difference between these two critical areas of your *cyber defence toolset* is vital because a comprehensive solution is needed to ensure success in both facets.

## **3. Threat intelligence tools**

In the battle against cybersecurity threats, security professionals utilize a wide array of threat intelligence tools and services to safeguard critical assets. Some of the most prevalent threat intelligence tools include:

- *Cybersecurity programs* - These programs usually help detect and manage threats.
- *Threat Intelligence Suppliers* are third-party companies that provide intelligence-gathering services to monitor potential threats.
- *Reverse Engineering Malware* where *Incident response teams* can address malware by reverse engineering the threat, analyzing it, and developing strategies to defend against it.
- *Web Proxy*, which is used to inspect inbound traffic. Web proxies can prevent new infections, for example, if someone unintentionally visits a malicious website.

## **4. Threat Intelligence Analyst**

*A Cyber Threat Intelligence Analyst* detects, reports and proposes measures to mitigate cyber threats. The job of the Threat Intelligence Analyst is vital to today's large companies, especially those that may be vulnerable to cybercriminals [3].

*Cyber Threat Intelligence* (CTI), or simply Threat Intelligence, is a discipline that uses concrete evidence and context analysis to mitigate cyber threats. In other words, cyber threat intelligence produces information about computer, network and IT threats that allows them to be prevented or mitigated.

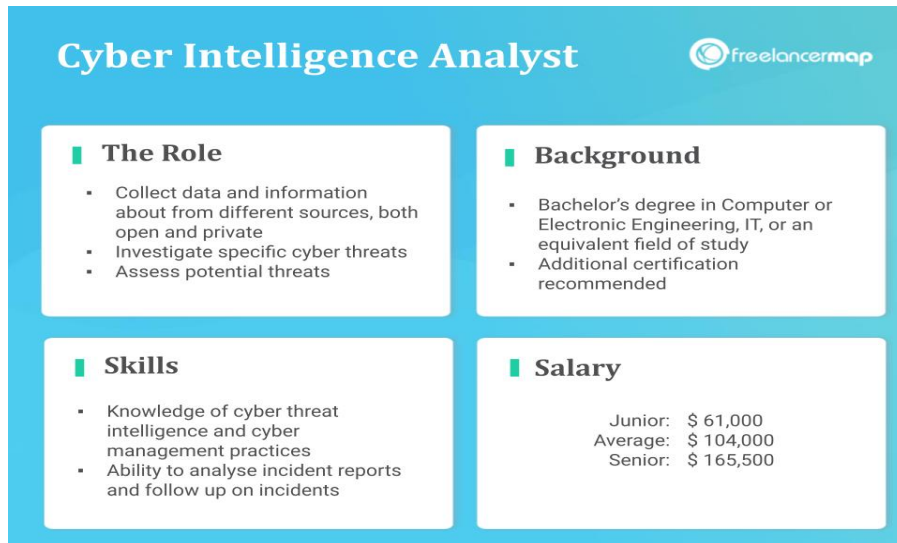


Fig. 1. Cyber intelligence analyst. (Source: <https://www.freelancemap.com/blog/what-does-cyber-threat-intelligence-analyst-do/#:~:text=Cyber%20E2%80%8B%E2%80%8Bintelligence%20experts,and%20third%2Dparty%20data%20sources> [4])

This approach is different from cybersecurity and far from replacing it—it complements it. It has to do with the evolution of cybercrime and its perpetrators. Cybercriminals are becoming more sophisticated. Sometimes, their interests are not merely economic but political. For this reason, cyber intelligence is a very relevant discipline for both the public and private sectors.

Cyber Intelligence is based on data collection and information analysis, but it goes much further than that. *It uses threat information to identify patterns and trends.* The company makes decisions and defines mitigation and prevention plans based on threat intelligence.

With the data collected with Threat intelligence, companies can make more informed security decisions and be proactive about avoiding potential cyber threats. *These essential aspects can be implemented in one of the leading cybersecurity tools' security operation centres.*

## 5. Security Operations Center (SOC)

Security Operation Center (SOC) is a centralized function that uses people, processes, and technology to monitor and enhance an organization's security posture continuously. *It aims to prevent, detect, analyze, and respond to cybersecurity incidents* [5].

*The Security Operations Center (SOC) is the central command post* for gathering telemetry from an organization's IT infrastructure, including networks, devices, appliances, and information stores, regardless of location. With the increasing number of sophisticated threats, gathering context from various sources is crucial when it comes to threat intelligence.

The SOC is responsible for correlating all monitored events within the organization and must determine how to manage and respond to each event.

#### *Security operations staff and organizational structure*

The primary role of a security operations team, often part of a security operations centre (SOC), is to continuously monitor, detect, investigate, and respond to cyber threats. These teams safeguard various assets, including intellectual property, personnel data, business systems, and brand integrity. As the hands-on component of an organization's cybersecurity framework, security operations teams are central in working together to monitor, assess, and protect against cyberattacks.

SOCs have typically been built around a hub-and-spoke architecture. This model includes various components such as vulnerability assessment solutions, governance, risk, and compliance (GRC) systems, application and database scanners, intrusion prevention systems (IPS), user and entity behaviour analytics (UEBA), endpoint detection and remediation (EDR), and *threat intelligence platforms (TIP)*.

A SOC manager typically oversees the Security Operations Center (SOC), which consists of incident responders, SOC analysts (at levels 1, 2, and 3), *threat hunters*, and incident response managers. The SOC reports to the Chief Information Security Officer (CISO), who then reports to either the Chief Information Officer (CIO) or directly to the Chief Executive Officer (CEO).

#### *Essential functions performed by the SOC*

a. Planning of available resources. ( This involves an account of available resources.)

The SOC is responsible for two types of assets: the *various devices, processes, and applications* it is charged with safeguarding and the *defensive instruments* it has available to utilise to ensure this protection.

The Security Operations Center (SOC) is crucial in safeguarding devices and data by maintaining visibility and control throughout the entire network, from the device level to the cloud. Without this pervasive visibility and control, potential blind spots in network security may be exploited. Thus, the *primary objective of the SOC is to attain a comprehensive understanding* of the organization's threat landscape, encompassing diverse endpoints, servers, on-premises software, third-party services, and the traffic traversing among these entities.

The SOC must comprehensively comprehend all available cybersecurity tools and internal workflows to fulfil its protective mandate. This adept understanding enhances operational agility and fosters maximal efficiency within the SOC.

b. Preparation and Preventative Maintenance

Even the most well-equipped and agile response processes cannot prevent problems from occurring. The SOC implements preventative measures to help keep attackers at bay, which can be divided into two main categories. There comes the role of *cyber intelligence*.

#### *Preparation*

Team members should stay informed on the *latest security innovations, trends in cybercrime, and the development of new threats in the future*. This research by *cyber intelligence* can create a security roadmap for the company's cybersecurity efforts and a disaster recovery plan for worst-case scenarios.

#### *Preventative Maintenance*

This step includes all actions to prevent more complex, more successful attacks, including regularly maintaining and updating existing systems and firewall policies, patching vulnerabilities, and allowlisting, blocklisting, and securing applications.

c. Continuous Proactive Monitoring

The Security Operations Center (SOC) tools constantly monitor the network for any



abnormalities or suspicious activities 24/7. This continuous monitoring allows the SOC to be immediately alerted to any emerging threats, providing them with the best opportunity to prevent or mitigate harm. Monitoring tools can include a Security Information and Event Management (SIEM) system, an Endpoint Detection and Response (EDR) system, and even a Security Orchestration, Automation, and Response (SOAR) or an Extended Detection and Response (XDR) system. The most advanced tools can utilize behavioural analysis to help systems differentiate between regular day-to-day operations and actual threat behaviour, reducing the manual triage and analysis required by humans.

d. Alert Ranking and Management

When monitoring tools issue alerts, it is the responsibility of the SOC to carefully examine each alert, dismiss any false positives, and assess the severity of any actual threats and their potential targets. This enables them to prioritize emerging threats effectively, first addressing the most critical issues.

e. Threat Response

The activities that come to mind for most individuals when considering the SOC are as follows: Once an incident is verified, the SOC functions as an initial responder, carrying out tasks such as closing or segregating endpoints, ending harmful processes (or preventing them from running), removing files, and more. The objective is to react to the required degree while minimizing the impact on business continuity as much as feasible.

f. Recovery and Remediation

After an incident, the Security Operations Center (SOC) will work to restore systems and recover any lost or compromised data. This may involve wiping and restarting endpoints, reconfiguring systems, or deploying viable backups to counter ransomware attacks. This will return the network to its pre-incident state.

g. Log Management

The SOC collects, maintains, and regularly reviews the organization's network activity and communications logs. This information assists in establishing a standard for "normal" network activity, can uncover potential threats, and can be utilized for resolving issues and conducting investigations following a security incident. Many Security Operations Centers (SOCs) utilize a Security Information and Event Management (SIEM) system to combine and analyze data from various sources such as applications, firewalls, operating systems, and endpoints, all of which generate internal logs.

i. Root Cause Investigation

After an incident, the Security Operations Center (SOC) is responsible for determining precisely what occurred, when it happened, how it happened, and why. To carry out this investigation, the SOC utilizes log data and other information to trace the issue back to its origin and prevent similar problems in the future. These analyses are part of the *cyber intelligence* process.

k. Security Refinement and Improvement

Cybercriminals are constantly improving their tools and tactics. The Security Operations Center (SOC) *must implement improvements to stay ahead of them continuously* through the implementation of *cyber intelligence*. This involves putting the plans outlined in the Security Road Map into action. The refinement process can also include hands-on practices such as red-teaming and purple-teaming.

j. Compliance Management

Many SOC processes are guided by established best practices, but compliance requirements govern some. The SOC must regularly audit their systems to ensure compliance with regulations issued by their organization, industry, or governing bodies, such as GDPR,

HIPAA, and PCI DSS. Adhering to these regulations helps protect the sensitive data entrusted to the company, safeguarding against reputational damage and legal challenges resulting from a breach.

### *Optimizing a security operations model*

While dealing with incidents occupies much of the SOC's resources, the chief information security officer (CISO) is responsible for the broader scope of risk and compliance. To connect operational and data depository across these functions, an effective strategy requires an adaptable security architecture that allows organizations to implement optimized security operations. This approach improves efficiency through integration, automation, and orchestration, reducing the number of labour hours required while enhancing the organization's information security management posture.

An optimized security operations model involves adopting a security framework that integrates *security solutions and threat intelligence* into daily processes. SOC tools, such as centralized and actionable dashboards, help consolidate threat data into security monitoring dashboards and reports. This keeps operations and Management informed about evolving events and activities. SOC teams can effectively oversee the overall risk posture by integrating threat management with other risk and compliance management systems. These configurations offer continuous visibility across systems and domains and can leverage actionable intelligence to enhance the accuracy and reliability of security operations. Centralized functions lessen the workload of manual data sharing, auditing, and reporting. In this process, *cyber intelligence is an essential part of threat management*.

A comprehensive assessment is essential to operationalizing threat management. In addition to evaluating defences, an organization should assess its processes and policies. What are the organization's strengths and weaknesses? What is its risk posture? What data is being collected, and how much is being utilized?

While each organization is unique, specific core capabilities and best practices for security operations represent prudent care today. *A reasonable threat management process begins with a plan*. The role of *cyber intelligence* involves discovering anomalies through baseline calculations, normalization, and correlation. Then, the anomalies are triaged based on risk and asset value, followed by analysis and scoping. Threat management processes prioritize and characterize cases for incident response programs. A well-defined response plan is crucial for containing a threat or minimizing damage from a data breach."

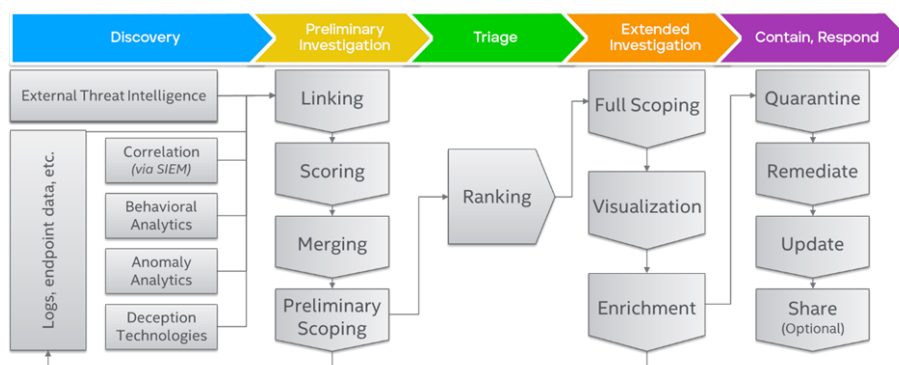


Fig. 2 Threat management plans integrate and structure many security and IT operations processes. (Source: <https://www.trellix.com/en-us/security-awareness/operations/what-is-soc.html>, [5])

Adequate visibility and threat management will draw on many data sources, but sorting out valuable and timely information can take time and effort. The most valuable data is event data generated by countermeasures, IT assets, and indicators of compromise (IoCs) produced internally through malware analysis and externally via *threat intelligence feeds*. Additionally, system data available from sensors (e.g., host, network, database, etc.) is crucial.

These data sources are not just inputs to threat management; they provide context and make the information valuable and actionable for more precise, accurate, and speedy assessment throughout the iterative and interactive threat management process. Accessing and effectively using the correct data to support plans and procedures is a measure of organizational maturity. In a "mature" scenario, a workflow would involve handing off the proper information or permitting direct action within operational consoles and across products. This workflow integrates IT operations security, security teams, and tools into incident response when a critical event occurs.

All these assessments will help *prioritize where increased investment or friction reduction is needed to align threat management implementation with goals*. Experts and penetration tests can help benchmark strategy and organizational maturity and evaluate security responses against attacks to increase an organization's ability to detect and contain malicious events. By comparing against peer enterprises, this vetted review can help justify and explain the need to redirect or invest in cybersecurity operations resources.

## 6. Discussion

*Cyber intelligence, commonly referred to as Cyber Intel, is a crucial component in augmenting the capabilities of a Security Operations Center (SOC).* Its applications are diverse and critical for enhancing security measures in an organization.

Firstly, Cyber Intel aids in threat identification and prioritization by utilizing intelligence feeds to discern and prioritize emerging threats. These feeds provide valuable data on known threats, indicators of compromise (IOCs), and tactics, techniques, and procedures (TTPs) utilized by threat actors.

This proactive approach allows analysts to hunt for threats within the network based on current threat landscapes and potential attack vectors.

Regarding incident response, Cyber Intelligence contributes to enhanced detection by enabling SOC's to integrate it into their capabilities. This integration allows for a comparison of anomalies and potential threats against known threat signatures and behaviours, leading to quicker threat identification. During an incident, Cyber Intelligence provides crucial context about the threat actor's motives, methods, and history, enabling a more effective response.

Cyber Intelligence plays a pivotal role in vulnerability management, as it assists in identifying and assessing an organization's systems and software vulnerabilities. Furthermore, intelligence on actively exploited vulnerabilities helps SOC's prioritize patching and remediation efforts.

Moreover, in strategic decision-making, Cyber Intel facilitates risk assessment, aiding SOC's in allocating resources by evaluating the risk level of various threats. Insights from Cyber Intel inform the development of security policies and procedures, ensuring they are aligned with current threat landscapes.

In security awareness and training, Cyber Intel is integral for designing and implementing effective security awareness programs for employees, utilizing intelligence on social engineering tactics and phishing campaigns. Moreover, it supports continuous training for SOC analysts on the latest threats and mitigation strategies.

Additionally, Cyber Intel can be integrated with Security Information and Event Management (SIEM) and Security Orchestration, Automation, and Response (SOAR)

platforms to automate threat detection and response workflows. This integration enables automated systems to use Cyber Intel to block or mitigate threats in real-time, thereby reducing the burden on SOC analysts.

Furthermore, internal and external collaboration is facilitated by sharing Cyber Intel within the organization and participating in information sharing and analysis centres and other industry groups. These collaborative efforts enhance insights into threats and enable collaboration on defence strategies.

In terms of compliance and reporting, Cyber Intel helps SOC teams ensure compliance with industry regulations and standards by providing insights into required security measures and threat landscapes. Additionally, it supports the creation of detailed reports for stakeholders, demonstrating the current security posture and ongoing efforts to mitigate risks.

Moreover, Cyber Intel provides insights into future threat trends, aiding SOC teams in long-term strategy and Planning. Additionally, it supports the justification of investments in new technologies, training, and resources to improve the organization's security posture.

In summary, integrating Cyber Intel into a SOC enhances its ability to detect, respond to, and mitigate threats more effectively, ultimately leading to a more robust and proactive cybersecurity posture.

## 7. Conclusions

Security Operations Center (SOC) teams must always stay one step ahead of attackers. In recent years, this has become increasingly difficult. The capabilities of Cyber Intel support this. The top three challenges faced by every Security Operations Center (SOC) team are:

1. Shortage of cybersecurity skills: A survey by Dimensional Research found that 53% of SOC teams need help hiring skilled personnel. This has resulted in many SOC teams needing more staff and the advanced skills to promptly and effectively identify and respond to threats. According to the (ISC) Workforce Study, the cybersecurity workforce needs to grow by 145% to close the skills gap and better defend organizations worldwide [6].

2. Too many alerts: As organizations implement new tools for threat detection, the volume of security alerts continues to grow. This has overwhelmed security teams and led to alert fatigue. *Additionally, many alerts must provide more intelligence or context for investigation, leading to false positives.* False positives drain time and resources and divert attention from actual incidents.

3. Operational overhead: Many organizations use disconnected security tools, requiring security personnel to translate security alerts and policies between environments. This results in costly, complex, and inefficient security operations.

For many SOC teams, detecting malicious activity within the network is like finding a needle in a haystack. They are often required to gather information from multiple monitoring solutions and navigate thousands of daily alerts. As a result, critical attacks are only noticed once it is too late. These processes can be optimized based on the effective use of cyber intelligence.

## Acknowledgement

The scientific publication is in fulfillment of the objectives of the National Scientific Program "Security and Defense", financed by the Ministry of Education and Science of the Republic of Bulgaria in implementation of the Decision of the Council of Ministers No. 731 of October 21, 2021 and according to Agreement No. D01-74 /19.05.2022.

## References

1. “What is the Difference Between Threat Intelligence and Cybersecurity?” Accessed: Apr. 21, 2023. [Online]. Available: <https://www.netscout.com/what-is/difference-between-threat-intelligence-and-cybersecurity>
2. “Cybersecurity vs. Cyber Intelligence,” Netsweeper. Accessed: Apr. 21, 2023. [Online]. Available: <https://www.netsweeper.com/filter/industry/cybersecurity-vs-cyber-intelligence/36753>
3. Campana N., “What Does A Cyber Threat Intelligence Analyst Do? | IT Security | Jobs,” Freelancer Blog. Accessed: Apr. 21, 2023. [Online]. Available: <https://www.freelancemap.com/blog/what-does-cyber-threat-intelligence-analyst-do/>
4. N. Campana, “What Does A Cyber Threat Intelligence Analyst Do? | IT Security | Jobs,” Freelancer Blog. Accessed: Jul. 25, 2024. [Online]. Available: <https://www.freelancemap.com/blog/what-does-cyber-threat-intelligence-analyst-do/>
5. “What is SOC (Security Operation Center)?,” Check Point Software. Accessed: Apr. 21, 2023. [Online]. Available: <https://www.checkpoint.com/cyber-hub/threat-prevention/what-is-soc/>
6. “Security Operations Center (SOC) Tools | AT&T Cybersecurity.” Accessed: Apr. 21, 2023. [Online]. Available: <https://cybersecurity.att.com/solutions/security-operations-center/building-a-soc/soc-tools>

# GAP Analyses in Cybersecurity

**Nedko Tagarev**

University of National and World Economy, Sofia, Bulgaria  
*ntagarev@unwe.org*

**Abstract.** This article, part of a series on cybersecurity management, addresses the challenge of managing Cybersecurity as a process and a product. It explores the connections between the current situation, standards, and policies for self-control in the cybersecurity work environment while also highlighting the concept of GAP analyses for managing the main threats and defence mechanisms. The article provides practical example that demonstrate the real-world application of these concepts in different organizations, thereby assuring readers of the practicality of the solutions presented. GAP analysis is a commonly employed methodology across diverse fields including business, education, and project management. It plays a crucial role in assessing the disparity between the present state of affairs and the intended objectives. The acronym "GAP" encompasses "Goal, Achievement, and Performance," representing the fundamental elements involved in the analysis process.

**Keywords:** GAP analyses, Cybersecurity

## 1. Introduction

A major **problem** in cybersecurity management is that when a new threat or exploit method emerges, new tools are often deployed, also requiring the addition of new personnel. This, in turn, increases costs and complicates management processes. This imposition of new tools is often done without an assessment of available resources, assets, and processes and without considering the current and desired state of the security system. A managerial solution to such problems can be GAP analysis.

GAP analysis, a versatile tool widely used in various fields such as business, education, and project management, can also be effectively applied in cybersecurity. It aids in evaluating the difference between the current state of business and the desired goals, with 'GAP' representing 'Goal, Achievement, and Performance.' This process is adaptable, allowing for the deployment of protection against various cybersecurity approaches. Figure 1 outlines the main steps in this type of analysis, instilling confidence in its effectiveness.

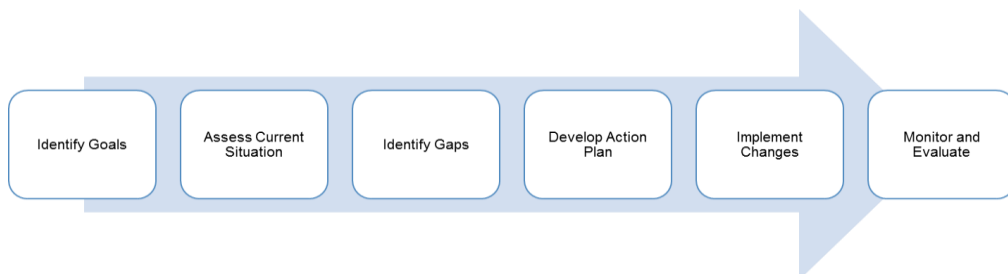


Fig. 1. The main steps in GAP analyses

The first step, **Identify Goals**, involves setting specific and measurable goals or targets that need to be achieved. These goals could be related to performance, efficiency, quality, or any other relevant aspect of the organization or project. Cybersecurity can be managed as a project. So, the following questions have to be answered: *How much? How many? Is it quantifiable? How will I know when it is accomplished?* [1].

The next step, **Assessing the Current Situation**, involves evaluating the current situation or performance level compared to the set goals. This process involves gathering data and conducting surveys and interviews to collect information about the current state. This includes researching the security products, the customers it serves, the geographical locations it reaches, and the benefits it offers its employees. The information gathered can be quantitative (such as financial records) or qualitative (such as surveys or feedback from key stakeholders) [2].

After analyzing the data, we have to **Identify the Gaps**. The gaps are discrepancies between the current state and the desired goals are identified. These gaps could be related to skills, resources, processes, performance metrics, or other relevant factors. In order to address the issue, it is crucial to define the gap and its specific characteristics clearly. Additionally, it is vital to examine the reasons behind the occurrence of the gap thoroughly. Asking targeted, business-specific questions and providing honest answers will help [3].

Once the gaps are identified, the next step is to **create an action plan** to bridge these gaps. This could involve implementing new strategies, acquiring additional resources, providing training, or making other necessary changes to improve performance and achieve the desired goals. A practical action plan should outline the specific actions, responsibilities, resources, timelines, and expected outcomes for each gap or opportunity. It is essential to prioritize the actions based on their impact, urgency, and feasibility and ensure they align with the organization's goals and objectives. Additionally, the actions should be SMART - specific, measurable, achievable, relevant, and time-bound [4].

With the action plan in place, the **necessary changes are implemented** to address the identified gaps and move closer to the desired goals. This is parallel to the next step, **Monitoring and Evaluation** is essential. Continuously monitor progress and evaluate the effectiveness of the change throughout the implementation process will provide data for good or bad practices. This continuous monitoring and evaluation ensure the sustainability of the process and allow for adjustments to be made as needed to ensure that the goals are met efficiently. This involves:

1. Execution of Plans: Carrying out the action plans developed to close the gaps.
2. Performance Tracking: Regularly measuring progress against predefined metrics to evaluate how well the organization moves towards its desired state.
3. Be ready to change your strategies based on feedback and performance data. Implementing effective and continuous monitoring is crucial to close the gap and ensure that the organization is progressing in the right direction [5].

GAP analysis is a valuable tool for identifying areas for improvement and guiding decision-making processes to help organizations or individuals reach their objectives effectively. It provides a structured approach for assessing the current situation, setting goals, and developing actionable strategies to bridge the gaps between the two.

## 2. Basic instructions for the creation of a GAP chart

A GAP analysis Chart offers a visual framework that illustrates the differences between actual and desired performance. All types of organizations can benefit from using a GAP analysis template as it helps to visualize performance gaps, which can then be shared with

employees and teams. A well-designed gap analysis chart(template) clearly outlines problem areas where resources are not being used or allocated effectively [6].

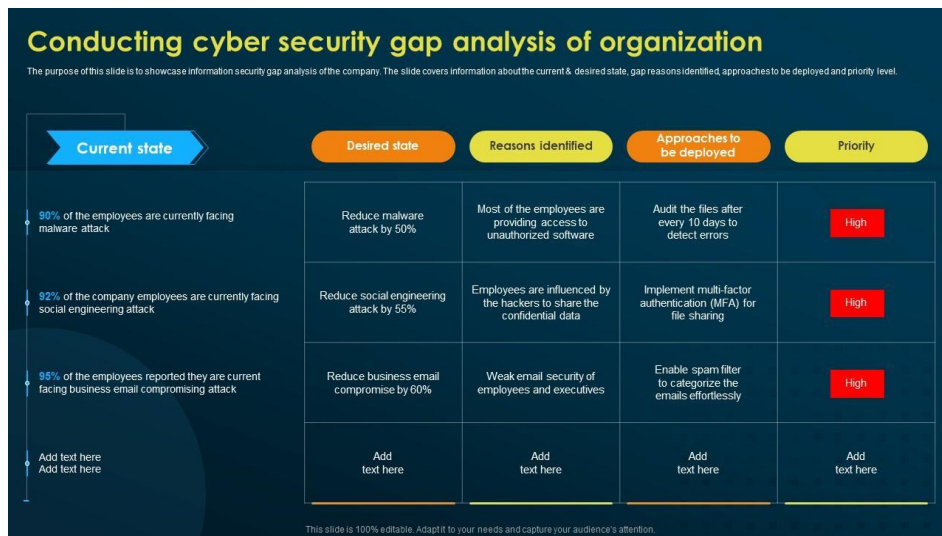


Fig. 2. Example Chart (Source: <https://www.slideteam.net/conducting-cyber-security-gap-analysis-of-organization-implementing-security-awareness-training.html> [7])

The chart contains rows and columns for specific goals or objectives, the current state (As Is) and the desired state (To Be). Under each state column, you would list the relevant aspects or factors that contribute to achieving the goals. By comparing the current state with the desired state, you can identify the gaps that need to be addressed to move from the current situation to the desired situation. It is important to note that this is a simplified representation, and a GAP analysis chart may contain additional details such as specific metrics, assessment criteria, action items, timelines, responsible parties, and more, depending on the complexity and context of the analysis.

Key benefits of using a gap analysis template include:

1. *Enhanced Customer Satisfaction:* Gap analysis helps organizations pinpoint performance gaps, allowing them to improve customer experience and satisfaction.
2. *Competitive Advantage:* Organizations can gain a competitive edge by identifying and addressing gaps.
3. *Improved Comprehension:* A gap analysis template visually represents the disparities between actual and desired performance, making it easier to understand problems and solutions.
4. *Contribution to Employee Benefit Programs:* Gap analysis is commonly utilized to develop employee benefit programs.

### 3. GAP analyses for Cybersecurity

When conducting a GAP analysis for Cybersecurity, it is crucial to confidently compare your organization's current cybersecurity measures with the desired or recommended measures



to ensure adequate protection against cyber threats. To structure your GAP analysis chart for Cybersecurity, refer to the following structure in Table 1.

Table 1. GAP analyses for cybersecurity basic chart

Goals	Current State (As Is)	Desired State (To Be)
Security Policies and Procedures	Policies exist but need to be fully enforced or updated regularly, and procedures need to be more consistent and complete.	Comprehensive policies covering all aspects of Cybersecurity. Regular updates and enforcement.
Network Security	Basic firewall and antivirus software in place. Some network segmentation. We have limited monitoring and logging.	Network has advanced firewall and intrusion detection systems, robust network segmentation and traffic monitoring.
Employee Training and Awareness	Basic cybersecurity awareness training is in place.	Regular training on Cybersecurity best practices. They are phishing awareness programs.
Incident Response	Ad-hoc incident response process. There needs to be a clear communication plan or escalation procedures.	A formal incident response plan is in place—regular drills and simulations.
Data Protection	We have limited data encryption. Data retention policies need to be more well-defined.	Encryption is used for all sensitive data. Strong data retention policies.
Vendor and Third-Party Risk Management	Limited assessment of vendor and third-party security risks.	Comprehensive assessment of vendor and third-party security risks.

- Each row represents a different aspect of Cybersecurity.
- Under the "Current State (As Is)" column, you would assess the existing measures or practices related to each aspect.
- Under the "Desired State (To Be)" column, you would outline the ideal or recommended practices for each aspect.
- The comparison helps identify the gaps between the current and desired states, guiding the development of an action plan to improve cybersecurity posture.

#### 4. Example with GAP analyses and access control

Access control is critical to Cybersecurity, which focuses on determining who can access and use resources in a computing environment. It ensures that only authorized users can access specific resources, thus protecting data from unauthorized access and potential breaches. Here is a detailed exploration of access control in Cybersecurity.

A few essential methods are commonly used to identify users. One way is through user IDs and unique identifiers like usernames or account numbers. Another method is through biometric identifiers such as fingerprints, retinal scans, or facial recognition.

Different approaches can be taken for authentication purposes. The most commonly used method is passwords, where users provide a secret word or phrase. Another method is Multi-Factor Authentication (MFA), which uses multiple methods like passwords, one-time passcodes (OTPs), and biometrics to verify the user's identity. Biometric verification, such as using fingerprints, facial recognition, or retinal scans, is also used for authentication.

In terms of authorization, there are several ways to control access to resources. Role-Based Access Control (RBAC) grants access based on the user's role within an organization. Attribute-Based Access Control (ABAC) allows access based on specific attributes, like department or clearance level. Discretionary Access Control (DAC) gives the resource owner the authority to decide who has access. At the same time, Mandatory Access Control (MAC) ensures that access levels are strictly defined by the organization and cannot be changed by users.

Finally, accountability is an essential aspect of user management. Logging and auditing are crucial in tracking who accesses what resources and when ensuring that actions can be traced back to specific users.

Access controls can be categorized into two main types:

Physical Access Control involves using locks, biometric scanners, and security guards to restrict physical access to computing resources.

On the other hand, Logical Access Control employs methods such as passwords, encryption, firewalls, and network security protocols to protect access to digital resources. These measures safeguard sensitive information.

There are several options regarding methods and technologies for ensuring security. Strong, regularly updated passwords can help reduce the risk of unauthorized access. Additionally, biometric security measures, such as fingerprint scanners and facial recognition, offer enhanced security by uniquely identifying individuals based on their biological traits.

Another approach is using smart cards and tokens, which are physical devices that provide access credentials and can add an extra layer of security to the access control process. Access Control Lists (ACLs) are also important, as they define which users or system processes can access objects and what operations they can perform, thereby regulating access to sensitive information.

Lastly, employing Public Key Infrastructure (PKI), which uses cryptographic keys and digital certificates, can provide secure access control by ensuring that only authorized users can access certain information. These methods and technologies all play a crucial role in maintaining the security and integrity of systems and sensitive data.

Adhering to best practices for access control and security is essential. The Principle of Least Privilege dictates that users should only be granted the minimum access necessary for their job functions. Additionally, conducting periodic reviews of access rights is crucial to ensure they are aligned with current job responsibilities. Strengthening security through implementing multi-factor authentication (MFA) whenever possible is highly recommended. Educating users about secure access practices and potential risks is also paramount.

Furthermore, having well-defined procedures for responding to unauthorized access attempts or breaches is crucial for effective incident response. Achieving the right balance between security and usability is crucial. More relaxed access controls could significantly hamper productivity. Managing access in dynamic environments and continuously updating access controls to reflect changing roles and responsibilities within an organization also pose significant challenges. Lastly, guarding against insider threats and preventing employees with legitimate access from misusing their privileges are ongoing concerns that require attention.

Table 2. Example of GAP analyses based on access control

Goals	Current State (As Is)	Desired State (To Be)
User Access Management	User access is managed inconsistently across systems—manual user provisioning and de-provisioning processes.	We have a centralized user access management system with role-based access control—automated provisioning and de-provisioning.
Authentication Methods	Basic password-based authentication is in place, but MFA is not used.	Multi-factor authentication (MFA) is implemented for all sensitive systems and applications.
Authorization Policies	Access rights are assigned based on job roles but not consistently reviewed, and there is limited role-based access control.	Granular access controls are based on the least privilege principle—regular access rights review and updates.
Access Monitoring	Limited monitoring of user access and activity. Basic logging is in place.	Continuous monitoring of user access and activity with real-time alerts for suspicious behaviour.
Privileged Access Management	Shared administrative accounts and passwords among IT staff—lack of accountability for privileged actions.	Segregation of duties for administrative tasks. Individual accounts with elevated privileges. Regular rotation of privileged account credentials.

- Each row represents a different aspect of access control.
- Under the "Current State (As Is)" column, you would assess the existing measures or practices related to each aspect.
- Under the "Desired State (To Be)" column, you would outline the ideal or recommended practices for each aspect.
- The comparison helps identify the gaps between the current and desired states, guiding the development of an action plan to improve access control measures.

## 5. Discussion

The benefits of conducting a cybersecurity gap analysis are numerous. First and foremost, it allows organizations to enhance their security by identifying and addressing vulnerabilities, thereby reducing the risk of cyber threats. Additionally, it helps ensure regulatory compliance by empowering organizations to fulfil legal and regulatory requirements effectively, ultimately helping them avoid penalties and fines. Furthermore, a cybersecurity gap analysis aids in risk management by allowing organizations to prioritize security efforts based on a comprehensive risk assessment, thus focusing on critical areas. This process also facilitates resource optimization by enabling organizations to allocate resources effectively, ensuring the efficient use of time and budget. Finally, conducting a cybersecurity gap analysis provides management with actionable insights, ultimately guiding informed decision-making regarding security.

## 6. Conclusions

A significant problem in cybersecurity management is deploying new tools with consideration of available resources and processes, leading to increased costs. GAP analysis can be used to evaluate and bridge the gaps between the current state and desired goals in Cybersecurity.

A GAP chart illustrates the differences between actual and desired performance in organizations. It includes specific goals, current state (As Is), and desired state (To Be). Gap analysis helps identify where resources could be used more effectively.

***When conducting a GAP analysis for Cybersecurity, it is crucial to compare your organization's current cybersecurity measures with the desired or recommended measures for adequate protection against cyber threats. The analysis includes security policies, network security, employee training, incident response, data protection, and vendor and third-party risk management. This helps identify gaps and guide the development of an action plan to improve cybersecurity posture.***

## Acknowledgement

The scientific publication is in fulfillment of the objectives of the National Scientific Program "Security and Defense", financed by the Ministry of Education and Science of the Republic of Bulgaria in implementation of the Decision of the Council of Ministers No. 731 of October 21, 2021 and according to Agreement No. D01-74 /19.05.2022.

## References

1. "Gap Analysis, Goal-Setting, and How to Succeed." Accessed: Jul. 23, 2024. [Online]. Available: <https://www.linkedin.com/pulse/gap-analysis-goal-setting-how-succeed-arootahcoach>
2. "What Is a Gap Analysis?," Investopedia. Accessed: Jul. 23, 2024. [Online]. Available: <https://www.investopedia.com/terms/g/gap-analysis.asp>
3. J. Westland, "How to Conduct a Gap Analysis: Definition, Steps & Example," ProjectManager. Accessed: Jul. 23, 2024. [Online]. Available: <https://www.projectmanager.com/blog/gap-analysis-project-management>
4. "How do you integrate quality gap analysis and action plan with your strategic goals and objectives?" Accessed: Jul. 23, 2024. [Online]. Available: <https://www.linkedin.com/advice/1/how-do-you-integrate-quality-gap-analysis-action>
5. Abbas T., "How to Do Gap Analysis in Change Management? - Explained," CMI. Accessed: Jul. 23, 2024. [Online]. Available: <https://changemanagementinsight.com/gap-analysis-in-change-management/>
6. "Gap Analysis Template - Voxco." Accessed: Jul. 23, 2024. [Online]. Available: <https://www.voxco.com/blog/gap-analysis-template/>
7. "Conducting Cyber Security Gap Analysis Of Organization Implementing Security Awareness Training." Accessed: Jul. 23, 2024. [Online]. Available: <https://www.slideteam.net/conducting-cyber-security-gap-analysis-of-organization-implementing-security-awareness-training.html>

# Application of Digital Tools in the Educational Process

**Katia Kirilova**

University of National and World Economy, Sofia, Bulgaria  
*katia.kirilova@unwe.bg*

**Abstract.** The present study is devoted to the problems related to the selection and implementation of digital educational platforms. The process is complex and difficult and very much depends on benchmarking. In order for this analysis to be effective, it is necessary to define the requirements for the software tools in advance. In conclusion, conclusions from the empirical analysis are formulated.

**Keywords:** Digital tools, Software platforms, Educational process.

## 1. Theoretical background

The Bulgarian education system functions in compliance with all national and European requirements, strategic and conceptual documents. In all of them, the issues of the quality of education, the provided educational services and the digitization of processes are widely advocated [1]. In order to be able to achieve the necessary level of digitization of the educational sphere, significant financial resources are needed, as well as highly qualified experts to study, compare, select and implement different technologies in different educational institutions. The factors that largely determine the quality of education offered by schools and institutions can be grouped into the following groups (Fig. 1).

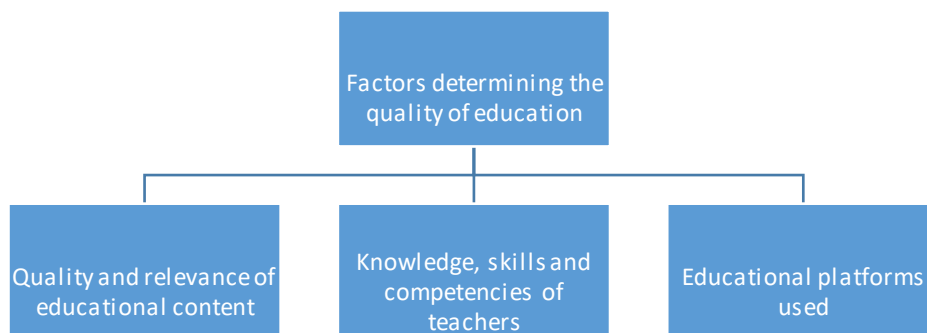


Fig. 1. Factors determining the quality of education

Many factors determine the quality of educational services offered. Among them is the educational content and its relevance. To maintain relevance, it is necessary to periodically update the content, as well as to reflect in it various innovations in individual fields, results of new research and scientific projects. A very important component is the knowledge, skills and competences of the teachers. They should also be periodically updated, through various courses and additional qualifications. Last but not least are the functionalities of the educational platforms used, which are analyzed in the present study.

## 2. Research methodology

The learning process is difficult and lengthy. In some cases, this is a long investment, the results of which become visible after some time. All this necessitates and requires that modern software solutions be used to the maximum extent in this long process. In fig. 2 presents the process of implementing similar solutions in the educational process.

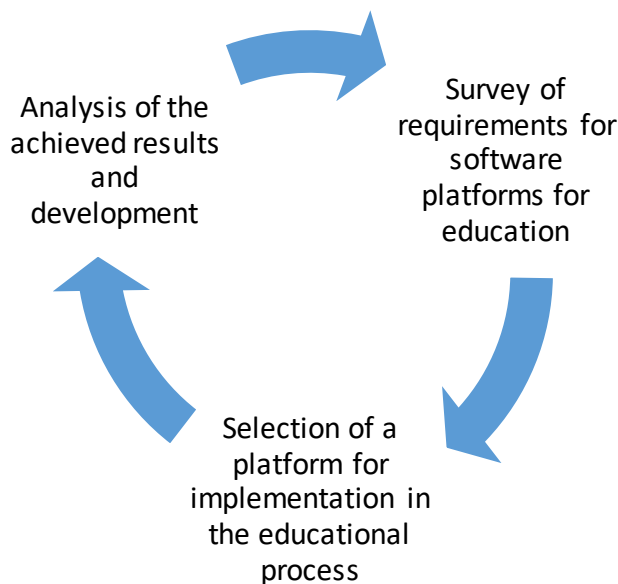


Fig. 2. Phases and stages of research

The main stages are:

- Survey of requirements for software platforms for education;
- Selection of a platform for implementation in the educational process;
- Analysis of the achieved results and development.

The first stage is related to determining the requirements for educational platforms. These requirements must be in accordance with the peculiarities of the educational process on the one hand and with the digital infrastructure in the educational organization on the other. If software solutions are purchased that do not meet the parameters of the educational infrastructure, they will not be able to be implemented and the final result of the implementation will be lost. In the second phase, it is necessary to define the requirements for educational platforms and to perform a comparative analysis. Based on this analysis, a specific platform can be selected for implementation in the educational institution. The third stage is related to the implementation of the chosen platform, its monitoring and maintenance. It is important to note that the implementation of the chosen platform will also take time, even if it uses a cloud infrastructure. The observation is related to the future development of the functionality of the platform.

### 3. Empirical study

As stated above, the implementation of software platforms in education is a complex task. Achieving high efficiency in its implementation largely depends on the strategic choice of functionalities and software solution. For these reasons, a very careful and in-depth comparative analysis of possible deployment platforms based on pre-defined selection criteria is required. Different manufacturers implement different functionalities of the offered platforms. In view of the objectives of the present study, the Capterra platform [2] was used, which offers a wide range of software products with their characteristics. In fig. 3 shows a basic diagram of the software tool comparison platform.

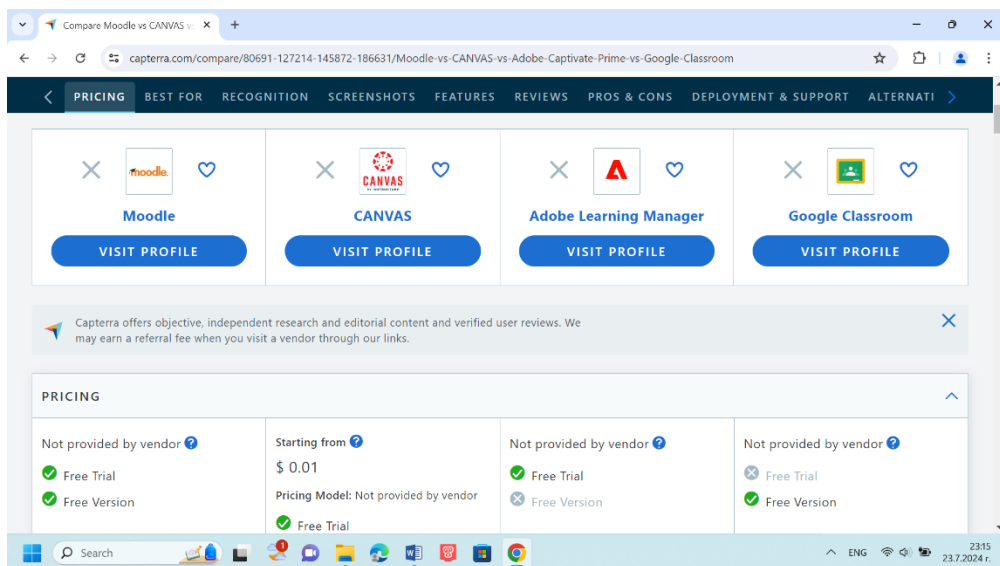


Fig. 3. Capterra software comparison platform

For the purposes of this study, the software tools Moodle, Canvas, Adobe Learning Manager, Google Classroom were compared. Capterra offers a description of the analyzed platforms by a total of 46 main characteristics, among which: Academic/Education, Access Controls/Permissions, Activity Dashboard, API, Assessment Management, Asynchronous Learning, Blended Learning, Built-in Course Authoring, Certification & Licensing, Collaboration Tools, Compliance Tracking, Content Library, Content Management and other.

These features may not be applicable to all software platforms examined. This is a normal process because all software is designed, developed and implemented with specific business processes in mind. Although all analyzed platforms are from the field of education, their specificity is different and their users have specific informational needs. For these reasons, it is difficult to establish a match on all analyzed criteria for the studied software applications. It can be assumed that this software, which has the largest set of analyzed characteristics, can cover the requirements to the greatest extent (Fig. 4).

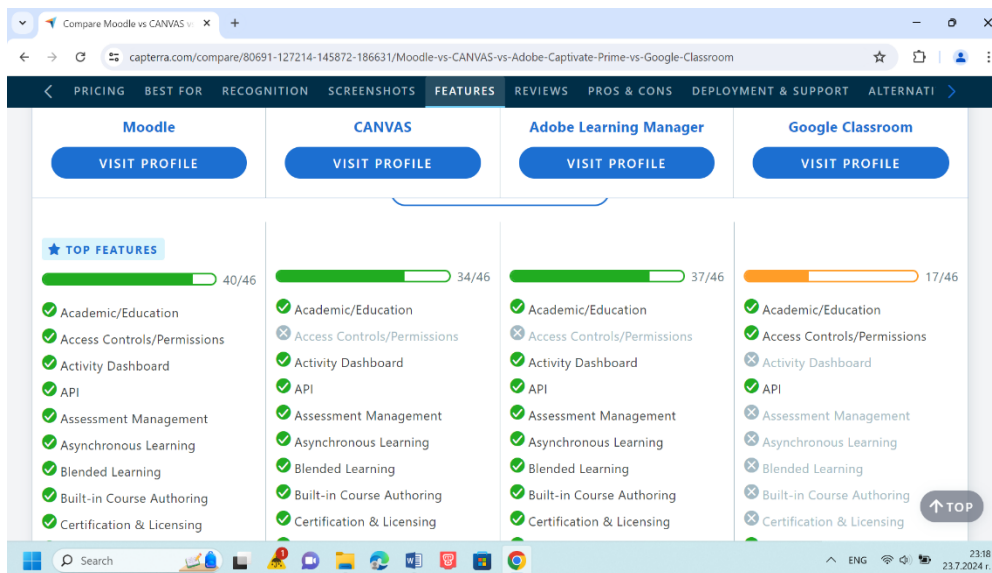


Fig. 4 Comparison on the main features

The results of the benchmarking show that Moodle covers about 40 out of 46 characteristics. Canvas offers 34 out of 46 functionalities. Google Classroom offers only 17 of the 46 features described in the platform, and Adobe Learning Manager 37 features. The explanation for these results is that Google Classroom is inherently a different kind of learning platform, offering a different type of functionality, and its qualities are in no way belittled.

## 4. Conclusions

As a result of the development of the present study, the following more significant conclusions can be formulated:

- It is very difficult to find educational platforms that offer all the necessary functionalities. For this reason, the choice of an implementation platform is always based on a choice of different strategic alternatives with some compromise in some of the functionalities of one or another software;
- The quality of the educational process and the achievement of high results largely depend on the choice of platform.

## References

1. <https://www.mon.bg/mon/strategii-i-politiki/strategicheski-dokumenti-2/>
2. <https://www.capterra.com/learning-management-system-software/>



# The Role of Data Management and Open Science in Supporting Digital Technologies at Universities

**Plamen Milev**

University of National and World Economy, Sofia, Bulgaria

*pmilev@unwe.bg*

**Abstract.** This paper explores the roles of data management and open science in enhancing digital technologies at universities. It examines how these practices support the integration of advanced tools like big data analytics, machine learning, and immersive technologies such as virtual reality (VR) and augmented reality (AR). Effective data management ensures data accuracy, security, and accessibility, facilitating robust research and efficient operations. Open science promotes transparency, collaboration, and the democratization of knowledge, accelerating innovation and enhancing research reproducibility. The paper reviews current practices, challenges, and benefits, highlighting the synergies and conflicts between data management and open science. Despite challenges such as funding constraints, cultural resistance, technical barriers, and data privacy concerns, adopting robust data management and open science practices is essential for leveraging digital technologies in higher education. Recommendations for universities include investing in infrastructure, fostering a culture of openness, developing clear policies, and promoting collaboration. These strategies are crucial for advancing research capabilities and educational outcomes in the digital age. This work underscores the importance of continuous adaptation to keep pace with technological advancements and ensure effective support for digital transformation at universities.

**Keywords:** Data Management, Open Science, Digital Technologies, Virtual Reality (VR), Augmented Reality (AR).

## 1. Introduction

In an era where digital technologies are rapidly transforming the landscape of higher education, universities are increasingly reliant on robust data management practices and the principles of open science to enhance their research capabilities and operational efficiencies. Data management, which encompasses the collection, storage, and utilization of data, is pivotal in ensuring that data is accessible, reliable, and secure. Concurrently, the open science movement advocates for transparency, accessibility, and collaboration in scientific research, promoting the sharing of research outputs and methodologies. The convergence of data management and open science presents a unique opportunity for universities to not only streamline their research processes but also to foster innovation using emerging digital technologies. This paper explores the intricate relationship between data management and open science and examines how their integration can support and enhance digital technologies at universities, with a particular focus on emerging immersive technologies such as virtual reality (VR) and augmented reality (AR).

The aim of this paper is to support the thesis that by effectively implementing data management and open science practices, universities can significantly enhance their support for digital technologies. This improvement will advance research capabilities, foster

collaboration, and enrich the educational experience through the application of emerging immersive technologies such as virtual reality (VR) and augmented reality (AR).

## **2. Data management and open science in universities**

Data management refers to the systematic process of collecting, storing, organizing, and maintaining data to ensure its accuracy, accessibility, and reliability. In the context of universities, data management includes data governance, quality control, security, storage, and sharing [1-4]. Effective data management ensures that data is reliable and easily accessible, which is crucial for supporting research and administrative functions in universities. Currently, universities face several challenges in data management, including fragmented systems, resource constraints, compliance issues, and cultural barriers. Different departments often use disparate systems, leading to data silos that hinder collaboration and data sharing [5]. Additionally, universities may lack the funding and expertise needed to implement advanced data management solutions [6]. Navigating complex regulatory requirements related to data privacy and security further complicates data management efforts [7]. Cultural resistance to adopting new practices can also impede progress [6]. Despite these challenges, effective data management offers numerous benefits. Enhanced research capabilities arise from improved access to high-quality data, enabling more robust and reproducible research [8]. Streamlined data processes can reduce redundancies and administrative burdens, thereby increasing operational efficiency [9]. Reliable data also supports informed decision-making at all levels of the institution [10], and proper data management ensures adherence to regulatory standards, mitigating risks [3].

Open science is a movement that advocates making scientific research, data, and dissemination accessible to all levels of society, emphasizing transparency, collaboration, and democratization of knowledge. Its core principles include open access to research publications, open data, open methodology, open peer review, and open educational resources [11-15]. In academia, open science enhances the visibility, impact, and integrity of research. It promotes collaboration among researchers across institutions and disciplines [16], increases transparency, and enhances the reproducibility and credibility of research findings [17]. Open science also democratizes access to scientific information, fostering public engagement and expanding access to knowledge [18]. Furthermore, it accelerates innovation by enabling faster dissemination and application of research discoveries [19]. However, challenges in implementing open science practices in universities include limited funding for open access publishing, insufficient incentives for researchers to share data, intellectual property concerns, and cultural resistance [4, 11, 20, 21]. Developing and maintaining the infrastructure to support open science initiatives also poses technical challenges [4].

Effective data management is foundational to the success of open science (Fig. 1). Well-managed data is easily accessible, facilitating data sharing and reuse [4]. Ensuring data is accurate and reliable enhances the credibility of open science outputs [8]. Properly documented data, including metadata and thorough documentation, makes data easier to understand and use by other researchers [22]. Additionally, data management practices help researchers comply with funder and publisher open access requirements [11]. While data management and open science share common goals, there can be tensions. Both promote transparency, collaboration, and efficient use of resources, and open data initiatives rely on robust data management practices to ensure data quality and accessibility [19]. However, conflicts may arise regarding data privacy, intellectual property, and the costs associated with open access publishing. Balancing these concerns requires careful policy and infrastructure development [20].

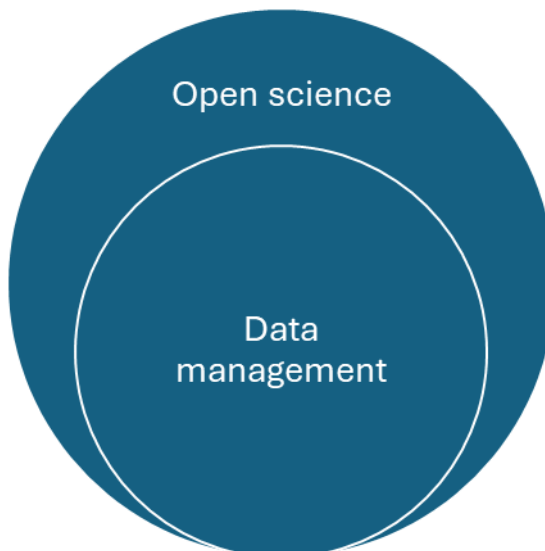


Fig. 1. Open science as a broader concept

Open science encompasses a wider range of practices and principles aimed at making all aspects of scientific research transparent and accessible. This includes not only data management but also open access to research publications, open sharing of research methods and protocols, transparent peer review processes, and the use of open educational resources.

### **3. Supporting digital technologies with data management and open science**

The integration of data management and open science practices can significantly enhance the support for digital technologies at universities. Digital technologies, such as data analytics, machine learning, and emerging immersive technologies like virtual reality (VR) and augmented reality (AR), require robust data infrastructure and transparent research methodologies to thrive. Data management plays a crucial role in supporting these technologies by ensuring that data is accurately collected, securely stored, and readily accessible for analysis and application. For example, effective data management can facilitate the use of big data analytics in educational research, helping universities understand student behavior and improve learning outcomes [9]. Similarly, secure and well-organized data storage solutions are essential for the deployment of VR and AR technologies in teaching and research [3]. Immersive technology can effectively support the process of teaching complex theories, concepts through object visualization and further used for technology education [23]. Open science further enhances the support for digital technologies by promoting the sharing of data, research methods, and findings. This openness accelerates innovation and collaboration, as researchers can build on each other's work more efficiently. For instance, open access to large datasets and advanced analytical tools enables researchers to develop more sophisticated machine learning models, which can then be applied to various academic disciplines [16]. Moreover, the principles of open science ensure that the research outputs from digital technology projects are transparent and reproducible, which is vital for maintaining scientific integrity [17].

Despite the potential benefits, universities face several challenges in integrating data management and open science practices to support digital technologies (Fig. 2). These challenges include:

- Funding and resources: implementing advanced data management systems and supporting open science initiatives require significant financial and human resources, which many universities may lack [11].
- Cultural resistance: there is often resistance to change within academic institutions, where traditional practices may be deeply ingrained. Encouraging faculty and staff to adopt new data management and open science practices can be difficult [21].
- Technical barriers: developing and maintaining the technical infrastructure needed to support data management and open science can be complex and costly. This includes the creation of data repositories, ensuring data security, and maintaining compliance with regulatory standards [4].
- Data privacy and intellectual property: balancing the openness of data with the need to protect sensitive information and intellectual property rights is a significant challenge. Universities must develop policies that address these concerns while promoting transparency and collaboration [20].

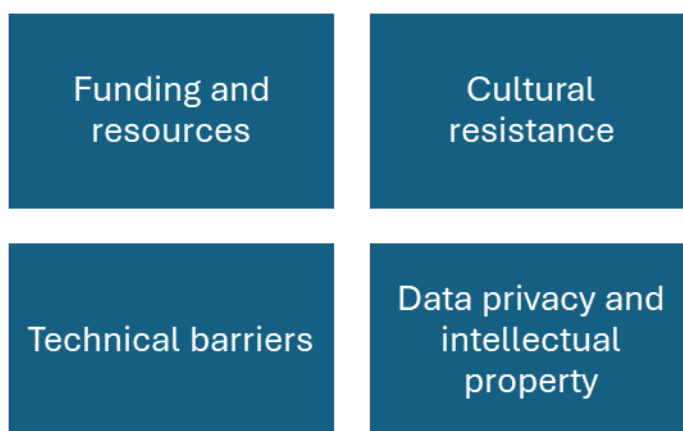


Fig. 2. Challenges in integrating data management and open science practices to support digital technologies

To address these challenges and fully realize the potential of data management and open science in supporting digital technologies, universities should consider the following strategies (Fig. 3):

- Invest in infrastructure: universities should allocate resources to develop and maintain robust data management systems and open science infrastructure. This includes investing in data repositories, security measures, and compliance tools [8].
- Promote a culture of openness: encouraging a culture that values openness and collaboration is essential. Universities can achieve this by providing training and incentives for researchers to adopt open science practices [18].
- Develop clear policies: establishing clear policies on data management, open access, and data privacy can help mitigate risks and ensure compliance with regulatory standards. These policies should be regularly reviewed and updated to keep pace with technological advancements [7].

- Foster collaboration: universities should promote interdisciplinary and inter-institutional collaboration to leverage the full potential of data management and open science. Collaborations can lead to the sharing of best practices, resources, and knowledge, which can enhance the development and application of digital technologies [16].



Fig. 3. Future directions in integrating data management and open science practices to support digital technologies

The process of researching and writing this paper has highlighted the critical role of data management and open science in supporting digital technologies at universities. Through the review of existing literature, it has become evident that these practices are not just complementary but synergistic, each reinforcing the other to create a more efficient and transparent research environment. One key learning is the importance of a holistic approach to data management and open science. This approach involves not only the implementation of technical solutions but also the cultivation of a supportive institutional culture. Universities need to provide training, resources, and incentives to encourage researchers and staff to adopt these practices. Additionally, clear policies and guidelines are essential to navigate the complexities of data privacy, intellectual property, and compliance. As digital technologies evolve, so must the strategies for managing and sharing data. Continuous learning and adaptation are necessary to keep pace with technological advancements and to ensure that universities can fully leverage the benefits of data management and open science.

#### 4. Conclusion

In conclusion, the integration of data management and open science practices is crucial for supporting digital technologies at universities. Effective data management ensures that data is accurately collected, securely stored, and readily accessible, which is essential for leveraging digital technologies like big data analytics, machine learning, and emerging immersive technologies such as virtual reality (VR) and augmented reality (AR). By implementing robust data management systems, universities can enhance research capabilities, streamline operations, and support data-driven decision-making. Open science further amplifies these benefits by promoting transparency, collaboration, and the democratization of knowledge. It ensures that research data, methods, and findings are accessible to all, accelerating innovation and fostering interdisciplinary collaboration. This openness is particularly important for the development and application of advanced digital technologies, as it enables researchers to build on each other's work and develop more sophisticated tools and models. Despite the

potential benefits, universities face several challenges in implementing these practices, including funding and resource constraints, cultural resistance, technical barriers, and concerns about data privacy and intellectual property. Addressing these challenges requires strategic investment in infrastructure, fostering a culture of openness, developing clear policies, and promoting collaboration both within and between institutions. By effectively integrating data management and open science practices, universities can not only support the development and application of digital technologies but also enhance the overall quality and impact of their research. This integration is essential for advancing knowledge, driving innovation, and improving educational outcomes in the digital age.

## Acknowledgement

The author acknowledges with gratitude funding provided by the Bulgarian National Research Fund under administrative contract KP-06-H65/5 of 12.12.2022.

## References

1. Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM*, 53(1), 148-152.
2. Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of management information systems*, 12(4), 5-33.
3. Stallings, W. (2007). *Data and computer communications*. Pearson Education India.
4. Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, 63(6), 1059-1078.
5. Kim, S. Y., Yi, H. J., & Huh, S. (2019). Current and planned adoption of data sharing policies by editors of Korean scholarly journals. *Science Editing*, 6(1), 19-24.
6. Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., ... & Frame, M. (2011). Data sharing by scientists: practices and perceptions. *PloS one*, 6(6), e21101.
7. Ienca, M., & Vayena, E. (2020). On the responsible use of digital data to tackle the COVID-19 pandemic. *Nature medicine*, 26(4), 463-464.
8. Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3(1), 1-9.
9. Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage.
10. Provost, F., & Fawcett, T. (2013). *Data Science for Business: What you need to know about data mining and data-analytic thinking*. "O'Reilly Media, Inc."
11. Piwowar, H., Priem, J., Larivière, V., Alperin, J. P., Matthias, L., Norlander, B., ... & Haustein, S. (2018). The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ*, 6, e4375.
12. Rice, R. (2019). Supporting research data management and open science in academic libraries: a data librarian's view. *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare*, 72(2), 263-273.
13. Miedema, F. (2022). Open science: The very idea (p. 247). Springer Nature.
14. Ross-Hellauer, T. (2017). What is open peer review? A systematic review. *F1000Research*, 6.
15. Bliss, T. J., & Smith, M. (2017). A brief history of open educational resources. *Open: The philosophy and practices that are revolutionizing education and science*, 9-27.
16. Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., ... & Yarkoni, T. (2015). Promoting an open research culture. *Science*, 348(6242), 1422-1425.
17. Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., Percie du Sert, N., ... & Ioannidis, J. (2017). A manifesto for reproducible science. *Nature human behaviour*, 1(1), 1-9.
18. Fecher, B., & Friesike, S. (2014). *Open science: one term, five schools of thought* (pp. 17-47). Springer International Publishing.

19. Vicente-Saez, R., & Martinez-Fuentes, C. (2018). Open Science now: A systematic literature review for an integrated definition. *Journal of business research*, 88, 428-436.
20. Bezjak, S., Clyburne-Sherin, A., Conzett, P., Fernandes, P. L., Görögh, E., Helbig, K., ... & Tennant, J. (2018). The open science training handbook.
21. McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., ... & Yarkoni, T. (2016). How open science helps researchers succeed. *elife*, 5, e16800.
22. Sansone, S. A., Rocca-Serra, P., Field, D., Maguire, E., Taylor, C., Hofmann, O., ... & Hide, W. (2012). Toward interoperable bioscience data. *Nature genetics*, 44(2), 121-126.
23. Dimitrov, K. (2023). A debate about emerging immersive technologies in the context of "Higher education 4.0". *Trakia Journal of Sciences*, 21(1), 242-247.

# Spectral Analysis the Wave Signal of Electronic Sphygmomanometer

**Lubomir Lahtchev**

Institute of Robotics, Bulgarian Academy of Sciences, Sofia, Bulgaria  
*llahtchev@abv.bg*

**Abstract.** The wave signal of electronic sphygmomanometer, recorded in personal computer provides whole information for the wave character of the artery blood flow. The corresponding function can be processed by discrete Fourier transformation described by a mathematic method in the paper. The algorithm consists of reducing the digital function to an acceptable size, discrete Fourier transformation and its graphic presentation. The frequency response covers wide range of the human spectrum and bears specific wave picks at given human frequency components. On their base, a general issue is formulated and suggested.

**Keywords:** Electronic sphygmomanometer, Spectral analysis, Wave blood pressure signal.

## 1. Introduction

The recognition of heart abnormalities is a task of important performance. “One way to detect abnormalities in the cardiovascular system especially in the heart is trough ECG reading.” [1]. But there are situations at which such an ECG is not available. The sphygmomanometer is an acceptable unit for reading the arterial blood pressure. The duo control sphygmomanometer of Hartmann provides wave arterial signal and it is used in this manuscript.

The Fourier transformation is a mathematic tool for researches in the physics and engineering. There are transformations on the material axis, on the periodic function and on the discrete function. The last one is applied in engineer practice and in the current exploration.

In the current research a new non invasive tool, like electronic sphygmomanometer provides stationary well repeatable input data about accidental blood pressure. It is spontaneous and in the same time precise measurement. The input frequency domain of tension sensor up to 500 Hz of recording covers physiological frequencies of 62 Hz, what could serve for more accurate exploration. The stationary of the measurement process is presented in [2].

The mathematic spectral characteristics are described in [3] and one of them is the discrete Fourier transformation. In the computer literature, for example [4], the discrete Fourier transformation is presented for the microprocessor applications.

Following that algorithms the developed solution is programmed in VBA for Excel – a high level computer language for applications and the results are graphically exposed.

## 2. Methods

The mathematic model of the discrete Fourier transformation is presented as a sum [5] as follows:



$$X_s = 1/N \left( \sum_{i=0}^{N-1} x_i * e^{jin} \right) \quad (1)$$

Each one discrete image represents a sum of the real and imaginary components of the all data averaged by their number.

$$e^{jin} = \cos(in) + j \sin(in) \quad (2)$$

The equation 2 manifests the complex entity of the Fourier transformation. Thus, two procedures of summation are going in parallel: the sum of real addends and the imaginary addends. The corresponding process can be described by the matrix multiplication:

$$\begin{pmatrix} X_0 \\ X_1 \\ X_2 \\ X_3 \end{pmatrix} = \begin{pmatrix} e^0 & e^0 & e^0 & e^0 \\ e^{0*1} & e^{1*1} & e^{2*1} & e^{3*1} \\ e^{0*2} & e^{1*2} & e^{2*2} & e^{3*2} \\ e^{0*3} & e^{1*3} & e^{2*3} & e^{3*3} \end{pmatrix} * \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix} \quad (3)$$

where the square matrix has real and imaginary components.

This procedure can be reduced by the mathematic property  $e^{in} = e^{in \bmod (N)}$  or if  $i*n > N-1$  the order begins from the start. This operation can be completed with integer division, whose result is residual of it: for example,  $9\%4=1$ ,  $6\%4=2$  and so on. The real and imagine results can be summed in the complex area, following the Pythagoras theorem.

$$S_x = (1/N) * (X_{c2} + X_{s2})^{(1/2)} \quad (4)$$

The spectrogram contains inversed relatively to the start number  $x_0$  values of the above result for each one image of the complex domain.

$$S_\Sigma = x_0 - S_x \quad (5)$$

If the spectrum must be verified by back Fourier transformation, one and the same ort plane will be found for every image.

Finally, the graphics of spectrogram shows the distribution of wave and phase results, characterizing the wave properties of the wave signal until domain of 62.5 Hz.

The frequency range of a BP measurement can be computed by the next expression:

$$F_x = \frac{R_f * t / \Delta t}{N} \quad (6)$$

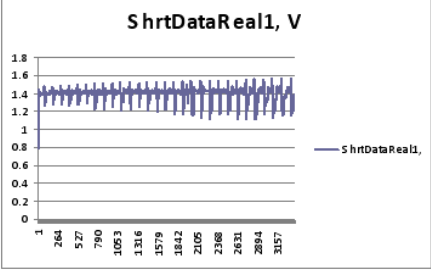
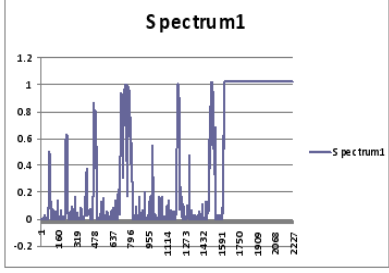
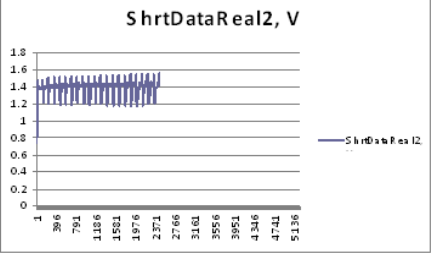
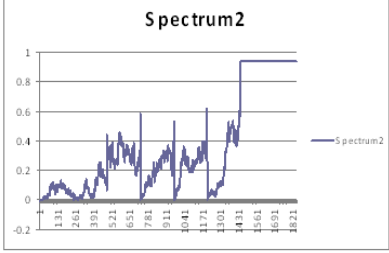
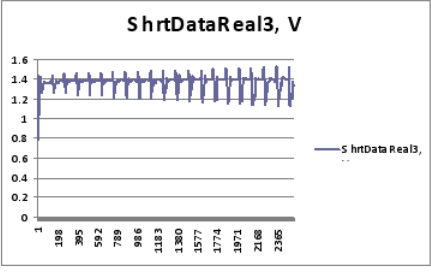
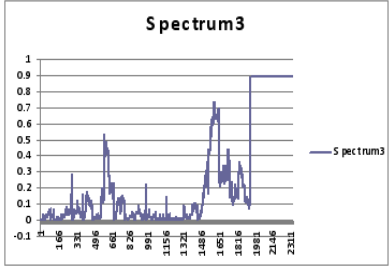
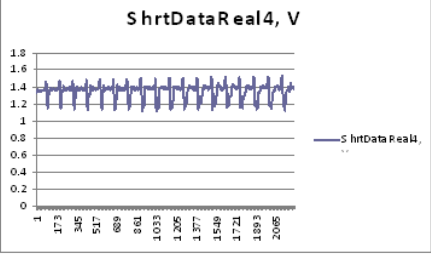
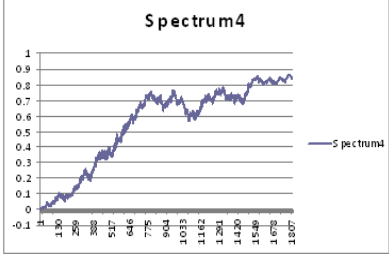
where  $R_f$  – frequency domain of wave signal,  $t$ ,  $\Delta t$  – time moment and time discrete of the array,  $N$  – length of the array at maximum scanning frequency for a minute. This formula also works for transformation from frequency to time domain and from integer numbers to frequency Hz.

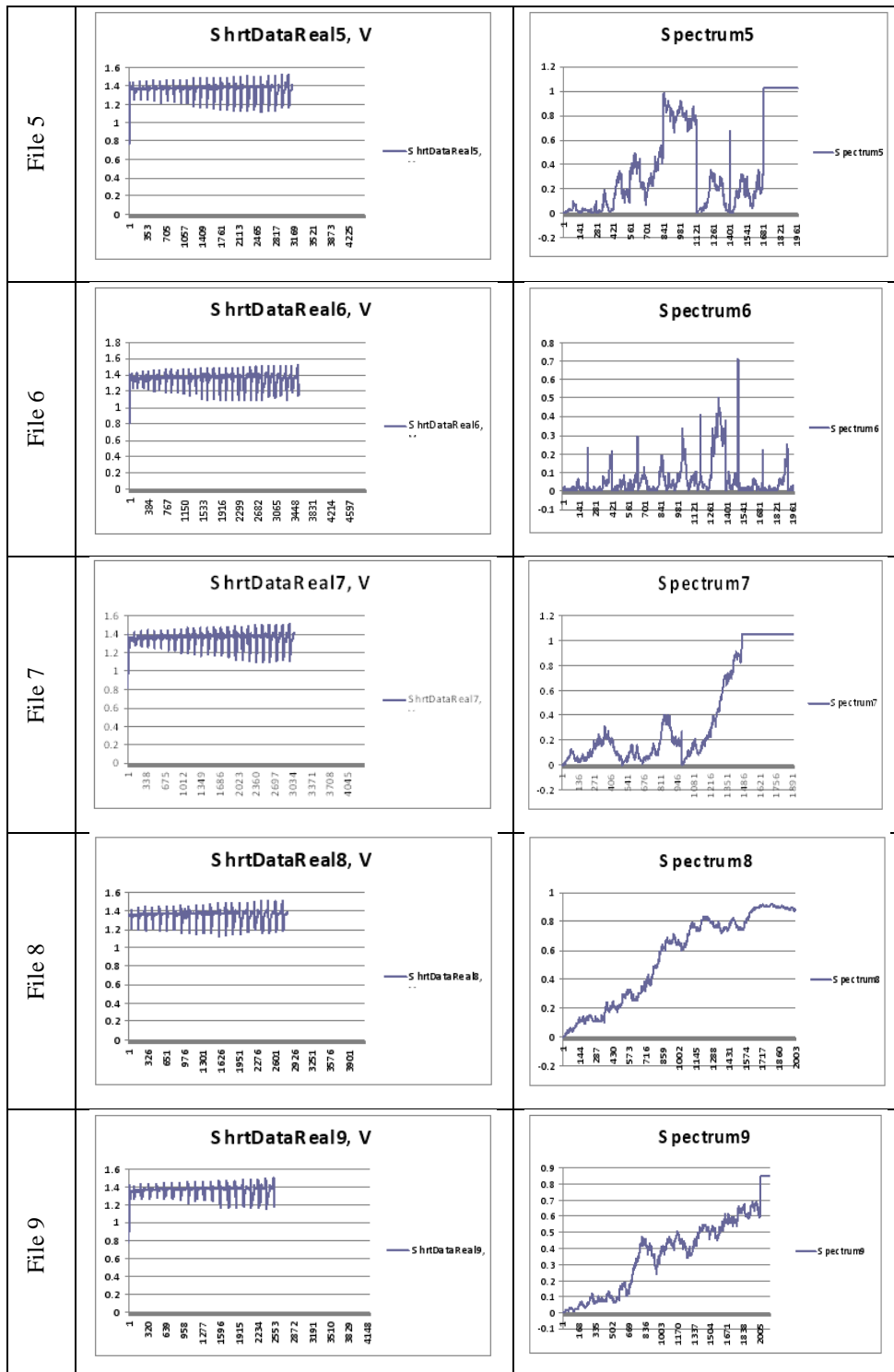
### 3. Results

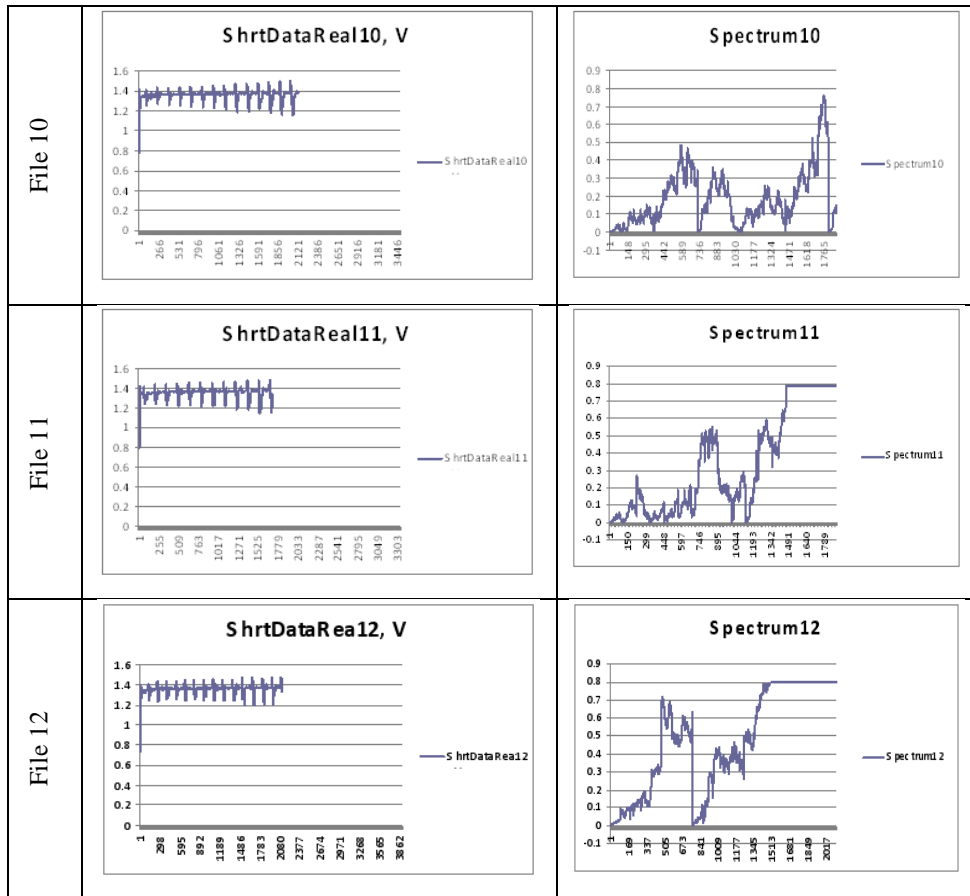
The spectral method is applied over twelve independent files of different input discrete data arrays for the microphone wave signal of electronic sphygmomanometer at measurement the BP. The program is written on VBA for Excel. It is one and the same for all files.

The algorithm includes: reduction of data by reading of every 4-th values; computation the table of trigonometric functions, computation the spectral images by recursion and inversion. The results of it are exposed on table 1. Thus the software works until 5 minutes for each one spectral result of a file.

Table 1. Original signal in time domain and spectral graphic of the wave signal

File	Real wave signal in time domain	Spectral result in frequency domain
File 1		
File 2		
File 3		
File 4		





The spectrograms vary and can not be base for classification. They vary by array length of input data, number of extremes, corresponding frequencies, and widths of groups with extremes. Thus, they show the variability of wave signal the artery. On the other side they contain special quantitative information. They can be described quantitatively by new parameters shown on table 2. It is build with one and the same software program for all files.

Table 2. Number description of spectral parameters: groups, extremes and frequencies.

File No	Group, No	Extreme, No	Amplitude, V	Beginning, No	Beginning, Hz	End, No	End, Hz	Duration, No	Duration, Hz
1	1			80	0.66675	93	0.766791	13	0.1
1		1	0.506517	84	0.700114				
1	2			402	3.350544	422	3.517238	20	0.1667
1		1	0.3766	410	3.42				
1	3			465	3.87563	503	4.192347	38	0.3167
1		1	0.813	488	4.06				

1	4			703	5.86	816	6.8	113	0.94
1		1	0.94	719	5.99				
1	5			1488	12.40201	1544	12.86876	56	0.4667
1		1	1.019151	1510	12.58538				
2	1			486	4.05	727	6.05833	241	2.0083
2		1	0.45337	577	4.808333				
2	2			784	6.533333	794	6.616667	10	0.0833
2		1	0.167603	788	6.566667				
2	3			813	6.775	967	8.058333	154	1.2833
2		1	0.53928	967	8.058333				
2	4			1013	8.441667	1208	10.06667	195	1.625
2		1	0.627143	1208	10.06667				

File No	Group, No	Extreme, No	Amplitude, V	Beginning, No	Beginning, Hz	End, No	End, Hz	Duration, No	Duration, Hz
2	5			1323	11.025	1446	12.05	123	1.025
2		1	0.609111	1444	12.03333				
3	1			587	4.891667	683	5.691667	96	0.8
3		1	0.536937	594	4.95				
3	2			973	8.108333	974	8.116667	1	0.0083
3		1	0.223354	973	8.108333				
3	3			1505	12.54167	1914	15.95	409	3.4083
3		1	0.365468	1843	15.35883				
3	4			1922	16.01667	1923	16.025	1	0.0083
3		1	0.127214	1923	16.025				
4	1			241	2.008333	1818	15.15	1577	13.142
4	1	1	0.639563	700	5.833333				
4	1	2	0.681176	728	6.06667				
4	1	3	0.677354	750	6.25				
4	1	4	0.727117	773	6.441667				
4	1	5	0.756281	804	6.7				
4	1	6	0.711971	838	6.98333				
4	1	7	0.716604	847	7.058333				
4	1	8	0.692607	881	7.331667				
4	1	9	0.80216	1505	12.54167				
4	1	10	0.851195	1563	13.81667				
4	1	11	0.84361	1658	13.81667				

4	1	12	0.844711	1731	14.425				
5	1			709	5.908	1127	9.39	418	3.482
5	1	1	0.990309	854	7.11667				
5	1	2	0.92288	893	7.441667				
5	1	3	0.917779	993	8.275				
5	1	4	0.877568	1102	9.18333				
6	1			844	7.033333	851	7.091667	7	0.0583
6		1	0.192773	849	7.075				
6	2			1018	8.48333	1062	8.85	44	0.3667
6		1	0.338904	1030	8.58333				
File No	Group, No	Extreme, No	Amplitude, V	Beginning, No	Beginning, Hz	End, No	End, Hz	Duration, No	Duration, Hz
6	3			1281	10.675	1394	11.61667	113	0.9417
6		1	0.503884	1340	11.1667				
6	4			1501	12.50833	1502	12.51667	1	0.0083
6		1	0.715814	1502	12.51667				
6	5			1716	14.3	1718	14.31667	2	0.0167
6		1	0.221874	1717	14.30833				
6	6			1907	15.89	1928	16.06667	21	0.1767
6		1	0.25345	1921	16.00833				
7	1			748	6.23333	762	6.35	14	0.1167
7		1	0.173058	754	6.283333				
7	2			802	6.683333	986	8.216667	184	1.5333
7	2	1	0.273691	818	6.81667				
7	2	2	0.338392	829	6.908333				
7	2	3	0.377168	844	7.033333				
7	2	4	0.233408	937	7.808333				
7	2	5	0.212866	964	8.03333				
7	2	6	0.272052	985	8.208333				
8	1			355	2.958333	2006	16.71667	1651	13.758
8	1	1	0.583768	846	7.05				
8	1	2	0.684337	906	7.55				
8	1	3	0.907919	1716	14.3				
8	1	4	0.912482	1726	14.38333				
9	2			586	4.883333	2020	16.8333		
9	2	1	0.182783	608	5.06667				

9	2	2	0.279497	725	6.041667				
9	2	3	0.32655	754	6.283333				
9	2	4	0.44387	840	7				
9	2	5	0.286428	954	7.95				
9	2	6	0.294619	966	8.05				
9	2	7	0.537377	1442	12.01667				
9	2	8	0.614344	1692	14.1				
9	2	9	0.694229	1939	16.15833				
File No	Group, No	Extreme, No	Amplitude, V	Beginning, No	Beginning, Hz	End, No	End, Hz	Duration, No	Duration, Hz
10	1			765	6.375	998	8.316668	233	1.9417
10	1	1	0.360766	849	7.075				
10	1	2	0.332431	931	7.758333				
10	1	3	0.284	973	8.108333				
11	1			715	5.958	1019	8.491667	304	2.5337
11	1	1	0.358281	749	6.241667				
11	1	2	0.540284	838	6.98333				
11	1	3	0.646014	1444	12.0333				
11	1	4	0.675968	1471	12.25833				
12	1			296	2.466667	779	6.491667	483	4.025
12	1	1	0.606373	698	5.816667				
12	1	2	0.634631	779	6.491667				
12	4			892	7.433333	1510	12.58333	618	5.15
12	4	1	0.298925	944	7.86667				
12	4	2	0.430234	991	8.258333				
12	4	3	0.443933	1023	8.525				
12	4	11	0.784816	1451	12.09167				
12	4	12	0.784	1493	12.44167				
12	4	13	0.79844	1510	12.58333				

The size of signal array is taken from point B – stop the pump to point G – end deflating the cuff and is reduced four times by reading every fourth data. The Fourier transformation is applied to the reduced array with accounting the double reduction of the frequency when the periodic variation is introduced. Every file has different size of the array and accordingly different frequency spectrum from 24 Hz to 38 Hz with prevail of frequencies over 30 Hz.

The human body organs vibrate with frequency of 6 Hz, 7 Hz, and 8 Hz [5]. It can be expected that frequencies of 3 Hz, 3.5 Hz, 4 Hz, can also provide useful support of the body

vibration. On the other side, extremes with frequencies of 12 Hz, 14 Hz, and 16 Hz confirm the availability of vibrations 6 Hz, 7 Hz, 8 Hz.

Table 2 represents the group of extremes and the extracts of extremes with corresponding frequencies for each one file. Thus, we can accomplish the analysis of results.

We can see that file 1 has extremes with frequencies 5.86 Hz, 5.99 Hz, 6.317 Hz, 12.4 Hz, 12.58 Hz. Obviously, the artery is of frequency around 6 Hz, confirmed by spectral extreme of frequency 12.4 Hz. Following the same context we can reveal that file 2 has extremes of frequencies 6.567 Hz and 12.033 Hz. File 3 has extremes of frequencies 4.95 Hz, 8.1 Hz, 12.5 Hz, and 16.016 Hz. File 4 has extremes of frequencies 6.067 Hz, 6.98 Hz, 8.01 Hz, 12.54 Hz, 14.93 Hz. File 5 has extremes of frequencies 5.908 Hz and 7.11 Hz. File 6 has extremes of frequencies 7.075 Hz, 8.58 Hz, 12.516 Hz, 16.008 Hz. File 7 has extremes of frequencies 6.23 Hz, 7.03 Hz, 8.03 Hz, 12.05 Hz. File 8 has extremes of frequencies 6.1 Hz, 7.05 Hz, 8.13 Hz, 12.11 Hz, 14.3 Hz, 15.74 Hz. File 9 has extremes of frequencies 6.04 Hz, 7 Hz, 8.05 Hz, 12.01 Hz, 14.1 Hz, 16.15 Hz. File 10 has extremes of frequencies 6.375 Hz, 6.61 Hz, 7.075 Hz, 8.1 Hz. File 11 has extremes of frequencies 5.958 Hz, 6.24 Hz, 6.983 Hz, 12.03 Hz. File 12 has extremes of frequencies 5.817 Hz, 6.49 Hz, 7.86 Hz, 12.09 Hz. It can be seen that despite some deviations like files 6, 7, 10 and 12 as whole extremes exist everywhere inside the searched ranges and especially for files 1, 2, 3, 4, 9. Extremes of frequencies 6 Hz, 7 Hz and 8 Hz in the spectrums reflect the vibration of human artery like a healthy normal organ. Thus the sphygmomanometer Hartmann Duo Control with a wave signal provides verified work of the human artery.

In parallel, extremes of deviations inside the frequencies 9 Hz, 10 Hz and 11 Hz can also be observed within files 1, 2, 4, 5, 6, 7, 8, 9, 11, 12. The corresponding extremes do not prevail over extremes of 6 Hz, 7 Hz and 8 Hz and consequently their character does not fluctuate importantly.

Following these observations, a general issue can be formulated. In parallel to the variable differences of the wave spectrums of the wave sphygmomanometer signal at measurement the blood pressure the healthy frequencies 6 Hz, 7 Hz, 8 Hz and supporting them frequencies 12 Hz, 14 Hz, 16 Hz are a main factor for the normal work of the human body. The availability of extremes of frequencies 9 Hz, 10 Hz, 11 Hz within the same signal contribute some caution at the diagnosis of uncertainties. These circumstances encourage explorations in the pointed scientific domain.

#### 4. Conclusions

The measurement of the heart frequency with sphygmomanometer Hartmann Duo Control provides an array of discrete number data, when it is connected to the personal computer through analog digital converter, for example DT 9810. This array can be used for spectral analysis in the time domain of measurement, respectively from p. B – end of pump work to p. G – full air deflation of the cuff. The corresponding frequency spectrum contains spectral extremes. For the healthy human the extremes of frequencies 6 Hz, 7 Hz, 8 Hz are observed everywhere independently of form of the spectrum. Often these extremes are supported by extremes of frequencies 3 Hz, 3.5 Hz, 4 Hz and 12 Hz, 14 Hz, 16 Hz inside the same spectrum data array. In parallel, extremes of frequencies 9 Hz, 10 Hz, 11 Hz are also observed and they do not respond to the expected healthy conditions. These frequencies do not prevail over the extremes of the healthy ranges and they can only bear caution character. Consequently, availability of variability of the artery frequencies, in parallel to the extremes inside the healthy frequencies 6 Hz, 7 Hz, 8 Hz and 12 Hz, 14 Hz, 16 Hz allow to issue the normal healthy condition of the individual. For achievement, the pointed purpose a mathematic method for discrete spectral analysis of Fourier transform is suggested. It is accomplished over



wave signal from sphygmomanometer Duo Control of Hartmann. The computer programs work with random files of the same Excel format. The frequency range of spectral periodograms is defined by elementary time period of the discrete values of the blood pressure or 0.008 – 0.016 seconds. For different files, the array of values varies by number of discrete. The software consumes less than 5 minutes for spectrum computation and less than a minute for number description of every file.

## References

1. Purvanti E., A. K. Nastiti, A. Supardi (2014) Abnormality Classification Using Fourier Transforms Method and Neural Networks. Indonesian Journal of Tropical and Infectious Disease, vol. 5, No 2, May – Aug. 2014, Research report.
2. Lahtchev L. (2013) Verification of Hypothesis about Blood Pressure by Statistic Estimations and Distributions. Proc. of Int. Conference Automatics and Informatics 2013, 3 – 7 Oct, Sofia, Bulgaria, pp. I 33 – I 36.
3. Jablon C., J.-C. Simon (1983) Application des Modeles Numeriques en Physique. Birkhäuser Verlag Basel, 1978, In Russian, Moscow, “Nauka”, 1983, 236 p.
4. Yuen C.K., K.G. Beauchamp, G.P.S. Robinson (1986) Microprocessor Systems and Their Application to Signal Processing. Academic Press, 1982, Second Printing. In Russian. Moscow, “Radio i Svyaz”, 1986, 296 p.
5. Horbenko I. G. (1978) Sound, Ultrasound, Infrasound. In Russian. Moscow, “Znanie”, 1978, 159 p.

# **Crowdfunding as an Alternative Source of Investments in the Conditions of Economic Digitalization: Relevance of Use in the Republic of Moldova**

**Irina Dorogaia<sup>1</sup>, Alexandr Scutari<sup>2</sup>**

<sup>1</sup> Academy of Economic Studies of Moldova, Chisinau, Republic of Moldova

<sup>2</sup> BALCANICA SRL, Chisinau, Republic of Moldova  
*dorogaia.irina.ion@ase.md, alscut@gmail.com*

**Abstract.** The relevance of the topic is associated with the increasing popularization of alternative sources of financing of entrepreneurial activity. Under the current conditions, the search for means to finance entrepreneurial activity is a priority task for both entrepreneurs and state institutions, oriented towards supporting existing and potential businesses. The purpose of the article is to describe the application of crowdfunding as one of the tools for alternative financing of projects in the current realities of building a digital economy in Moldova. The article describes the legal aspects of crowdfunding regulation as a new for Moldova legal institute of non-banking lending using the possibilities of online platforms. The paper considers the economic and legal essence of crowdfunding, the model of functioning of this institution, the opinion on the main advantages and disadvantages of this alternative type of lending, as well as outlines the prospects of development and possibilities of application of the new financial technology in our country.

**Keywords:** Crowdfunding, Crowdlending, Crowdfunding platforms, Digital economy, P2P lending.

## **1. Introduction**

In recent years, with the growth of digitalization and digitalization, especially intensified during the pandemic, there has been a societal demand for increased mobility, including of financial services and products. It has become relevant to receive services via the Internet and applications on electronic devices in real time, preferably with a choice of service providers in one place.

The sources of innovation financing in the modern economic model of the Republic of Moldova are financial institutions, organizations, funds, as well as the population, which provide money for innovation projects.

The aim of the research is to study the peculiarities of crowdfunding as a phenomenon of alternative financing, to clarify the legislative framework for this process and to create conditions for the development of processes related to crowdfunding in the Republic of Moldova.

## **2. Diversity of financing models**

The private group is dominated by alternative investor-assisted financing models. We distinguish two groups:

- 1). 3F (Family, Friends, Fools) - sponsors (family, friends and "naive") financing the initial stage of the project from their own cash [1].

- 2). Crowdinvestors - backers, an association of sponsors who are able to invest in projects through crowdfunding platforms [3].

At the same moment, there is a need to organize a controlled market for digital assets that will have the advantages of assets while providing an acceptable level of investor protection (including judicial), reducing their credit risks, and identifying the issuers and owners of digital assets. Responding to the needs of the market, society and economy for new alternative financial instruments that allow to obtain higher returns at a controlled level of risks, the concept of "collective platforms for attracting investments" was introduced in the Moldovan legal framework.

This collective cooperation of individuals and legal entities investing their resources through the Internet to support and develop startups and innovative projects was called "crowdfunding" or "crowdfunding". Its main essence is that the organization of raising funds for their implementation is carried out in small batches from an unlimited number of users [2].

Crowdfunding platforms - electronic platforms functioning in accordance with the sectoral legislation of RM and using investment platforms, which carry out activities to organize the attraction of investments by providing loans (crowdfunding), acquiring shares in the social capital of enterprises (crowdinvesting) and utilitarian digital rights (UCR) [4].

Their purpose is to bring together investment attractors and investors, ensuring the safety of settlements and legal purity of the transaction.

It should be noted that the formation of institutional conditions for the development of non-banking financing system was also envisaged by the governmental program of economic development, the Strategy of Financial Market Development of the Republic of Moldova until 2030.

### **3. Fundamentals of legal regulation of Crowdfunding**

Currently, many countries are adopting and enacting sectoral regulations governing the activities of crowdfunding platforms due to the fact that this financing mechanism is relatively new.

In addition, these legal relations are often partially regulated by regulatory legal acts governing the issues of protection of the rights of consumers of financial services, licensing of activities in the market of financial services, etc. At the same time, a number of states have already developed special legislation regulating fundraising activities using online services.

For example, France was one of the first states to adopt special sector-specific legislation to regulate crowdfunding within the EU.

In October 2014, a law facilitating the operation of crowdfunding platforms came into force. Investment portals and intermediary resources were divided into **3 categories**:

- platforms for loans with or without interest (IFP);
- resources for the realization of shares (CIP);
- resources for the provision of investment services (PSI).

All three types of platforms must be registered with the Association for the Certification of Intermediaries in Insurance, Banking and Finance (ORIAS) [5].

In 2018, the European Union took the initiative to harmonize the legislation of EU Member States by presenting a draft Regulation on the regulation of providers of crowdfunding services for entrepreneurial purposes. As a result, the EU Directive (Regulation (EU)) was adopted 2020/1503 - on providers of crowdfunding services in the EU [6].

In order to streamline the relations arising from the borrowing of funds through online services, as well as to increase the transparency of the financial market in this crowdfunding segment, in July 2023, the Parliament of Moldova adopted Law No. 181/2023 "On Collective

Financing Services" and on amending certain legislative acts of the Republic of Moldova", which regulates the relations arising from investing and attracting investments using collective platforms. This law limits the maximum value of the invested amount in the projects of applicants at the level of 1 million euros, and for individuals the amount of crediting at the level of 7500 euros. Also, the law stipulates that in order to increase the level of control, the National Commission of Financial Market should keep a register of platform operators, while the amount of own funds of this operator should be at least 500 thousand lei. [7].

Participants of the crowdfunding platform are: directly the operator of the platform, which must be registered in the register of NCFM of Moldova; borrowers - organizations and individuals, placing information about their projects or obtaining a loan on the platforms in order to attract investments; investors - legal entities and individuals, investing in one or more projects in accordance with the established limits.

It should also be noted that the loans attracted are not insured, and platforms are not liable for the obligations of companies attracting financial resources. Taking into account that platforms usually host "young projects" (startups), the probability of failure to repay the invested capital is extremely high. To eliminate the risk of default, operators of crowdfunding platforms, in addition to the legal verification of the project, are required to more thoroughly assess the solvency of its business model, as well as to introduce restrictions on the maximum amount of investment in projects at early stages of development [7].

To assess the creditworthiness of borrowers, platforms need to conduct high-quality scoring based on modern innovative technologies of fast online analysis without losing the quality of expertise (KYC&AML system).

In order to reduce investor risks, in addition to the formation of reserve funds, based on the experience of crowdfunding platforms operating in the EU, it is necessary to provide for the possibility of introducing a collateral mechanism.

Nevertheless, in order to maximize the potential of crowdfunds, it is necessary to further improve external conditions and regulatory mechanisms. Given the current challenges of the local market, it is necessary to create an adequate legislative framework focused on "soft" regulation that does not hinder the development of effective crowdfunding models. Regulatory regulation should be aimed, first of all, at increasing confidence in crowdfunding technologies, protecting the rights of investors, lenders, entities raising funds, as well as specialized platforms [8].

#### **4. Description of the crowdfunding platform business model**

The main objective of the formation and implementation of crowdfunding is the need to raise money for the implementation of a common idea or project. The general description of the mechanism of attracting and granting loans is as follows:

a. a funder (a legal entity or an individual) places an application on the crowdfunding platform to attract a loan. The borrower determines the terms and conditions of attracting a loan by filling in the relevant fields on the service's website, or agrees to the proposed terms and conditions. The borrower must also familiarize himself with the rules of its operation posted on the website of the online borrowing service, provide the necessary information (scanned copies of documents, description of the financed project (if any), account details for crediting funds, etc.);

b. based on the received application and information, the platform operator makes a decision on the possibility of placing on its website the borrower's proposal to attract financing and performs a preliminary assessment of the borrower and the financed project (if any) in accordance with the criteria and methods independently determined by the service operator;

c. potential investors (backers), having studied the application and information about the borrower and the project being implemented by the borrower on the service's website, make a decision on granting loans by taking appropriate actions (filling in fields, providing information, confirming agreement with the terms of the loan and familiarizing themselves with the rules of the platform's activities, etc.);

d. if there is a potential investor (group of investors) who has the ability to provide the necessary loan amount, the funds are transferred directly to the account of the funder;

e. the funder fulfills its obligations to investors on payment of interest and loan amount by transferring the respective monetary amounts;

f. both parties provide information on all facts of granting and receiving the loan, its repayment, as well as on cases of non-fulfillment of their obligations to the platform operator and counterparties in real time, using the services and functionality of the platform. [4]

The operator of the crowdfunding platform carries out its activities in accordance with the rules of use of the investment platform approved by it, which must comply with the requirements established by the National Bank and posted on the platform's website in the public domain on the Internet.

Thus, the advantage of crowdfunding for SMEs is time saving, as the process of raising the necessary amount of money requires less documentation due to the use of digital technologies. Despite the fact that crowdfinancing is not currently considered a competitor to bank loans, the convenience and speed of borrowing through platforms will allow crowdfinancing to take a certain market share in the lending sector in the medium term. Despite a number of significant advantages over traditional methods of financing and its growing popularity, crowdfinancing has some drawbacks, such as lack of public awareness; incompetence of the project author in terms of promotion; mistrust of the author and his project on the part of a potential investor, etc.

In modern conditions, crowdfunding is an innovative tool for the development of Moldova's digital economy, and the popularity of this particular method of financing is explained by the speed and accessibility of collecting the necessary amount of money without the direct involvement of bank lending.

## **5. Crowdfunding Industry 2023: Stats, Insights, and Future Horizons**

The European crowdfunding industry has experienced significant growth and has become an integral part of the global financial landscape. Alternative investment platforms have played a crucial role in this revolution, providing both entrepreneurs and investors with greater access to capital. This summery explores the thriving crowdfunding ecosystem in Europe, shedding light on key statistics and important insights.

Alternative investment platforms have gained popularity worldwide, as investors are drawn to their potential for high returns and diversification. Currently, there are 510 platforms operating globally, as reported by p2pmarketdata.com. These platforms not only offer financial benefits but also provide investors with the opportunity to make a positive impact on society and the environment through their investments.

In the United States, a similar trend has emerged, with alternative investment platforms experiencing a surge in popularity. This can be attributed to the growth of equity crowdfunding and real estate crowdfunding. Platforms such as AngelList and Fundrise have gained widespread recognition and have attracted a significant number of investors who are seeking non-traditional investment opportunities. [11]

This discrepancy in the number of crowdfunding platforms per capita across Europe can be attributed to various factors. Big and mature European economies tend to host most of them, including the UK (94), Italy (66), France (63), and Germany (46). However, if we correct for the market size, the Baltic states stand out. Estonia hosts the most platforms per capita – about 22 per 1 million people, followed by Latvia (4.7) and Lithuania (3.9). By comparison, in the UK, there are only 1.5 platforms per million people, in Italy – 1.1 and Germany – 0.6. Finally, countries in Central, Eastern, and South-Eastern Europe seem to lag in the crowdfunding craze. Some big markets like Poland, Hungary, and Romania have almost non-existent alternative investment markets (Fig.1).

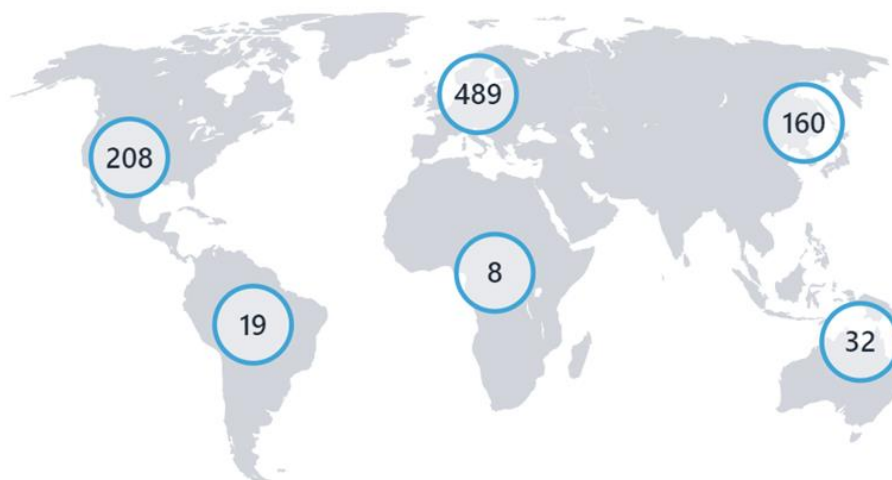


Fig. 1. Number of platforms worldwide (n=914) [12]

However, by investment volume, the US still likely surpasses the entire old continent. According to the Cambridge Centre for Alternative Finance, the US was by far the largest market in the world in 2020, having funded \$73.62 billion of total volume (corresponding to 65% of the global market share). Europe trailed far behind with \$22.6 billion, over half of which came from the UK alone (\$12.6 billion).

Globally, investment crowdfunding generally offers high-risk high-return opportunities. Looking at investment segments, equity platforms (including start-up and real estate equity) account for almost half of all platforms, while peer-to-peer lending trails at around one-third share of all platforms (Fig.2.). The growth of the “other” segment, including mostly crypto and collectible investments is also notable [12].

However, investment products vary across markets. For example, Europe is the cradle of debt investments, while the US rocks the venture sector. Furthermore, the US is much more prone to purely digital investments: almost a third of all alternative investments there go to crypto and collectibles.

The global crowdfunding market has experienced significant growth, with projections set to climb further. From \$13.4 billion in 2022, the market is expected to reach \$15.54 billion in 2023, reflecting a Compound Annual Growth Rate (CAGR) of 16%. Although the Russia-Ukraine war temporarily impacted global economic recovery from the COVID-19 pandemic, crowdfunding remains resilient. Economic sanctions, inflation, and supply chain disruptions have driven inflation across various sectors. However, the crowdfunding market is projected to surge to \$27.81 billion by 2027 at a CAGR of 15.7%. [13]

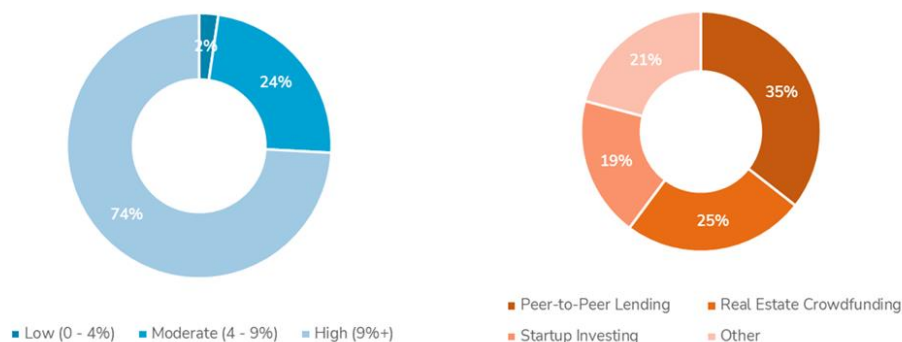


Fig. 2. Share of worldwide platforms by expected return and investment segment [12]

In 2022, monthly funding volumes decreased quite sharply in the first quarter, driven mostly by the invasion of Ukraine and the resulting declines in consumer lending in the resale lending platforms with exposure to the Russian and Ukrainian markets.

Funding volumes in the second quarter and throughout the summer were mostly dragged down by historically the largest lender – Mintos. Mintos went from funding almost €200 million worth of loans in October 2021 to just €35 million in July 2022. Business lending also entered a usual summer slowdown in July-August.

The end of the year ended on an optimistic note with funding volumes finally climbing back to roughly the levels of late 2021, fueled by both Mintos' rebound and the renewed activity in the business sector.

When we analyze the crowdfunding industry over a longer timeframe, we observe two significant trends. Firstly, the sector is dynamic and susceptible to unexpected events. For example, during the first wave of the Covid-19 pandemic from February to April 2020, investments in crowdfunding declined, and the overall market shrank by approximately 23% compared to 2019. Additionally, in 2022, there was another downturn in the industry, although less severe, due to challenging global macroeconomic conditions. However, these setbacks actually contribute to shaping and fortifying the emerging crowdfunding sector.

Secondly, despite the recent challenges, the industry has been experiencing exponential growth, with funding volumes more than doubling between 2018 and 2022. This positive trend indicates the resilience and potential of the crowdfunding sector, even in the face of adverse circumstances.

**Key Players:** The crowdfunding arena is a dynamic and rapidly evolving sector of the global financial market. The European region, in particular, has seen a significant surge in crowdfunding activities, driven by the increasing popularity of alternative investment platforms. The Europe Crowd Lending and Crowd Investing Market size is expected to grow from USD 13.21 billion in 2023 to USD 14.41 billion by 2028, at a CAGR of 1.76% during the forecast period (2023-2028). The continent holds a significant position in this landscape, housing 379 of these platforms.

The United Kingdom holds a commanding position in the European crowdfunding scene, boasting a total of 59 platforms. The UK's regulatory environment, which is supportive of financial innovation, has played an important role in fostering the growth of these platforms. Investors in the UK have been attracted by the potential for high returns that are projected from 4 to 9% on average.

Talking about interest rates they can vary significantly depending on the type of investment, for example: On average, P2P lending platforms in the UK have offered interest

rates ranging from 3% to 7% or even higher in some cases and Real Estate Crowdfunding interest rates can range from 5% to 10% or even more. Some British platforms from real estate crowdfunding and peer-to-peer lending are: Shojin, Proplend, Money&Co, The Money Platform, and more.

Italy is another European country that has made significant strides in the alternative investment sector. With 48 platforms, Italy has a vibrant and diverse alternative investment scene. Italian investors have shown a keen interest in peer-to-peer lending and equity crowdfunding, driven by the potential for attractive returns that range from approximately 5% to 10%, depending on the specific platform and the risk profile of the loans.

Italian alternative investment platforms have also been instrumental in driving economic growth in the country. By providing a source of funding for startups and small businesses, these platforms are supporting innovation and job creation in the Italian economy. Some of the important players in this sector include: Invest-t, Business Lending, Mamacrowd, WeAreStarting, and more.

France, with 45 platforms, has a dynamic and evolving alternative investment market. Real estate crowdfunding has been particularly successful in the country, with several platforms offering investors the opportunity to invest in property projects across the country. French alternative investment platforms have also been at the forefront of the trend towards socially responsible investing, with several platforms offering investments in renewable energy projects and socially responsible businesses. Peer-to-peer lending and startup investing are the most popular categories in the country with platforms like: myOptions, Solylend, Tudigo, Proxima, and more.

**Types of Crowdfunding Platforms:** Crowdfunding in Europe can be segmented into various categories, each with its unique statistics:[1]

- **Peer-to-peer lending:** This segment has 161 platforms in the EU, with a minimum investment threshold of just €1. Leading this space is Mintos (Latvia), having facilitated approximately €9 billion in funding. Other significant players include Peerberry (Czech Republic) with €2 billion and Oplyn (Italy) with €1.7 billion in funding.
- **Real estate crowdfunding:** This thriving segment has 131 platforms in the EU. The minimum investment required is just €1. Noteworthy platforms include Estateguru (Estonia) with €742 million in funding, Kuflink (UK) with £279 million, and CapitalRise (UK) with £204 million.
- **Startup investing:** This sector comprises 77 platforms in the EU. With a minimum investment threshold of €1, these platforms cater to a diverse pool of investors. Italy stands out with 19 platforms, featuring players like Hensoo, 200crowd, and Activant.

**Regional Trends:** The European crowdfunding market is characterized by a wide range of niche platforms catering to specific investment interests. Equity crowdfunding, in particular, is on the rise, offering investors an opportunity to take ownership stakes in startups and small businesses.

The Baltic region, comprising Estonia, Latvia, and Lithuania, has emerged as a burgeoning hub for alternative investment platforms in Europe. With a total of 44 platforms, making up 12% of the EU's total. The Baltics are a testament to the rapid growth and popularity of crowdfunding in the region. The success of the crowdfunding market in the Baltics can be attributed to several factors, including a solid digital infrastructure, a conducive regulatory environment, and a culture of innovation and entrepreneurship.

Estonia stands out as the leader in the Baltic crowdfunding scene, boasting an impressive 26 platforms. This small nation is renowned for its digital-first approach to governance and business, making it a fertile ground for the growth of digital platforms,



including crowdfunding. Explore here some of the Estonian platforms: Income Marketplace, Swaper, IUVO Group, Bondora, Monestro, Reinvest24, and more.

Latvia and Lithuania, with 7 and 11 platforms respectively, are also making their mark in the crowdfunding arena. These countries have embraced digital innovation, creating a conducive environment for the growth of alternative investment platforms.

In Latvia, peer-to-peer lending platforms have gained significant popularity, providing an alternative source of personal and business financing. Meanwhile, Lithuania has seen the growth of equity crowdfunding platforms, offering investors the opportunity to take an equity stake in Lithuanian startups. Explore some of the Latvian and Lithuanian platforms: Twino, Debitum, Mintos, Lande, Heavy Finance, Letsinvest, NEO Finance and more.

Reward-based crowdfunding in Europe is experiencing significant growth due to various factors. The credibility and acceptance of crowdfunding have increased, making it a more attractive option for entrepreneurs and investors alike. The accessibility of crowdfunding platforms and the influential role of social media in promoting campaigns have also contributed to its popularity.

The COVID-19 pandemic has further accelerated the growth of crowdfunding, as businesses sought alternative funding sources to navigate the economic challenges. This trend highlights the resilience and adaptability of the crowdfunding industry.

Looking ahead, the future of crowdfunding in Europe appears promising. Mobile crowdfunding apps are gaining traction, providing investors with the convenience of investing on the go. The rise of the gig economy and freelance work is expected to drive the demand for crowdfunding, as individuals seek funding for their projects.

Notably, the focus on impact investing and market consolidation is anticipated to shape the crowdfunding landscape. Investors are increasingly seeking opportunities to make a positive social or environmental impact through their investments, aligning with the growing interest in sustainable and responsible investing. Furthermore, market consolidation is likely to occur as larger crowdfunding platforms expand their reach and acquire smaller competitors.

Another noteworthy trend is the integration of crowdfunding platforms into lending management systems. This integration streamlines and enhances the management of crowdfunding campaigns, improving the experience for both investors and fundraisers. At Fintech Market, we have strived to support this trend by partnering with various crowdfunding platforms. Our Loan Management System offers the convenience of integrating crowdfunding services into your company, alongside our risk management solution, Decision Engine.

In conclusion, the crowdfunding industry in Europe is experiencing growth fueled by factors such as legitimacy, accessibility, social media influence, and the response to the pandemic. The future of crowdfunding in Europe looks promising, with mobile apps, the gig economy, impact investing, and market consolidation shaping the industry. The integration of crowdfunding platforms into lending management systems is also an emerging trend that enhances the crowdfunding experience for investors and fundraisers.

## **6. Conclusion**

Summarizing the content of this article, we believe it is possible to make the following main conclusions:

1) Crowdfunding platforms are one of the instruments of collective financing and mechanisms for stimulating the transformation of the national economy of Moldova on its path of digitalization.

2) The envisaged regulation of the activity of crowdfunding platforms corresponds to the global trend, in which with the development of digital technologies, the development and

regulation of alternative mechanisms for attracting and providing funds to banks are intensified.

3) The described mechanisms of platforms' operation are in line with the general rules of activity on the financial market of the Republic of Moldova (licensing of activity or inclusion in the register, supervision (control) over the activity, setting requirements for financial market participants, etc.). They allow controlling the cleanliness and transparency of financial operations in the financial market sector, which previously remained outside the regulator's influence.

4) Despite the existence of problems that can be regulated at the legislative level, the mechanism of alternative capital attraction through investment platforms, if properly and competently used, contributes to a partial solution of key socio-economic problems. This is primarily due to the infrastructural role of crowdfinance platforms, on the one hand, stimulating the development of innovative projects, and on the other hand, increasing the investment activity of economic entities and population.

5) Crowdfunding is a method of raising capital to sponsor endeavors and companies. Fundraisers can use online platforms to gather money from a sizable audience. Startup enterprises and expanding businesses most frequently use crowdfunding to obtain alternative capital. It is a creative method of obtaining finance for new endeavors, enterprises, or concepts. It may also be a means of creating a network of support for the offering. One can access new customers and acquire helpful industry information by leveraging the power of the Internet community.

6) Crowd-technologies are a good example of achieving synergy between the need of business structures in financial resources and the desire for profitable investment of capital on the part of qualified and unqualified investors.

7) The European crowdfunding landscape is a vibrant and dynamic space, characterized by innovation, growth, and an ever-evolving array of investment opportunities. As we continue to navigate the digital age, crowdfunding is poised to play an increasingly significant role in the world of finance.

8) The crowdfunding market is witnessing remarkable growth, with an array of innovative solutions and expanding applications. The ability to leverage social networks and online channels has made crowdfunding an accessible and efficient way to gather funds from a diverse and global community. This comprehensive market report provides valuable insights and data to empower investors, startups, and businesses to thrive in the dynamic world of crowdfunding.

## References

1. Dibrova A. Understanding the essence of alternative finance: theoretical aspects // New Challenges of Economic and Business Development. – 2017. – Vol.4. – pp. 133-143.
2. European Commission. Guide on crowdfunding/ What is Crowdfunding? [https://single-market-economy.ec.europa.eu/access-finance/guidecrowdfunding/what-crowdfunding/crowdfunding-explained\\_en](https://single-market-economy.ec.europa.eu/access-finance/guidecrowdfunding/what-crowdfunding/crowdfunding-explained_en)
3. Fiyaksel E.A., Solokhin A.A., Sokolov I.N. Crowdinvesting. Research review and intermediate development results. Innovative Economy. vol. 7, pp. 42-54, 2017 (in Russian).
4. Nazarkin V. Crowdfunding - raising money for your projects. (in Russian). <https://neiros.ru/blog/business/kraudfanding-sobiraem-dengi-na-svoi-proekty/>
5. Kleshcheva, S. A. International experience of regulating crowdfunding as a tool for attracting investments. (in Russian). [https://rep.polessu.by/bitstream/123456789/14676/1/17Kleshcheva%20S.A\\_Mezhdunarodny%20opyt%20regulirovaniia%20kraudfandinga.pdf](https://rep.polessu.by/bitstream/123456789/14676/1/17Kleshcheva%20S.A_Mezhdunarodny%20opyt%20regulirovaniia%20kraudfandinga.pdf)
6. Kuznetsov V. A. Crowdfunding: current regulatory issues. Money and credit, vol. 1, pp. 65-73, 2017.

7. Law of the Republic of Moldova No. 181/2023 "On collective financing services". [https://www.legis.md/cautare/getResults?doc\\_id=138188&lang=ro](https://www.legis.md/cautare/getResults?doc_id=138188&lang=ro)
8. Khoroshev, R. What is missing from the law on crowdfunding? Vedomosti. (in Russian). <https://www.vedomosti.ru/opinion/articles/2019/12/20/819236-hvataet-zakone>
9. Global Crowdfunding Market Size, Status and Forecast 2020–2026. <https://www.absolutereports.com/global-crowdfunding-market-15043142>
10. Kirby, E., Worner, S. Crowd-funding: An infant industry growing fast // Staff Working Paper of the IOSCO Research Department. <https://www.iosco.org/research/pdf/swp/Crowd-funding-An-Infant-Industry-Growing-Fast.pdf>
11. Explore European Alternative Investment Platforms. <https://fintech-market.com/blog/explore-european-alternative-investment-platforms>
12. Crowdfunding Industry Market Update 2023 [Survey of 70 Platforms] <https://p2pmarketdata.com/articles/crowdfunding-statistics-worldwide/>
13. Crowdfunding Industry Report 2023: Empowering Global Causes Through Social Networks // GLOBE NEWSWIRE <https://finance.yahoo.com/news/crowdfunding-industry-report-2023-empowering-081800097.html?guccounter=1>

# A Combination of Investments in Renewable Energy and Heat Pumps in the Private Sector

**Lesan Timur**

Academy of Economic Studies of Moldova, Chisinau, Moldova  
*lesan.timur@ase.md*

**Abstract.** In recent years, Europe has faced an energy crisis, in terms of both the energy prices and the availability of energy supplies. The article discusses an example of the implementation of investments in solar panels, as well as a heat pump in a private household, and calculates the return on investment at the current tariff with net electricity metering.

**Keywords:** Electricity, Renewable energy, Investments, Energy prices, Net-metering

## 1. Introduction

Moldova does not have its own fossil fuel reserves and imports 96% of the required fuel, spending significant financial resources on this [1].

In the production of electricity and heat, not the most modern technologies are used and, for various reasons, not the most economical operating modes of the equipment.

Energy can be considered the artery for most sectors of the economy, and without its development, it is difficult to imagine the development of the entire country's economy.

In recent times, there has been an energy crisis in the world and in the region, as well as in Moldova, associated with rising energy prices, as well as the risks of a shortage of energy resources in Moldova. This is largely due to external factors, as well as the lack of funding for the energy sector from the state and economic agents. The development of energy technologies largely determines the rate of economic growth and its sectoral structure in the long term, affecting the value of relative and absolute production costs in the country. At the same time, economic growth is an important factor in the dynamics of energy demand. Just four decades ago, growth in energy consumption approximately equal to the growth in GDP production was considered not only normal, but also a natural element of development. This was largely due to the low cost of oil and other energy sources. The fundamental factors for changing the situation in recent decades have been a significant increase in relative energy prices, problems associated with nuclear energy, and the concern of many countries with energy security issues; finally, the need to preserve the planet's climate. High prices and energy policy factors turned on the traditional mechanism for solving problems - technological, i.e., efficiency factor on the consumption side and the method of energy production [2].

As a result of the complex, contradictory effects of demand and prices, humanity finds itself in a transformation of the entire world energy sector. Of course, the irreversibility of investments in long-term energy assets slows down all these processes, but the flow of inventions and commercially viable investments in new technologies is becoming abundantly clear.

With all the difficulties of the current situation in the world, world GDP, according to the IMF forecast, will grow by 3.2% in 2023, which also indicates a further increase in demand for energy [3].

In the context of rising prices and demand for energy resources globally and also in the Republic of Moldova, there is a need to study the necessity of investing in Moldova's energy sector. Investments in the energy sector can significantly accelerate the development of a significant number of sectors of the nationaleconomy and lead to growth in the country's GDP.

## 2. Research methodology

In recent years, due to the risk of insufficient electricity supply and even the cessation of deliveries, coupled with the corresponding increase in energy resource prices, there is a pressing need to increase electricity production in Moldova, primarily from renewable sources. To address this, we will explore the production of electricity from renewable sources using the example of a private household, employing a heat pump for home heating. According to Law No. 10 dated 26-02-2016, promoting the use of energy from renewable sources [4]. Producers of electricity from renewable sources can employ net metering, which involves feeding the generated electricity into the grid and then using it for their own needs. In the end, payment is made only for the difference in consumed electricity. If more electricity is fed into the grid than consumed, compensation can be requested at the average market purchase price.

We will conduct the analysis on the basis of the Peresnicenco Veaceslav Ilia household at the address Alexandru Cosmescu 43/A MD-2009, Chisinau.

This household has installed photovoltaic panels on the roof of their own house and replaced the gas boiler with a heat pump. The house's total area is 230 square meters, with the heated portion (excluding the garage) being 190 square meters.

The installation of solar panels was carried out by Altenergy DC, using Longi Solar panels paired with Huawei inverters [5]. The solar panels have a total capacity of 10.8 kW, with an inverter capacity of 10 kW. The heat pump is supplied by Midea [6].

## 3. Research results

The electricity generation for the year amounted to 10,940 kWh, with an increase during the summer months and a decrease during the winter months (Table 1).

Table 1. Amount of electricity generated from renewable energy sources in KWh

Electricity generation		Nov.2022	Dec.2022	Jan.2023	Feb.2023	Mar.2023	Apr.2023	Mai.2023	Jun.2023	Jul.2023	Aug.2023	Sep.2023	Oct.2023	Total
Indicator at the beginning		43	852	992	1 247	1 733	2 506	3 313	4 582	6 071	7 137	8 612	9 899	
Indicator at the end		852	992	1 247	1 733	2 506	3 313	4 582	6 071	7 137	8 612	9 899	10 983	
Electricity generation in kilowatt-hours (KWh)		809	140	255	486	773	807	1 269	1 489	1 066	1 475	1 287	1 084	10 940

Source: compiled by the author based on data from <https://eu5.fusionsolar.huawei.com/>

From the Table 2, we can observe that the household's electricity consumption for domestic needs (including the heat pump) amounted to 6771 KWh for the year. Comparing it with the electricity generation table, we see a difference of 4,169 KWh. As there was no accumulated generation at the beginning of the period, the household had to pay for the consumption difference of 2,482 KWh from December to February. If we consider that the average electricity consumption before installing the heat pump was 2400 KWh per year, then the difference between the total consumption and the consumption before the heat pump installation, namely 4,371 KWh, represents the electricity consumption for heating the house.

Table 2. Household electricity consumption and accumulated difference in kilowatt-hours (KWh)

Electricity Consumption		Nov.2022	Dec.2022	Jan.2023	Feb.2023	Mar.2023	Apr.2023	Mai.2023	Jun.2023	Jul.2023	Aug.2023	Sep.2023	Oct.2023	Total
Indicator at the beginning.		0	668	2007	2973	4172	4784	5344	5546	5808	6076	6351	6540	
Indicator at the end		668	2007	2973	4172	4784	5344	5546	5808	6076	6351	6540	6771	
Consumption kW		668	1339	966	1199	612	560	202	262	268	275	189	231	6 771
Accumulated difference in electricity kW		141	0	0	0	161	408	1475	2702	3500	4700	5798	6651	
Electricity payable kW			1 058	711	713									2 482
Average electricity consumption without heat pump		200	200	200	200	200	200	200	200	200	200	200	200	2 400

Source: compiled by the author based on data from <https://eu5.fusionsolar.huawei.com/>,  
<https://premierenergy.md/>

Analyzing the diagram (Fig. 1.), we observe that the monetary consumption of gas significantly exceeds the consumption of electricity after the investment in the heat pump. This indicates that the savings from gas usage outweigh the increased electricity consumption.

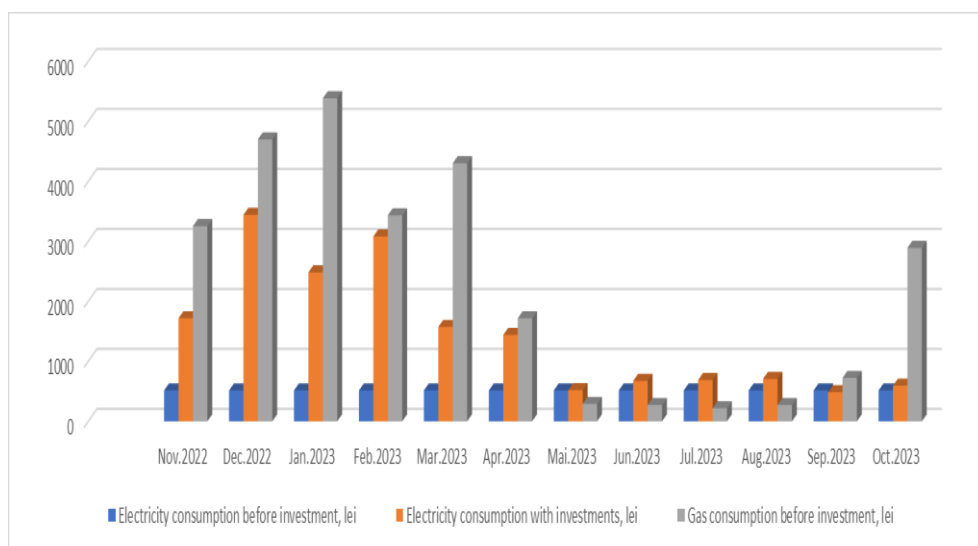


Fig. 1. Electricity consumption and generation by month, in kilowatt-hours (KWh)

Source: compiled by the author based on data from <https://eu5.fusionsolar.huawei.com/>,  
<https://premierenergy.md/>

According to the Table 3, we can see that the annual gas consumption amounted to 1519 cubic meters, and in monetary terms, at the average gas price at the end of 2023 of 0.924 euros per cubic meter, it amounted to 1404 euro.

Table 3. Amount of gas consumed by the household for a similar period last year

Gas consumption	unit	Nov	Dec	Jan	Feb	Mar	Apr	Mai	Jun	Jul	Aug	Sep.	Oct	Total
Gas consumption	cubic meter	180	260	298	190	238	95	16	15	12	15	40	160	1519
	Euro	166	240	275	176	220	88	15	14	11	14	37	148	1 404

Source: compiled by the author based on data from <https://my.moldovagaz.md/>

According to the diagram (Fig. 2.), we can observe that electricity generation almost always exceeds electricity consumption in kilowatt-hours (KWh), except during the winter months. Therefore, based on net metering, the household accumulates the generated electricity fed into the grid for later consumption during the winter months.

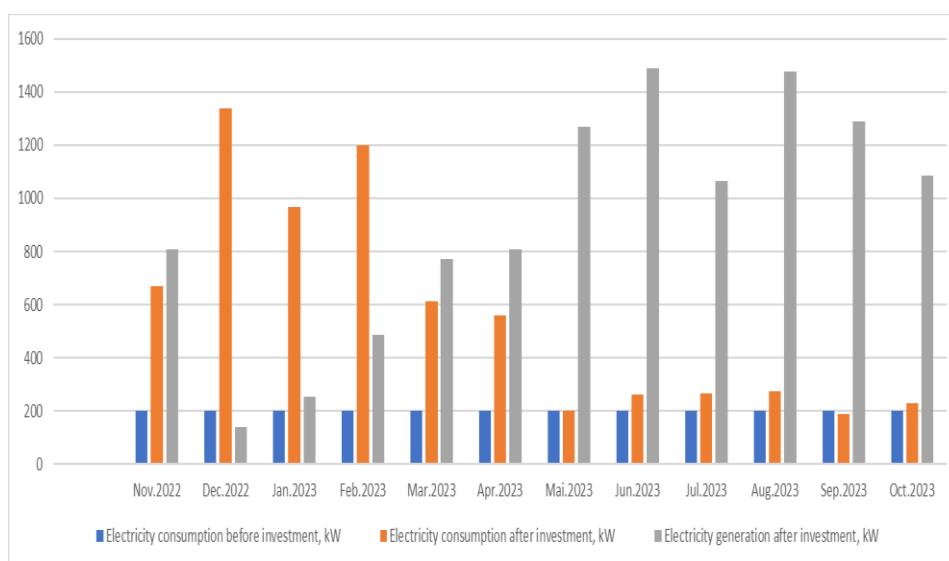


Fig. 2. Electricity and gas consumption before and after investments.

Source: compiled by the author based on data from <https://eu5.fusionsolar.huawei.com/>,  
<https://premierenergy.md/>

From the Table 4, we can see that selling all the generated electricity to the grid at the average market purchase price of 0,088 euros would result in a payback period of 10,33 years for the 10,000 euro investment in solar panels.

Analyzing Table 5, we have concluded that if you sell all the generated electricity to the grid and then retrieve electricity from the grid for your own needs (which is the electricity consumed before the installation of the heat pump), you not only earn income from selling electricity but also benefit from the price difference between the purchased electricity and the electricity obtained through net metering. So, we sell electricity to the grid at 0,088 euros per kWh [9], but purchase it at 0,131 euros per kWh [10]. As a result, with an annual consumption of 2400 kWh, we save 315 euros. Additionally, we earn 756 euros per year from the difference in the generated electricity. The payback period under such circumstances would be 9,34 years

Table 4. Calculation of the payback period for solar panels, assuming the sale of excess electricity to the grid

Indicators		1 year	2 year	3 year	4 year	5 year	6 year	7 year	8 year	9 year	10 year	11 year
Investing in Solar Panels,eur	10 000											
KW output		10 940	10 940	10 940	10 940	10 940	10 940	10 940	10 940	10 940	10 940	10 940
Sales to the electricity network,eur		968	968	968	968	968	968	968	968	968	968	968
Cumulative sales to the network,eur		968	1 936	2 904	3 872	4 840	5 808	6 776	7 745	8 713	9 681	10 649
Payback, years												10,33

Source: compiled by the author based on data from <https://eu5.fusionsolar.huawei.com/>,  
<https://premierenergy.md/>

Table 5. Calculation of the payback period for solar panels, considering the sale of the difference between previously consumed electricity and generated electricity

Indicators		1 year	2 year	3 year	4 year	5 year	6 year	7 year	8 year	9 year	10 year
Investing in Solar Panels,eur	10 000										
KW output		10 940	10 940	10 940	10 940	10 940	10 940	10 940	10 940	10 940	10 940
Electricity consumption before investment,kW		2 400	2 400	2 400	2 400	2 400	2 400	2 400	2 400	2 400	2 400
Sale of electricity difference KW to the grid		8 540	8 540	8 540	8 540	8 540	8 540	8 540	8 540	8 540	8 540
Sale of electricity difference to the network,eur		756	756	756	756	756	756	756	756	756	756
Energy savings,eur		315	315	315	315	315	315	315	315	315	315
Cumulatively selling online and saving,eur		1 071	2 142	3 214	4 285	5 356	6 427	7 498	8 569	9 641	10 712
Payback, years											9,34

Source: compiled by the author based on data from <https://eu5.fusionsolar.huawei.com/>,  
<https://premierenergy.md/>

This calculation illustrates (Table 6) what the household ultimately achieved. It made investments of 20,000 euros by installing solar panels and replacing the gas boiler with a heat pump. In the end, a portion of the generated electricity, through net metering, is used for internal consumption, mainly for heating. The surplus from the generation is compensated by the distribution grids, and the household also benefits from gas savings due to its absence. In the end, this household saves 1404 euros annually on gas, based on the gas price of 0,924 euros per cubic meter. It also saves on previously consumed electricity, with an annual consumption of 2400 kWh at a rate of 0,131 euros per kWh. Additionally, the household generates additional income from selling the surplus electricity to the grid, with 4169 kWh at an average market purchase price of 0,088 euros per kWh. The overall payback period is 9,58 years. It should be noted that at the time of the investment, the gas price for consumers was one-third higher, and the electricity price was also higher, potentially reducing the payback period compared to our calculations.



Table 6. Calculation of the payback period for solar panels, considering the sale of the difference between consumed electricity, generated electricity, and gas savings

Indicators		1 year	2 year	3 year	4 year	5 year	6 year	7 year	8 year	9 year	10 year
Investment in solar panels+heat pump,eur	20 000										
KW output		10 940	10 940	10 940	10 940	10 940	10 940	10 940	10 940	10 940	10 940
Electricity consumption before investment, kW		2 400	2 400	2 400	2 400	2 400	2 400	2 400	2 400	2 400	2 400
Electricity consumption after heating investments, kW		6 771	6 771	6 771	6 771	6 771	6 771	6 771	6 771	6 771	6 771
Sale of electricity difference KW to the grid		4 169	4 169	4 169	4 169	4 169	4 169	4 169	4 169	4 169	4 169
Sale of electricity difference to the network ,eur		369	369	369	369	369	369	369	369	369	369
Energy savings,eur		315	315	315	315	315	315	315	315	315	315
Gas saving,eur		1 404	1 404	1 404	1 404	1 404	1 404	1 404	1 404	1 404	1 404
Cumulatively selling online and saving, eur		2 088	4 176	6 265	8 353	10 441	12 529	14 618	16 706	18 794	20 882
Payback, years											9,58

Source: compiled by the author based on data from <https://eu5.fusionsolar.huawei.com/>,  
<https://premierenergy.md/>

Analyzing the Table 7, we observe that in Moldova as of the beginning of 2023, there were 263,390 private houses [12] with four or more rooms, which is closely comparable to the household in our example. If this method of investing in solar panels and heat pumps is applied to 10% of such households, the country could reduce its electricity consumption by 173 million kWh and gas consumption by 40 million cubic meters annually. In monetary terms, this would amount to over 55 million euros, which holds significant importance for a small republic like Moldova.

Table 7. Application of investment in solar panels and heat pumps to households in the Republic of Moldova with four or more rooms

Indicators	4 rooms and more
Individual houses, units	263 390
% of investment application	10%
Calculation of gas savings cubic m	40 008 941
Calculation of gas savings thousand euros	36 975,9
Calculation of energy savings thousand kW	63 214
Calculation of energy savings thousand euros	8 309,7
Sale of excess electricity to the grid, thousand kW	109 807
Sale of excess electricity to the network, thousand euros	9 717
<b>Total benefit thousand euros</b>	<b>55 002,3</b>

Source: compiled by the author based on data from <https://eu5.fusionsolar.huawei.com/>,  
<https://premierenergy.md/statistica.md>

## 4. Conclusion

Investments are a crucial element of the economic development of a country. In this context, the economic policy of the Republic of Moldova for the medium and long terms should be oriented towards fostering investment growth and maintaining it at a high level. This is because investments determine the production capacities of the country in the future and contribute to its economic growth potential. In this regard, countries compete internationally for the influx of investments. Investments by individuals in solar panels and the installation of heat pumps are a profitable decision, as they pay off within a 10-year period, exceeding the equipment's average lifespan, making it a sufficiently long-term choice. Considering the possibility of selling the gas boiler after dismantling, the payback period can be further reduced. There is also a possibility of further reduction in the prices of gas and electricity, while no decrease in investments is observed. This suggests that the payback period may have been extended. Moreover, investments by individuals contribute to reducing the purchase of electricity from external markets and also decrease gas purchases. Moldova, as known, does not possess its own gas reserves and produces electricity insignificantly, so reducing dependence on external suppliers enhances the energy security of the country. As it is known, the government does not provide grants and assistance for the purchase of equipment for individuals, but only for legal entities. Therefore, with a further decrease in the prices of energy resources, investments from private households may be reduced. We believe that the government should pay attention to this issue and develop programs to finance such investments.

Support measures from the government may include public-private partnerships, providing government loans and guarantees, compensating for investment expenses, granting subsidies, reducing tariffs, eliminating bureaucratic barriers, streamlining administrative procedures for investors, easing taxation, regulating, and committing to purchasing generated capacities at fixed prices, simplifying investment procedures through the Internet, and supporting e-commerce. Moreover, the increase in investments from individuals in solar panels and heat pump installations contributes to the development of new companies involved in the acquisition and installation of such equipment. Additionally, there is the possibility of partial or complete assembly of such equipment, which further promotes the economic growth of the state.

## References

1. Bikova E.V. (2013) National report "Problems and prospects for creating a favorable climate for increasing energy efficiency and energy saving in Moldova" (in Russian), [https://unece.org/fileadmin/DAM/energy/se/pdfs/ee21/EE21\\_Subregional\\_projects/MoldovaBicova-Rus02.pdf](https://unece.org/fileadmin/DAM/energy/se/pdfs/ee21/EE21_Subregional_projects/MoldovaBicova-Rus02.pdf)
2. Grigoriev L.M., Kudrin A.A. (2013). Economic growth and energy demand. Economic journal HSE, №3, 2013, (in Russian). <https://cyberleninka.ru/article/n/ekonomicheskii-rost-i-spros-na-energiyu/viewer>
3. Korovleva N. (2022). The IMF has lowered its forecast for global GDP growth to 3.2% in 2022. Economics and business / Global topics, Deutsche Welle Magazine, 26.07.2022, (in Russian) <https://www.dw.com/ru/mvf-ponizil-prognoz-rosta-mirovogo-vvp-do-32-v-2022-godu/a-62604828>
4. LAW No. 10 of 26.02.2016 on advancing the use of energy from renewable sources, (in Russian) [https://www.legis.md/cautare/getResults?doc\\_id=140494&lang=ru](https://www.legis.md/cautare/getResults?doc_id=140494&lang=ru)
5. LONGi (2023) Vertically Integrated Solar PV Value Chain. <https://www.longi.com/us/>
6. Romstal (2023) <https://romstal.md/ru/pompa-de-caldura-aerapa-incalzire-racire-split-ecoheat-by-midea-12kw-380v>
7. Huawei (2023) <https://eu5.fusionsolar.huawei.com/unisso/login.action#/view/stationVNE=35570544/overview>

8. Premierenergy (2023) <https://premierenergy.md/ru/optiuni-de-plata-ru/>
9. National Energy Regulatory Agency (2022) Raport privind activitatea Agenției Naționale pentru Reglementare în Energetică în anul 2022. <https://anre.md/storage/upload/administration/reports/1138/Raport%20privind%20Activitatea%20ANRE%20in%20anul%202022.pdf>
10. <https://tv8.md/ru/2023/05/07/v-moldove-podeshevet-elektroenergiya-nare-utverdilo-novie-tarifi/233749>
11. <https://my.moldovagaz.md/my-history>
12. National Bureau of Statistics. *I Fond\_locativ 2023* [online]. Available at <[https://statistica.gov.md/files/files/serii\\_de\\_timp/Fond\\_locativ/Fondul\\_locativ\\_1\\_ianuarie\\_2023.xls](https://statistica.gov.md/files/files/serii_de_timp/Fond_locativ/Fondul_locativ_1_ianuarie_2023.xls)>

x

## **Capital Structure of the Enterprise /Methodology and Methods for Analysis in the Context of Digitalisation/**

**Rositsa Ivanova**

University of National and World Economy, Sofia, Bulgaria

*Rlvanova@unwe.bg*

**Abstract.** Digital transformation is a strategic process where various digital tools and technologies are used, such as scanning, computer networks, internet, satellite, social media, etc. This enables the processing, storage and transfer of information to users, irrespective of their location. Digitalisation gives enterprises the opportunity to create new and to build-up on existing processes. Nowadays, digitalisation becomes a key priority for many enterprises, since by digitalizing the business processes in the course of their operations, enterprises have the objective opportunity to overcome the dependence on their physical location, the remoteness of suppliers, customers and markets, the existence of appropriate infrastructure, and a number of other factors that affect their business. Thus, enterprises may implement timely and adequate measures upon the occurrence of any adverse trends in the development of the dynamic and competitive market, which is characterized by a certain level of entropy. The main body of this publication explains the nature of digitalisation, the advantages of its introduction and the requirements in the process of enterprises' digitalisation. Based on this, opportunities for digitalisation of activities in the field of accounting and financial and business analysis of the enterprises are offered. The object of this study refers to the opportunities for digitalisation of the financial and business analysis of the enterprises' operations, and its subject matter covers the digitalisation of the analysis of the capital structure of non-financial enterprises in the economy of our country based on the data of their financial statements. The publication highlights the opportunities to link the information contained in the enterprise's balance sheet to the methodology and the methods for analysis of the enterprise's capital structure. The aim is to achieve connectedness – within a single system, among information provision, methodology and methods for analysis of the enterprise's capital structure, the analysis results and the adoption of justified and good decisions for its management. Within this system, data from the business analysis of development, the condition and the trends in the capital structure, with view of, and in relation to, the financial and investment policy of the enterprise, could be successfully integrated and used.

**Key words:** Digitalisation, Financial and business analysis, Capital structure, Methodology, Analysis, Decisions

### **1. Introduction**

Digital transformation refers to the integration of new digital technologies in all spheres of business life, thus changing the way enterprises from all areas and their related sectors in the national economy work. Long-term perspectives for their development are created, their competitive power and better positioning on the dynamic and competitive power are enhanced. The increase of the amount of information that enterprises receive (created either in or out of the enterprise), process, use and store in the course of their business, results in increase of the

responsibility for its management. In terms of digital transformation, enterprises need to comply with a number of requirements, the more important ones being:

1) increase of investments for improvement of digital knowledge and for digitalisation of business processes;

2) establishment of flexible, well-protected and secure infrastructure, for example, automation of processes related to enterprise's management, development of cloud platforms, optimization of costs for their support, etc.

3) use of cloud platforms for improvement of the product and service quality and competitive power, thus enhancing the effectiveness of business processes, increasing the profit from the overall business of the enterprises, leading to improvement of their profitability too;

4) automation of processes for making management decisions for the development of the enterprises and their business, both in operational and strategic aspect, which are based on objective evidence and not on suggestions. We support the opinion of Lazarova (2020, page 43, 51) that "leaders (managers and owners of enterprises – added by R.I.) should have to understand the strategic value of digital transformation", and that "the digital transformation requires to understand the need of investments, risk management and introduction of the relevant changes in order to gain greater benefit from the change."

## **2. Digitalisation**

The term "digitalisation" originates from English – "digital", which means digital/numeric. Digitalisation is a process of information transformation from analogue carrier (existing in different forms, such as text, pictures, voice, etc.) into digital form by means of electronic devices (scanners, cameras, etc.), so that the information can be processed, stored and transferred through digital circuits, devices and networks. Lazarova (2019, page 98) is at the opinion that "digitalisation is a set of actions for transformation of analogue information resources in digital copies and their storage, access, use, dissemination and security." Digitalisation is a symbiosis between the use of information and technologies for the purposes of increasing the productivity of human resources, which fundamentally determines the digital transformation. This allows to shift the focus from the physical assets traditionally held by the enterprises to the optimization of the way in which information, being considered as an intangible asset, is processed, analysed and shared along the entire value chain during the entire lifecycle of every individual product. Business digitalisation ensures vast opportunities to easily generate, collect, process and share the huge amount of data created by employees, clients, suppliers, processes, products, competitors, other enterprises, government authorities and institutions, bank system, etc., among individual enterprises, geographic areas and system domains. The generation of such big databases requires effective management, processing, storage, information security and considerable information and technological resources. From this point of view, cloud-based services should be applied in the enterprises' business.

As a strategy, digitalisation refers to determination of the direction for the enterprise's development, the arrangement of resources in specific order, and the assumption of the relevant obligations. We are strongly convinced in our opinion that the digital strategy should be integrated in the business strategy of each enterprise.

Cloud technology is a model that provides opportunities for network access to shared resources, such as internet networks, servers, data arrays depositories and various software applications with minimum involvement or management on the side of the service provider. Naturally, the cloud enables users (enterprises) to use the services so provided in their operations, thus saving financial resources to be invested in hardware and software.

Cloud technologies provide users with services, according to their needs, as well as services on demand. They can be used in the form of three types of models: public, private and hybrid cloud.

In its nature, the public cloud is an established cloud infrastructure prepared for a wide range of users in return of payment to this type of service provider. In contrast, the private cloud's infrastructure is provided to the users within a private organization and may be used only by them for specific purposes. The private cloud infrastructure is not shared with other organisations, i.e. with other users. The advantage of the private cloud is that it offers cloud environment with enhanced level of access control and improved information security. The hybrid cloud model allows users to store their most important applications and information arrays that require higher access security, confidentiality and reliable safety in their private cloud, and in the public cloud they could store those of their applications and data that are not subject to increased requirements in terms of their access and security.

Cloud services may be accessed via three models: IaaS (infrastructure as service); PaaS (platform as service); SaaS (software as service).

With IaaS model, users are mainly given server and network resource, while being responsible for ensuring the processes for operating system installation, data storage and backup. The model for provision of PaaS services generally refers to lease of hardware, operating systems, storage and network capacity via internet. It gives users the opportunity to rent virtualized servers and their related services to run existing or to develop and test new applications. As far as the SaaS model is concerned, services are provided by service providers and the users use these services to run applications in the cloud infrastructure. These applications may be accessed via web-based browsers. To our opinion, the appropriate model for the purposes of the enterprise's financial and business analysis is the SaaS model for running a cloud platform for analysis of the overall business of the enterprise within the mutual relations and dependences among the indicators within the system of indicators that are typical for its different aspects.

### **3. Opportunities for digitalisation of financial and business analysis**

Different users of the information in the enterprises' financial statements need information about the revenue, expenses, financial performance, financial position, financial stability, profitability, cash flows, etc. In the context of the modern environment in which enterprises carry out their business, the opportunities for faster introduction of new technological solutions significantly increase. Against this background, the opportunities for establishment of digital platform for financial and business analysis of enterprises also increase. The digitalization of the financial and business analysis makes it possible to extract, study and transform huge amount of information (big data), which the enterprise has available, processes, uses and stores as meaningful information, which is targeted, important and useful for the design and adoption of proper, justified and effective management decisions for its business and its development. We believe that it is absolutely possible to build-up and integrate a digital platform for financial and business analysis in a cloud environment for real-time digital analysis.

To our opinion, the cloud platform for financial and business analysis can be presented as individual modules that describe the methodologies and methods for analysis by separate items of the analysis content (see figure 1). For example, this is about the methodologies and methods for analysis of revenue, expenses, cost, provision, use and effectiveness of use of resources (fixed tangible and intangible assets, material resources and workforce), financial

performance, financial position, financial stability, profitability, cash flows and other items of the analysis of the enterprise's business.

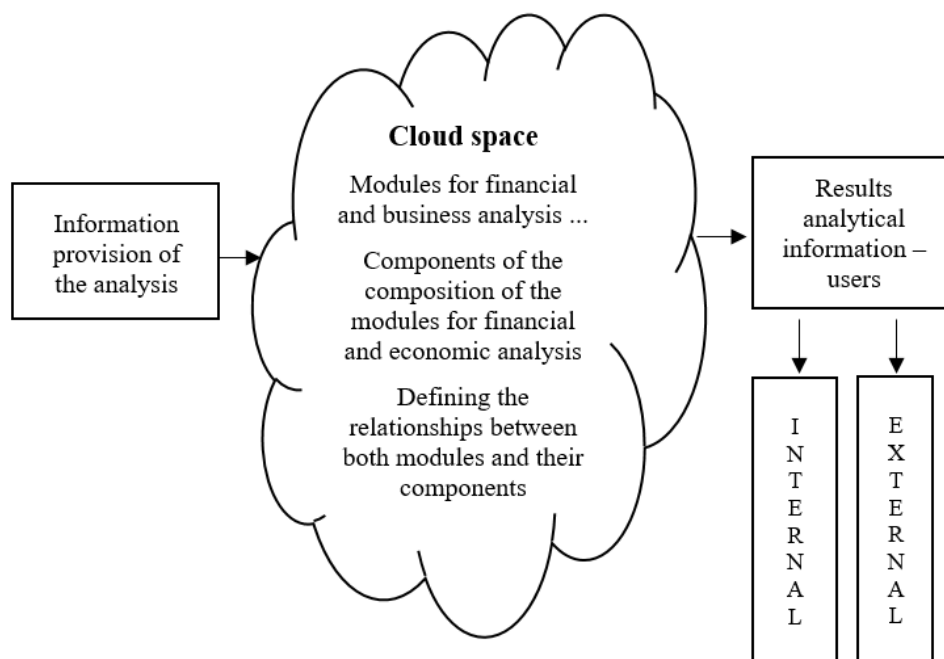


Fig. 1. Model of cloud technology for financial and business analysis - systems approach

We believe that the cloud platform for financial and business analysis of the enterprise can be defined as a system (sub-system within the system of the cloud space). To the opinion of Petrova (2018, page 138), the platform is based on “the traditional software products, which (the cloud platform – note of R.I.) is accessible via internet.” With view of the system approach, we can reasonably raise the question for the elements of the financial and business analysis system within the cloud space.

We think that the first element of this system is the input information mainly based on data from the company's business, i.e. this is the information provision of the analysis of the enterprise's overall business. Electronically, this information is transmitted to the cloud platform for financial and business analysis, where it can be processed and stored.

At least to our opinion, the second element of the cloud platform for financial and business analysis are the software tools (software), which are combined, unified, put under analysis of their effective use, and with the help of which it is possible to visualize the resultative analytical information. To our view, the software in the cloud platform for financial and business analysis may be based on a system of software solutions by individual modules of the common methodologies and methods of the analysis, some of which have been mentioned above. In real time, the cloud platform for financial and business analysis could visualize the relevant tables, which in general represent structured data reflecting individual facts and circumstances from the relevant field, and in this particular case – the specific items of the financial and business analysis content. These data are arranged in particular way and are organized with the aim to allow their processing and storage.

The third element of the cloud platform for financial and business analysis should be the output information. This is the resultative analytical information on the basis of which proper and justified management decisions could be designed and made. We should note that enterprises do not meaningfully use the huge amount of information that can provide them with crucial and useful data for the internal business processes, clients, suppliers, etc. And here, we can see an interesting fact. Within the huge information environment in which business is ran (big data operations), the enterprise's managers may overlook some information they do not know to exist, which can be extremely important and useful for the management. Actually, when making their decisions, the managers are led only by the information they are aware of. We believe that by using digital technologies and cloud solutions where analysis is made in real time, the issue with the information scope might be minimized and the awareness in the course of decision-making for the development of the enterprise and its business, from operational and strategic aspect, might be raised.

The greatest benefit of the cloud platform for financial and business analysis is the easy and quick access to the resultative analytical information in any part of the world, at any time and from any device. For this, only available internet is required.

#### **4. Possibilities for cloud platform for analysis of enterprise's capital structure**

The module for analysis of the capital structure within the system of modules for financial and business analysis of the enterprise's business in the cloud platform may be presented on figure 2.

The entire movement of information (input information, its processing, storage and obtaining output information) for the enterprise's capital structure (formation, change and management) is performed in virtual environment.

To our opinion (2019, page 379-387), the software module for analysis, including of the enterprise's capital structure, may comprise the following components based on the system of absolute and relative indicators and on the methodology and methods for analysis :

- 1) analysis of the overall capital structure;
- 2) analysis and assessment of the amount, composition, structure and dynamics of enterprise's equity;
- 3) analysis and assessment of the amount, composition, structure and dynamics of the enterprise's borrowings (liabilities);
- 4) models for analysis of capital turnover – of the entire invested capital, of equity, borrowings and fixed capital;
- 5) models for analysis of the gross and net profitability of the entire invested capital;
- 6) models for analysis of equity profitability. Analysis of the leverage effect;
- 7) models for analysis of fixed capital profitability.

We are at the opinion that comparability must be achieved between the input information that is necessary for the analysis of the capital structure, the methodology and methods for analysis, the resultative analytical information and the users (internal and external) of this information. For this purpose, software can be used, by applying the methodology and methods for analysis and by generating the resultative analytical information automatically and based on the balance sheet data. The quality and usefulness of the information generated as a result of the analysis depend on the quality of the input information – the accounting information for the capital created in the enterprise – both in general and by types of capital,



as well as its disclosure in the balance sheet and the explanatory notes to the annual financial statements of the enterprise.

We consider the opinion of Natchkova (2018, page 16) that “the true and fair presentation and disclosure of (information for equity and borrowings in the enterprise’s financial statements – added by R.I.) will improve the quality of the input information in the cloud platform for analysis of the capital structure, and on the basis of the obtained resultative analytical information, its effective management and the establishment of good capital structure may be achieved, which suggests that the required yield for the capital owners and positive leverage effect have been reached” appropriate.

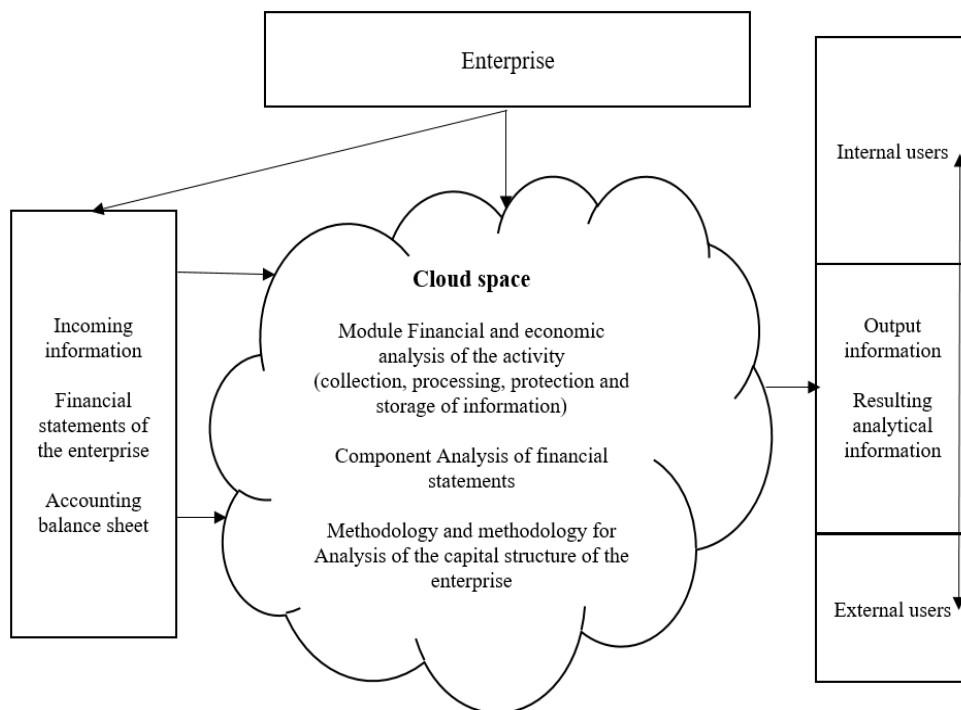


Fig. 2. Cloud platform model for capital structure analysis – systems approach

To our view, based on the above statements, we can come to the conclusion that the cloud application (platform) for analysis of the capital structure should be a compulsory element of the cloud platform for comprehensive financial and business analysis of the enterprise’s business.

## 5. Conclusion

A number of issues occur in the process of digitalization of accounting and financial and business analysis. The first of them refers to the way of accounting of digital technologies. We share the opinion of Petrova (2018, page 137) that “the organization of accounting of innovative (digital) technologies will be peculiar. As a product of intellectual work, they will definitely be classified in the group of intangible assets. The lack of adequate regulations and practices might cause some difficulties when drafting (designing – note of R.I) the accounting

policy (of the enterprise – note of R.I.) with regard to the recognition, measurement, and disclosure of the information about them in the financial statements.” Furthermore, we think that the accounting recognition and reporting of digital technologies is a challenge for the theory of the financial and business analysis too, which should design methods and methodology for their analysis that are practically feasible.

The issue of compatibility that should be achieved during the digitalization and the use of cloud services in business comes second. To the opinion of Lazarova (2020, page 11-12) some level of compatibility must be achieved “between the various information systems used for the relationships among the intracompany departments and between the company and the outer world.”

Thirdly, we can consider the issue of the economic benefit from the introduction of cloud solutions into the business, including into the accounting and the financial and business analysis. In order to analyse and assess this benefit, we can use a set of indicators that characterize the benefits, costs, risks, and future opportunities of the enterprise generated as a result of the investments made for business digital transformation (2011, page 3), including cloud solutions. The economic effect of the use of cloud technologies in business may be analysed through a system of indicators: total profit; total expenses; net cash flow; cumulative cash flow; net present value; return on investments (ROI); time period for repayment; number of users of cloud software. We fully agree with Lazarova (2020, page 45) that “the traditional key indicators for effectiveness are not sufficient and effective for the measurement of the effectiveness of the digital transformation” of business.

Despite the above issues, and probably the existence of any other similar and currently unsolved issues, the advantages of cloud platforms for users are significant. They may be summarized as follows:

- 1) they provide users with significant reserve of resources, whereas they (the users) pay only for the resources that have been used. Thus, optimization of expenses on materials, consumables, courier services and other costs is achieved, according to the enterprise’s needs in the cloud platform for the time being;
- 2) high level of hardware use, overall automation of activities while carrying out the financial and business analysis, and flexible use of the cloud platform for analysis of the business are achieved;
- 3) operational and continuous online access to the information base of the financial and business analysis is ensured;
- 4) monitoring by separate items of the financial and business analysis content is ensured, which is convenient for the management;
- 5) maintenance of convenient electronic records of documents and schedules by separate analysis items is offered;
- 6) high level of security of both the input information for the financial and business analysis system and of the resultative information generated by the analysis is achieved;
- 7) opportunities for creation of directories of enterprise’s clients and the related receivables, as well as of its suppliers and other creditors and the related payables are provided in the cloud platform. This is a way to optimize the opportunities for analysis of the capital structure, financial stability, solvency and liquidity of the enterprise. Issues that are crucial at all stages and phases of the enterprise’s development;
- 8) the use of cloud technology facilitates the achievement of business targets and strategy of the enterprise.

The advantages of cloud technologies require some proposals and solutions, and namely:

- 1) the number of enterprises, which optimize their operations and business by using cloud technologies is increases, i.e. we see a growth rate of the use of cloud services ;
- 2) the size of the data storage spaces is increased. This suggests and requires the cloud service providers to offer more data centers;
- 3) constant improvement of the security measures for personal data protection is necessary. Appropriate and effective security measures are required for the storage and transfer of data in cloud infrastructures, irrespective of the cloud type and the cloud service model;
- 4) creation of new clouds, such as: public cloud, combined cloud; internal cloud, etc. the public cloud may be defined as infrastructure shared by several enterprises, which form a community sharing common and naturally close interests, for example, information security, terms and conditions for use of cloud service, compatibility requirements, etc. Combined cloud refers to combined cloud technology established by joining a few clouds. The internal cloud, also known as cloud in clouds, is a network of cloud services established on the basis of standards and providing specific environment of cloud technology.

## References

1. Lazarova, V., Digitization and digital transformation in accounting, Sofia, magazine "Economic and Social Alternatives", UNSS, issue 2, 2019.
2. Lazarova, V., Digitization in accounting, Sofia, Avangard prima, 2020.
3. Nachkova, M., Specific objects of the financial audit in the contractual funds, Sofia, IDES Yearbook, 2018.
4. Petrova, P., Industry 4.0 and accounting: challenges and opportunities, Sofia, Scientific works of UNSS, item 3, 2018.
5. Chukov, K., R. Ivanova, Financial and economic analysis, Sofia, IC-UNSS, 2019.
6. Chukov, K., R. Ivanova, Financial and economic analysis, Sofia, IC-UNSS, 2022.
7. Herbert, L., Erickson, J., The ROI of cloud apps (for Sourcing & Vendor Management Professionals). June 23, 2011. [http://resources.idgenterprise.com/original/AST-0042573\\_Forrester\\_Report\\_The\\_ROI\\_of\\_Cloud\\_Apps.pdf](http://resources.idgenterprise.com/original/AST-0042573_Forrester_Report_The_ROI_of_Cloud_Apps.pdf)

## Specific Features of the Accounting of Insurers' Expenses

Maia Natchkova

University of National and World Economy, Sofia, Bulgaria  
*mnachkova@unwe.bg*

**Abstract.** This publication highlights the specific features of the accounting of insurer's expenses. Insurance expenses are specific depending on the type of insurance carried out - direct insurance, active or passive reinsurance. Special focus is also put on the types of the expenses that insurers establish, depending on the type of the insurance license they are holders of, and on the specific accounts to which established insurance expenses are booked on. The objective is to present, justify and outline the specific characteristics of the accounting of insurer's expenses in terms of content, type and ways of occurrence. Information on insurance and other expenses received during the year must be disclosed in the insurers' financial statements truthfully and accurately, in all material respects, in compliance with accounting rules, principles, concepts and conventions.

**Key words:** Insurers, expenses, accounting, financial statements.

### 1. Description of insurers' expenses

*A definition of expenses*, as economic category, may be found in the general framework of IFRS/ IAS. In the context of the recitals to IFRS and the general requirements to the preparation and presentation of financial assets, expenses are defined as decreases in enterprise's economic benefits during the accounting period in the form of outflows or depletions of resources (assets) or incurrences of liabilities that result in decreases in equity, other than those relating to distributions to equity participants. Spending expenses is not always related to a specific cash flow movement and making payments. However, in some business operations, specific payments or extra charges may be made, which cannot be classified as expenses. In such cases, no expenses are accounted for in the accounting books, but acquisition of assets the use of which is expected to bring future economic benefit; or loss of assets; or distributions among owners (for example, dividends). Thus, insurers' expenses may be identified. They are incurred in the ordinary course of insurance business, which is part of the financial services offered by financial enterprises. Insurers carry out their ordinary activity in a high-risk environment mainly due to the occurrence of insurance and financial risks. Therefore, the insurance business – which is strictly regulated as part of the financial services offered by the financial enterprises (Code of Insurance, (2023), paragraph 9, and supplementary provision, paragraph 1, item 43) – is subject to licensing by the state supervisory authority – Insurance Supervision Division of the Financial Supervision Commission (FSC). Expenses directly related to the insurers' licenses issued by FSC – for non-life and for life insurance, or for reinsurance.

*Insurers' expenses are recognised* in strict observance of the accounting principles of compliance (comparability of income and expenses), of accrual accounting, of prudence (conservatism), and of the "going concern", and in strict observance of all accounting rules, principles, concepts, conventions, and usual regulations and standards, as the main objective of any going concern is to generate income while incurring expenses.

*Insurers' expenses are classified* on the basis of criteria related to their economic nature, to the time of being recognized in the accounting books, to their connection with the ordinary business or to their functional purpose. The main criterion for classification of expenses is the nature of the business carried out by the insurers. According to this criterion, expenses are classified as:

1) Expenses that are directly related to insurance business – these are the expenses incurred in relation to the conclusion, administration, maintenance, renewal and termination of the insurance contracts and to the formation of insurance (technical provisions), which refer to the insurers' main obligations under the concluded insurance contracts. Insurers generate significant cash flows in insurance monetary funds that they manage and distribute. From these funds they pay compensations or amounts and ensure coverage of the insurance risks upon occurrence of events or realization of conditions as set out in the insurance contracts or the applicable law (Code of Insurance, Green Card International Agreement, etc.). Namely these statutory obligations of the insurers result in the formation of expenses due to their insurance business. In insurance, these expenses are called insurance costs. Depending on the type of insurance and according to the level of insurance risk and the liabilities assumed by the insurers, their expenses may be:

- passive reinsurance expenses – these are the expenses of assigners (direct insurers who have reinsured themselves at reinsurers). They have transferred (assigned) a portion of insurance premiums collected by them upon transferring the insurance risk to reinsurers and have become assigners that perform passive reinsurers;

- active reinsurance expenses – these are the expenses of reinsurers who have assumed the insurance risk transferred by the assigners and have received from them the assigned reinsurance premiums. They perform active reinsurance.

2) Expenses that are indirectly related to the insurance business – these are the expenses incurred by the insurers in relation to the support of their companies, the so called administrative expenses:

- expenses by economic items (expenses on materials, on hired services, amortization of fixed tangible and intangible amortizable assets, on salaries and social security of employees, on local taxes and charges, and other expenses related to their administrative activity);

- expenses by functional purpose (by types of activities) to which expenses by economic items relate and are comprised in.

3) Expenses related to the insurers' investment activity – these are the financial expenses. Insurers are institutional investors and the legislator obliges them to maintain their solvency, yield, liquidity and financial stability. Therefore, they make transactions on primary stock exchange markets with financial assets and financial instruments in which they invest their own funds (assets), generated from their equity and allocated insurance (technical) provisions under the insurance contracts they have entered into with insured persons and insuring parties. These transactions may give rise to financial expenses as a result of purchase and sale of financial assets and financial instruments, and of impairment losses upon change of their fair values, of interests, of broker fees and bank charges, of charged dividends for share holdings, etc. Insurers must account for their financial assets and financial liabilities at fair value, which is “related to continuous recognition of gains and losses from changes in the assets/ liabilities rates.” (Daskalov, 2020). Insurers must also disclose the expected credit losses. “With the adoption of IFRS 9 Financial Instruments... a new approach to the classification of financial assets has been introduced, by putting a strong focus on the contemporary model for recognition of expected credit losses. The introduced new accounting

framework for recognition of credit risk impairment is a key tool that directly affect the levels of the regulatory capital...” (Feschian, Andasarova, 2020)

## 2. Specificity of accounting of insurers’ expenses

The specificities of the accounting of insurers’ expenses refer to their content, types, measurement and recognition, periods of occurrence, methods, technique and methodology of accounting, presentation and disclosure in the insurers’ financial statements, and namely :

*1) Specific content and type of expenses* – they are defined by the legislator and the insurance regulatory authority and refer to the two insurance sectors – life insurance and non-life insurance. (Code of Insurance, 2023; Directive 2009/138/EC of the European Parliament and of the Council (Solvency II) and its provisions, 2009, and Ordinance No 53 of FSC, 2016). In terms of their content, insurers’ expenses are insurance expenses, related to their principal insurance business; investment expenses incurred as a result of the insurers’ business as institutional investors on the stock and money markets, and other administrative expenses of the insurers related to the held fixed tangible and intangible assets, investment properties, materials, employees, and those related to the used hired services, local taxes and charges due, secondments and staff training, etc. The specific **acquisition costs** exist only in insurance as a business. These are the expenses incurred by the insurers to cover the insurance risk and refer to the initial conclusion, subsequent overall administration, termination and possible renewal of insurance contracts. Acquisition costs are of biggest amount in comparison to the other expenses and are of greatest importance for the performance of the insurers’ ordinary activity. Most of these costs are related to the settlement of claims and payment of insurance compensations, or amounts to insurance policy holders or to third parties – beneficiaries of the insurances, upon liquidation of damages due to occurred insured events in non-life insurance, or upon entering of insurance contracts in life insurance due to the survival of the end of the contract. Furthermore, these are costs related to the definition of the amount, allocation, release, and accounting of insurance (technical) provisions.

*2) Basis for recognition of expenses* in the accounting books and the financial statements of insurers. Insurers’ expenses are recognized with view of the application of IFRS 17 Insurance Contracts. This is a new standard endorsed by the EU with Commission Regulation (EU) 2021/2036 of 19 November 2021, however its application in Bulgaria was delayed by 2023. Therefore, the first financial statements of insurers that already apply IFRS 17 are for the 2023 accounting year. This standard applies to specific types of insurance contracts, and not for a specific type of insurance company. The expenses incurred for the conclusion, maintenance and termination of insurance contracts are referred to, recognized and accounted for namely in relation to the types of insurance contracts covered by IFRS 17. Expenses are recognized in the Statement of Profits and Losses and Other Comprehensive Income of insurers together with the recognition of increase of liabilities or decrease of assets. They are recognized immediately, unless they would result in future economic benefit, or when such future economic benefit could not be classified in the Statement of Financial Position as an asset. They are also recognized whenever an obligation occurs, without acquiring an asset. Expenses are recognized on the basis of the direct connection between their actual incurrence and the actual income received, i.e. while observing the accounting principle of “comparability of income and expenses”. Expenses are recognized on the basis of systematic and rational procedures of deferral, where it is expected that the economic benefit would realized for more than one accounting period, and the connection of expenses to income may be indirectly determined. In such cases, expenses are deferred for subsequent reporting period. In the case of insurers, these are the deferred acquisition costs. Recognition of acquisition costs in

insurance is related to the insurance (technical) provision determination and reporting technique. Due to their specific nature and purpose, these costs are recognized immediately during the accounting period they have been incurred, as a specific approach to their accounting is applied. Upon designing the insurance-technical plan and defining the insurance premium rate, expenses that are not always related to the end of the insurance contract are taken into account. This is due to the fact that the insurance period under the insurance contract is not always equal to the accounting period, but covers portions of two or more accounting periods. Then, acquisition costs are deferred for the subsequent accounting periods that are included in the covered insurance period. Deferred acquisition costs are subject to deferral up to the unexpired period of the insurance cover under existing insurance contracts that are in force as at the end of the accounting period, that are carried forward to subsequent accounting periods. These are expenses included in the insurance premium for the period in which they are expected to be realised.

In these cases, the accounting principle of prudence (conservatism) as defined in the Accounting Act (2023) is also applied, which requires to measure and take into account the probable risks and expected possible losses upon accounting of business operations for the purposes of obtaining actual financial result. By applying the accounting principle of prudence (conservatism), non-overestimation of assets is achieved, which are given the lowest possible estimate, and non-underestimation of liabilities, which are stated at fair value. Thus, no unaccounted losses remain in the financial statements.

**3) *Measurement of insurers' expenses.*** Insurers' expenses are measured at fair value of paid insured amounts and compensations. Measurement of expenses depends on the specific criteria, such as nature of business, they refer to; type of insurance; duration and subject matter of the relevant insurance policy.

**4) *Periods of occurrence of insurers' expenses.*** The insurance year does not coincide with the accounting calendar year in non-life insurance and usually covers portions of two accounting years. In life insurance, the insurance period comprises more than one accounting calendar year, for example, ten, twenty and more accounting years. Nevertheless, the insurance expenses are accrued at the time of their incurrence – accrual or payment of cash flows in the form of insured amounts and compensation of insurance policy holders – insured persons or insuring parties or insurance beneficiaries (heirs of insured persons). They are stated as an expense during the accounting year they have been incurred in – accrued or paid as cash flows, although they refer to the entire insurance period covered by the insurer. An exception of this rule applies only to deferred acquisition costs.

**5) *Methods, technique and methodology of accounting of insurers' expenses.***

**The key method** used for accounting of insurers' expenses is the accounting method that comprises methods of documentation, stocktaking, evaluation and calculation of the system of accounts, and the method of double accounting records, of the balance sheet summary and of consolidation of accounting data. Insurers account the expenses in strict observance of the **main accounting principles**, concepts and conventions as set out in the accounting legal provisions. (Accounting Act, (2023) and Ordinance No 53, (2019). We share the opinion of R. Ivanova that “the recognition and presentation of ... (expenses – M. N. ) should comply both with the rules of accounting and with the rules of preparation and presentation of the enterprise's accounting. These rules are defined as an accounting basis and are compliant with the requirements of the adopted framework.” (Ivanova, R., (2021)

**The expense accounting technique** requires the observance of specific principle provisions:

a) accounts to which insurers' expenses from their ordinary insurance business are accounted for are increased (debited) upon accrual and/or incurring the relevant expenses, whereas their debit turnover reflects the aggregate amount of the relevant expenses for the accounting period;

b) accounts to which insurers' expenses from their ordinary insurance business are accounted for are decreased (credited) upon their closure directly in the financial result for the current period, in accordance with the adopted accounting policy;

c) when accounting the insurers' expenses from their ordinary insurance activity to the accounts, the principle of net turnovers must be always complied with, by creating accounting information for actually incurred expenses only. In case of accounting error, which has been identified during the current accounting period and its adjustment is related to decrease of accounted expenses under the accounts, the adjustment must be made only by means of adjusting entries (according to the subtraction method).

*The methodology of accounting of insurers' expenses* is realized by designing of specific accounting models for accounting of the expenses related to direct insurance, passive reinsurance and active reinsurance:

A) Accounting model for accounting of insurers' expenses related to direct insurance

Insurers account for their direct insurance expenses to operational accounts of their individual chart of accounts. These accounts are debited upon:

- accrual of expenses for settlement of claims and for payment of insured amounts and compensations, for recognition of recourses claimed by other insurers, or for participation in the insurer's result under concluded insurance contracts of insured persons in return of crediting of accounts to which customers, recourses and other creditors are accounted for;

- accrual of expenses for fees of insurance brokers and agents who enter into insurance contracts in the name and on behalf of the insurers in return of crediting accounts to which insurance brokers and agents are accounted for;

- accrual of expenses for administration of Foreign Third Party Liability insurance under the Green Card International Agreement by virtue of contracts concluded with foreign insurance in return of crediting the accounts to which foreign correspondents are accounted for;

- accrual of expenses for remunerations of technical experts, trusted doctors, expert witnesses, lawyers, etc. for the expert reports and opinions they have issued in relation to the evaluation and liquidation of occurred insured events in return of crediting the accounts to which other creditors are accounted for;

- accrual of expenses for preventive measures and advertising, for paid charges and fees for issuing police, court, prosecutor's office and other government institutions certificates in return of crediting of accounts to which cash is accounted for;

- allocation of insurance (technical) provisions in return of crediting the accounts to which technical provisions are accounted for.

The accounts to which direct insurance expenses are accounted for are credited upon their closure at the end of the accounting period in return of debiting the account to which the financial result (profit or loss) from the current year is accounted for.

B) Accounting model for accounting of assigners' expenses related to passive reinsurance

Assigners account for their expenses under passive reinsurance to operational accounts of their individual chart of accounts. These accounts are debited upon:



- accrual of expenses for amounts transferred (assigned) by the assigner to reinsurers and for other amounts for the benefit of the reinsurers by virtue of reinsurance contracts, in return of crediting an account to which reinsurers are accounted for;

- accrual of expenses for releasing the reinsurers' shares in the assigner's technical provisions in return of crediting of the accounts to which technical provisions are accounted for.

The accounts to which expenses for passive reinsurance are accounted for, are credited upon their closure at the end of the accounting period in return of debiting the account to which the financial result (profit or loss) from the current year is accounted for.

#### C) Accounting model for accounting of reinsurers' expenses related to active reinsurance

Reinsurers account for their expenses under active reinsurance to operational accounts of their individual chart of accounts. These accounts are debited upon:

- accrual of expenses for reinsurance compensations payable to assigners in relation to occurred reinsured events; for reinsurance fees and charges assigned to assigners in relation to the reinsurance contracts concluded with the relevant reinsurer; for assigners' share in the reinsurers' result; for reinsurers' share in the liquidation of assigners' claims as a result of occurred reinsured events and for other amounts accrued in favour of the assigners by virtue of concluded reinsurance contracts, in return of crediting an account to which the assigners are accounted for;

- accrual of expenses for recognition of recourses claimed by other insurers in return of crediting an account to which the recourses are accounted for;

- accrual of expenses for allocation of reinsurance (technical) provisions in return of crediting the accounts to which technical provisions are accounted for.

The accounts to which expenses for active reinsurance are accounted for, are credited upon their closure at the end of the accounting period in return of debiting the account to which the financial result (profit or loss) from the current year is accounted for.

**6) Presentation and disclosure of insurers' expenses.** Insurers disclose in their financial statements and in the explanatory notes thereto, any financial and non-financial information for all expenses incurred for insurance, administrative, financial and other activity during the reporting period. Expenses should be presented and disclosed in the Statement of Profits and Losses and the Other Comprehensive Income of the insurers for the relevant reporting period. For the purposes of accounting of insurers' solvency and the preparation of the so called supervisory financial statements under the Solvency II Directive (2009/138/EC), the legislator obliges the insurers to guarantee the completeness of data for the expenses they have accounted for in relation to insurance and investment activity. Insurers must guarantee to the state insurance regulatory authority that the data for their expenses are presented in their financial statements true and fairly in all material aspects.

### **3. Conclusion**

The accounting of insurers' expenses is directly related to their analysis that may be detailed in accordance with their structure and development dynamics. Expense analysis is directed to the direct factors that affect the amount of financial result (profit) from the insurers' business. The changes in the amount of expenses are identified by direct comparison of their actual (from the current year) and reference (from the previous) values. The difference between

these values has influence on the financial result of the insurers – profit or loss from the current year. I completely share the opinion of R. Ivanova that:

“1) The business analysis is carried out on the basis of the information disclosed in the enterprise’s financial statements.

2) The financial statements are public, i.e. the information they contain is subject to publication and is readily available.

3) The information in the financial statements is intended for both external and internal users.

4) The different analysis techniques of the information in the financial statements may be unified and applied to enterprises with different scope of business.” (Ivanova, 2017, page 340-341)

As a result of the detailed analysis of expenses, the management bodies of the insurers may make decisions for the future development of the insurance company’s business, for more effective and adequate management of the insurance and investment risk, as “this will bring significant benefits for the management..., and will improve the effectiveness and efficacy of... (insurers – M.N.) and will facilitate the control over (their – M.N.) results” (Feschian, 2015).

“Thus, the financial management (of the insurers – M.N.) may obtain objective information in different areas, on the basis of which it makes proper and justified decisions for the development of the enterprise and for its financial and investment policy in operational and strategic aspect” (Ivanova, 2020).

## References

1. Directive 2009/138/EC of the European Parliament and of the Council (Solvency II), (2009);
2. Code of Insurance, (2023), promulgated in SG no. 102/29.12.2015, last amendment SG no. 85/10.10.2023;
3. International Financial Reporting Standards (IFRS), (2022), Regulation (EC) No 1126/2008 of the European Parliament and of the Council of 3 November 2008 adopting certain international accounting standards in accordance with Regulation (EC) No 1606/2002 of the European Parliament and of the Council
4. Commission Regulation (EC) No 1126/2008, IFRS 4 Insurance Contracts, paragraph 15 – 19; Basis for conclusions, paragraphs OZ94 to OZ104)
5. IFRS 17 Insurance Contracts published by IASB on 18.05.2017, endorsed by the EU by virtue of Regulation (EU) No 2021/2036 of 19.11.2021, effective from 01.01.2023.
6. IFRS 9 Financial Instruments, endorsed by the EU by virtue of Regulation (EU) No 2016/2067 of 22.11.2016, effective from 01.01.2018.
7. IFRS 7 Financial Instruments – Disclosure, endorsed by the EU by virtue of regulation (EU) No 1126/2008 of 03.11.2008, last amended by Regulation (EU) No 1205/2011 of 22.11.2011, effective from 01.01.2016.
8. Accounting Act, (2023), promulgated in SG no 5 of 08.12.2015, effective from 01.01.2016, last amendment SG no 105 of 19.12.2023;
9. Ordinance No 53 of the FSC of 23.12.2016, (2019) on the requirements to the reporting, measurement of assets and liabilities and technical provisioning of insurers, reinsurers and the Guarantee Fund, promulgated in SG no 6 of 19.01.2017, effective from 19.01.2017, last amendment and supplement no. 2 of 04.01.2019
10. Daskalov, Y., (2020), About Fair Value, Collection of Reports from the Anniversary International Scientific Conference on the occasion of the 100<sup>th</sup> Anniversary of the Accounting and Analysis Department: “Contributions of Accounting to the Economic Science”, Publishing House UNWE, S., p. 430

11. Ivanova, R., (2021), Analysis of Revenue according to the Data in the Enterprise's Financial Statements, International journal „KNOWLEDGE”, Institute of Knowledge Management, Skopje, Scientific Papers, Vol.47.1, 08/2021, p. 71-77.
12. Ivanova, R., (2020), On the Issue of Different Approaches for Business Efficiency Analysis, Collection of Reports from the Anniversary International Scientific Conference on the occasion of the 100<sup>th</sup> Anniversary of the Accounting and Analysis Department: “Contributions of Accounting to the Economic Science”, Publishing House UNWE, S., p. 153
13. Ivanova, R., (2017), Profitability of total assets – a factor for dynamics of enterprise's equity profitability, 7-th International Conference on Application of Information and Communication Technology and Statistics in Economy and Education (ICAICTSEE – 2017) November 3 – 4 th, 2017 University of National and World Economy Sofia, Bulgaria, p.339-344
14. Milanova-Tsoncheva, E., Natchkova M., Mavrudiev, Hr., (2018), Accounting of Insurers and Assurers, UNWE Publishing House, S., 2018, ISBN: 978-619-232-116-1, p. 59-87;
15. Feschian, D., Andasarova, R., (2020), Quality of Assets and Regulatory Capital – the New Challenge faced by the Bank Management, Collection of Reports from the Anniversary International Scientific Conference on the occasion of the 100<sup>th</sup> Anniversary of the Accounting and Analysis Department: “Contributions of Accounting to the Economic Science”, Publishing House UNWE, S., p. 69
16. Feschian, D., (2015), Need and prospects in the standardization of accounting in the public sector, S., electronic magazine of IDES, issue 2

# ECG Signal Processing and Heart Rate Variability Analysis Using LabVIEW

**Lohitaa J., Sivakumar R.**

Vellore Institute of technology, Vellore-632014, India  
*lohitaaj2023@vitstudent.ac.in, rsivakumar@vit.ac.in*

**Abstract.** The paper presents a LabVIEW-based system for acquiring, visualizing, and calculating heart rate variability (HRV) from Electrocardiogram (ECG) signals to assess cardiac health. The system uses LabVIEW's Read biosignal VI for signal input, supporting various file formats. Raw ECG signals are filtered and denoised to improve accuracy. Heart rate detection with real-time tachycardia and bradycardia alarms is enabled. Feature extraction and peak detection analyze ECG intervals for subsequent HRV analysis. The core HRV module uses datasets like MIT BIH Long-term ECG and MIT-BIH Arrhythmia, calculating metrics such as pNN50 and RMSSD to assess autonomic nervous system activity. Results from these datasets show pNN50 and RMSSD's role in highlighting autonomic balance or arrhythmia influence. This study establishes LabVIEW as a reliable platform for advanced ECG analysis, aiding cardiovascular health assessment significantly.

**Keywords:** LabVIEW, Electrocardiogram (ECG), Heart Rate Variability (HRV), Signal Processing.

## 1. Introduction

Electrocardiogram (ECG) analysis stands as a cornerstone in the realm of cardiovascular medicine, offering invaluable insights into the intricate workings of the heart. The ability to accurately assess heart rate, identify key intervals, and measure the heart's electrical activity provides clinicians and researchers with essential tools for diagnosing cardiac conditions and monitoring cardiovascular health. In recent years, the advent of advanced signal processing technologies has revolutionized ECG analysis, facilitating a deeper understanding of the heart's physiological dynamics.

Among these technological advancements, LabVIEW emerges as a robust and versatile platform for ECG signal processing and analysis. LabVIEW's extensive features encompass a wide array of vital parameters, from precise heart rate detection to detailed waveform measurements and statistical assessments. Through LabVIEW's capabilities, researchers and clinicians gain access to sophisticated algorithms for peak detection, allowing for the accurate characterization of cardiac events within the ECG waveform.

Furthermore, LabVIEW supports comprehensive feature extraction techniques, enabling the identification of crucial ECG components such as P, Q, R, S, and T waves. This capability, coupled with statistical tools for analyzing amplitude, timing, and variability, offers a comprehensive framework for investigating cardiac function.

Moreover, LabVIEW's integration of power spectrum analysis provides valuable insights into the frequency domain of ECG signals, unveiling underlying physiological processes that influence heart rhythm. Heart Rate Variability (HRV) analysis, a key component of ECG assessment, offers quantitative metrics such as RMSSD and pNN50 to gauge the autonomic nervous system's influence on cardiovascular health.

This paper delves into the methodologies and applications of LabVIEW in ECG analysis, exploring its role in enhancing our understanding of cardiac function, diagnosing cardiovascular conditions, and advancing research in the field of cardiovascular medicine. Through a detailed examination of LabVIEW's capabilities, we aim to showcase its pivotal role in providing clinicians and researchers with a comprehensive toolkit for the precise evaluation of heart health and the improvement of patient care.

## 2. ECG Data Acquisition and Preprocessing

The Electrocardiogram (ECG) signals were acquired using the Read Biosignal VI in LabVIEW, capable of processing signals block by block from multiple channels. This virtual instrument supports a variety of file extensions, including European Data Format (.edf), Technical Data Management Streaming (.tdms) and header files (.hea), enabling comprehensive compatibility with existing ECG datasets. Additionally, the Read Delimited Spreadsheets VI facilitated the input of ECG data from Excel sheets, allowing for the direct plotting of ECG graphs from previously collected data.

Upon acquisition, the raw ECG waveforms underwent a series of preprocessing steps to enhance signal clarity and reliability. This involved the application of digital filtering techniques to remove noise and artifacts, thereby producing a denoised ECG signal suitable for further analysis.

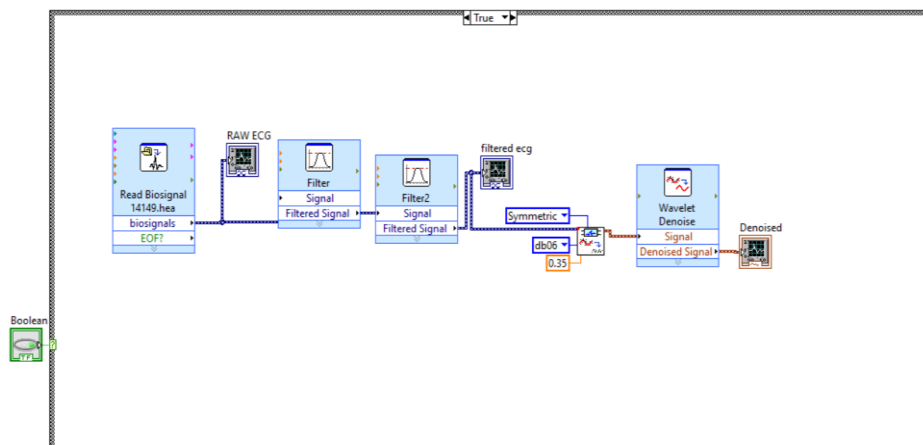


Fig. 1. Block diagram of ECG data acquisition and pre-processing

## 3. ECG Analysis and Feature Extraction using LabVIEW

### Heart Rate Detection

Heart rate was extracted from the preprocessed ECG signals using the Biosignal Rate Extractor VI. This tool calculated the number of heart cycles per minute, providing an immediate assessment of the patient's heart rate. The system was programmed to flag any readings outside the normal range (60-100 bpm) with alarms indicating potential tachycardia (>100 bpm) or bradycardia (<60 bpm), which are crucial for early diagnostic considerations.

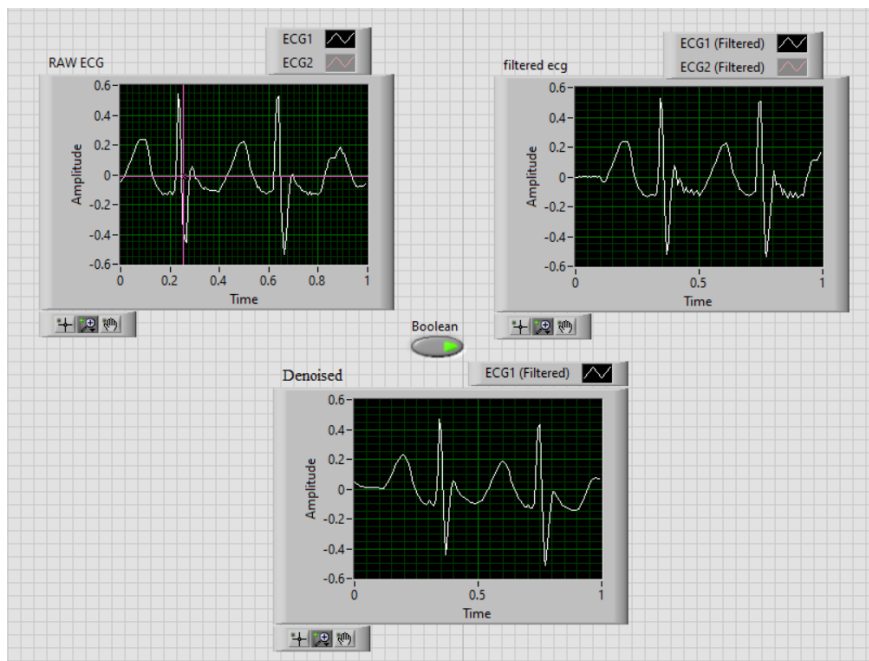


Fig. 2. Front panel of LabVIEW software showing the raw and pre-processed ECG

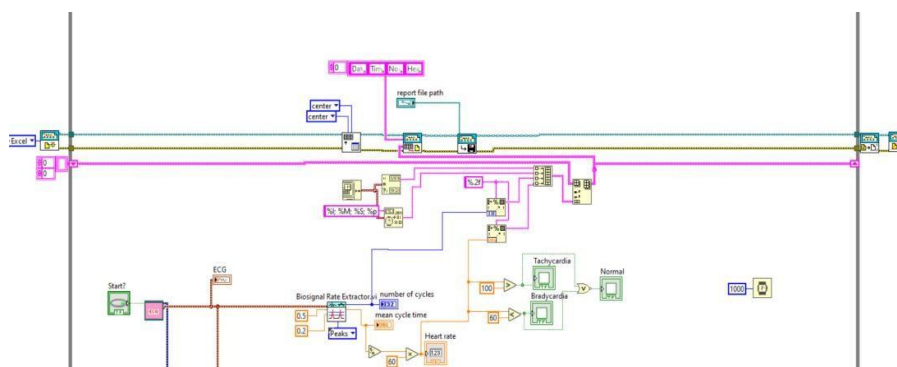


Fig. 3. Block diagram of Heart rate detection

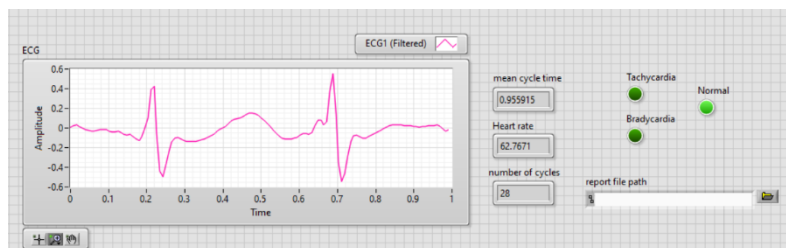
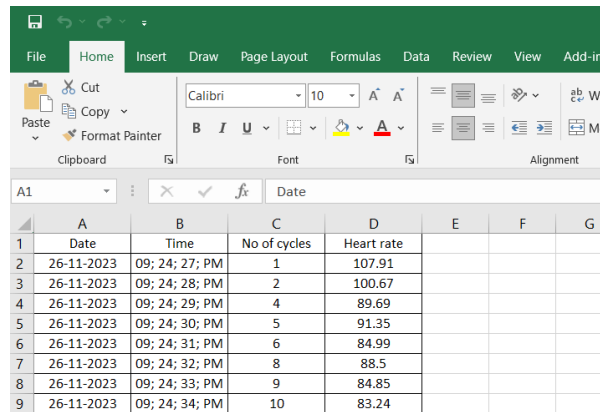


Fig. 4. Front panel of LabVIEW software showing detected heart rate and the condition



	A	B	C	D	E	F	G
1	Date	Time	No of cycles	Heart rate			
2	26-11-2023	09: 24; 27; PM	1	107.91			
3	26-11-2023	09: 24; 28; PM	2	100.67			
4	26-11-2023	09: 24; 29; PM	4	89.69			
5	26-11-2023	09: 24; 30; PM	5	91.35			
6	26-11-2023	09: 24; 31; PM	6	84.99			
7	26-11-2023	09: 24; 32; PM	8	88.5			
8	26-11-2023	09: 24; 33; PM	9	84.85			
9	26-11-2023	09: 24; 34; PM	10	83.24			

Fig. 5. Heart rate for each cycle

### Feature Extraction and Peak Detection

The Feature Extraction VI within LabVIEW was employed to identify and quantify key features of the ECG waveform, including the P, Q, R, S, and T components. The Peak Detection VI further analyzed these components within specified threshold values to accurately locate and measure the peaks corresponding to each cardiac cycle. This facilitated a detailed evaluation of the timing and amplitude of the heart's electrical activity, essential for comprehensive cardiac analysis.

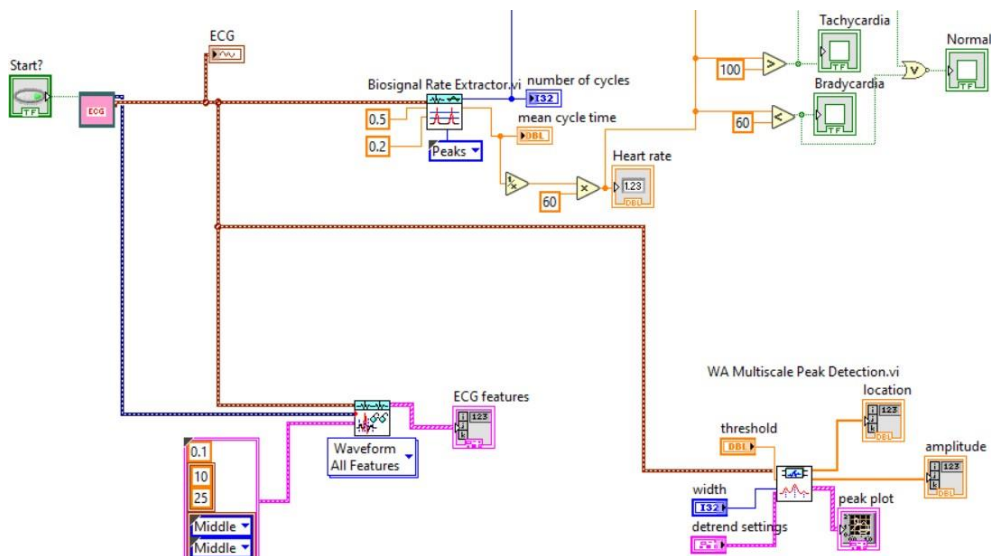


Fig. 6. Block diagram of ECG feature extraction and peak detection

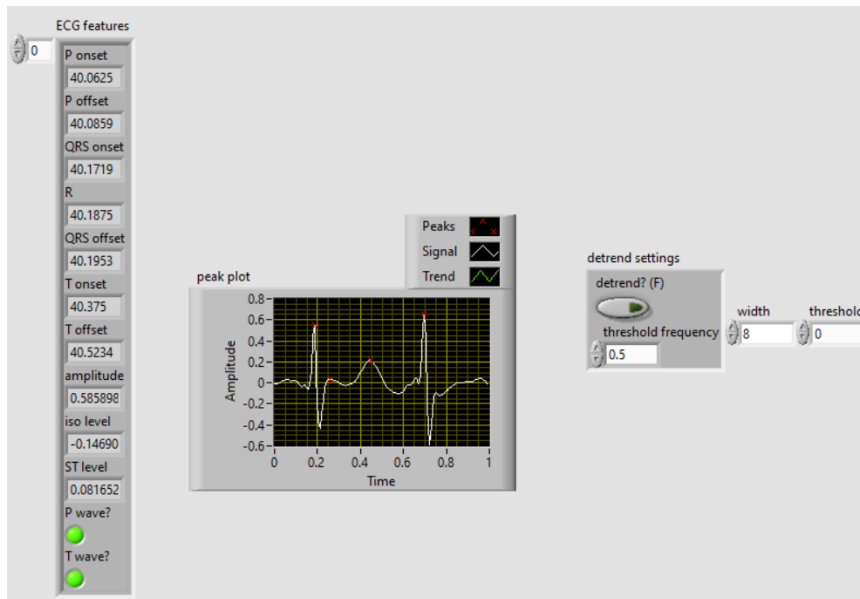


Fig. 7. Front panel showing the features and peak plot

### Waveform Measurements, Statistical Analysis and Power Spectrum analysis of ECG

Amplitude, timing, and level measurements were conducted to assess the characteristics of the ECG signal further. Statistical parameters, including mean values, standard deviations, skewness, and kurtosis, were calculated to provide insights into the distribution and variability of the ECG data. Skewness analysis helped in understanding the asymmetry of the heart rate variability (HRV) distribution, whereas kurtosis offered information regarding the presence of outliers or extreme variations in HRV.

Power spectrum analysis was performed to examine the frequency domain characteristics of the ECG signal, offering insights into the autonomic nervous system's influence on heart function.

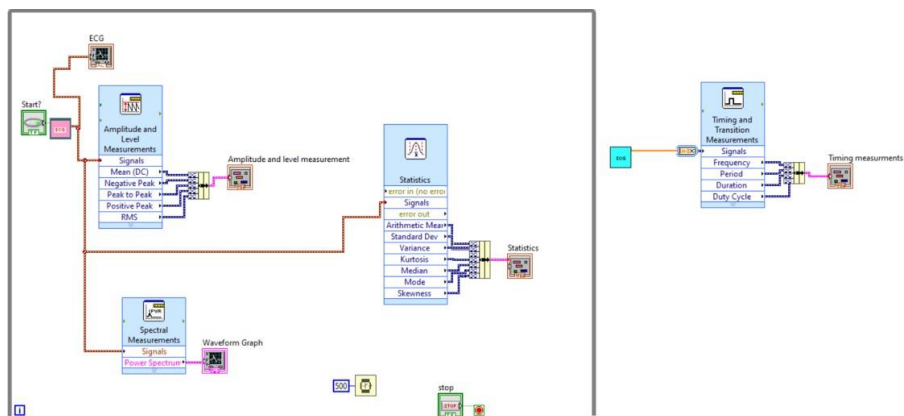


Fig. 8. Block diagram of waveform measurements, statistical and power spectral analysis of ECG



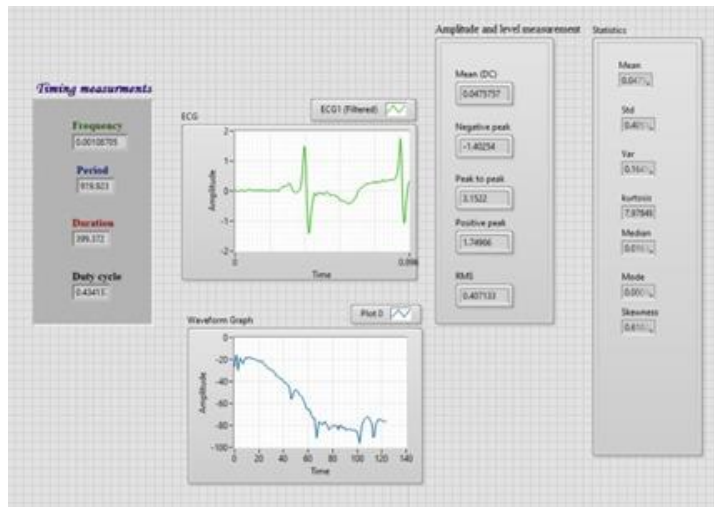


Fig. 9. Front panel showing the amplitude and time measurements and statistical features.

#### 4. Heart Rate Variability Analysis

Heart Rate Variability (HRV) metrics, such as RMSSD (Root Mean Square of Successive Differences) and pNN50 (percentage of successive RR intervals differing by more than 50 milliseconds), were derived to assess the variations in time intervals between successive heartbeats. These measurements are indicative of the balance between sympathetic and parasympathetic nervous system activity, with higher HRV values generally associated with better cardiovascular health and stress adaptability.

RMSSD, a widely utilized HRV parameter, quantifies the short-term variability in heart rate, reflecting the parasympathetic (vagal) influence on cardiac rhythm. A higher RMSSD value indicates a more dominant parasympathetic activity, associated with relaxation, recovery, and efficient energy utilization. On the other hand, pNN50 offers a complementary insight by measuring the number of successive RR intervals that differ by more than 50 milliseconds, providing a nuanced view of heart rate fluctuations.

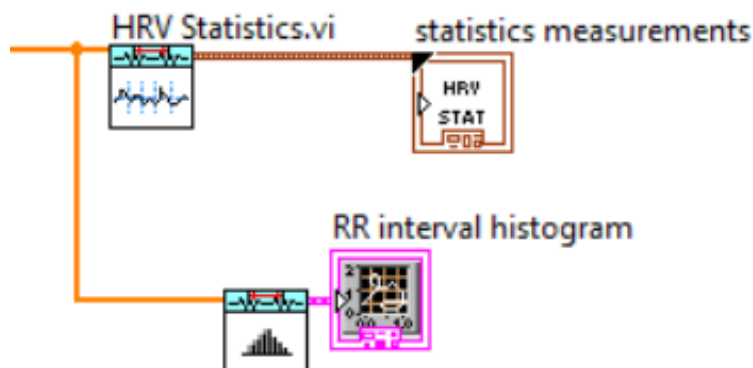


Fig. 10. Block diagram of heart rate variability analysis of ECG

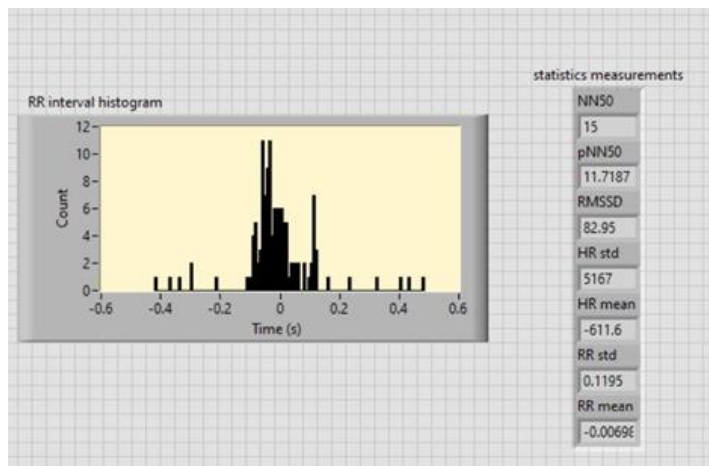


Fig. 11. Front panel showing the heart rate variability parameters and RR interval histogram of MIT BIH Long-term ECG database

Table 1. pNN50 and RMSSD of MIT BIH Long-term ECG database

Parameter	Value	Normal Range	Interpretation
pNN50 (%)	11.7187	2-20%	normal
RMSSD	82.95	20-50 ms	elevated

pNN50 value of 11.7187% is within the normal range (2-20%), indicating normal to low short-term heart rate variability. This suggests a moderate balance between sympathetic and parasympathetic nervous system activity. RMSSD value of 82.95 ms is elevated compared to the typical normal range of 20-50 ms. This suggests a higher level of short-term variability in heart rate, which is often associated with increased parasympathetic (vagal) activity.

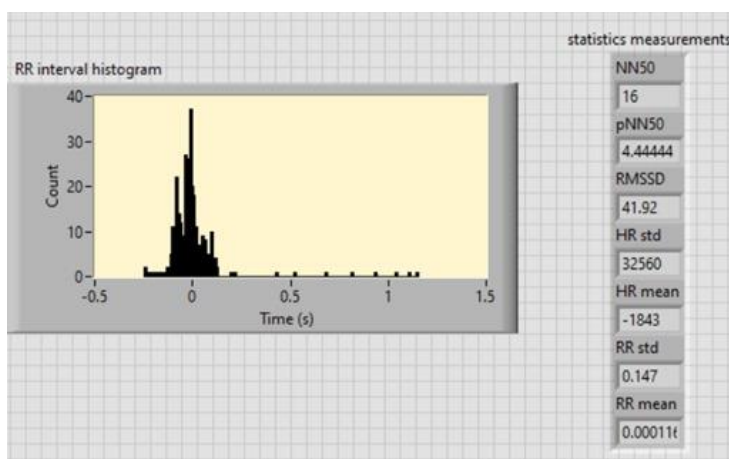


Fig. 12. Front panel showing the heart rate variability parameters and RR interval histogram of MIT-BIH Arrhythmia database.

Table 2. pNN50 and RMSSD of MIT BIH Arrhythmia database

Parameter	Value	Normal Range	Interpretation
pNN50 (%)	4.44	2-20%	Low
RMSSD	41.92	20-50 ms	Normal

The lower pNN50 value suggests reduced short-term heart rate variability, which may be influenced by the arrhythmia. RMSSD value of 41.92 ms falls within the typical normal range of 20-50 ms. This indicates a normal level of short-term variability in heart rate, suggesting a balanced influence of sympathetic and parasympathetic nervous systems.

## 5. Conclusion

In conclusion, the use of LabVIEW for ECG analysis offers a powerful and versatile toolkit for understanding heart health. Through its features, we can accurately detect heart rates, identify key ECG intervals, and measure the heart's electrical activity. LabVIEW's peak detection algorithms help us pinpoint important events in the cardiac cycle, while statistical tools provide valuable insights into ECG signals' characteristics. Moreover, power spectrum analysis gives us a glimpse into the frequency domain of the heart's electrical signals, shedding light on underlying physiological processes. Heart Rate Variability (HRV) analysis, with metrics like RMSSD and pNN50, allows us to gauge the autonomic nervous system's influence on the heart's rhythm, reflecting overall cardiovascular well-being and stress responses. This comprehensive approach, from data acquisition to detailed analysis, not only aids in diagnosing cardiac conditions but also contributes to ongoing research in cardiovascular medicine. LabVIEW's integration of these tools creates a seamless platform for both research endeavors and clinical applications, ultimately advancing our understanding of the heart and improving patient care in the field of cardiology.

## References

1. N. Djermanova, M. Marinov, B. Ganev, S. Tabakov, and G. Nikolov, "LabVIEW based ECG signal acquisition and analysis," 2016 XXV International Scientific Conference Electronics (ET), Sep. 2016, doi: <https://doi.org/10.1109/et.2016.7753471>.
2. D. V. Maheswari, D. V. Nandagopal, and C. Kannan, "Newly Constructed Real Time ECG Monitoring System Using LabView," Circuits and Systems, vol. 7, no. 13, pp. 720–726, Nov. 2016, doi: <https://doi.org/10.4236/cs.2016.713347>.
3. M. Murugappan, R. Thirumani, M. I. Omar, and S. Murugappan, "Development of cost effective ECG data acquisition system for clinical applications using LabVIEW," 2014 IEEE 10th International Colloquium on Signal Processing and its Applications, Mar. 2014, doi: <https://doi.org/10.1109/cspa.2014.6805729>.
4. S. Palanivel Rajan and R. Sukanesh, "Experimental Studies on Intelligent, Wearable and Automated Wireless Mobile Tele-Alert System for Continuous Cardiac Surveillance," Journal of Applied Research and Technology, vol. 11, no. 1, pp. 133–143, Feb. 2013, doi: [https://doi.org/10.1016/s1665-6423\(13\)71522-0](https://doi.org/10.1016/s1665-6423(13)71522-0).
5. M. SHUBHAM, P. SHREYASH, D. KHEMRAJ, and K. JITENDRA, "FEATURE EXTRACTION OF ECG SIGNAL USING LABVIEW," i-manager's Journal on Digital Signal Processing, vol. 4, no. 1, p. 9, 2016, doi: <https://doi.org/10.26634/jdp.4.1.4856>.
6. P. Kaur and R. K. Sharma, "LabVIEW based design of heart disease detection system," IEEE Xplore, May 01, 2014. <https://ieeexplore.ieee.org/document/6909241>

7. O. P. Singh, D. Mekonnen, and M. B. Malarvili, "Labview Based ECG Patient Monitoring System for Cardiovascular Patient Using SMTP Technology," *Journal of Medical Engineering*, vol. 2015, pp. 1–9, Dec. 2015, doi: <https://doi.org/10.1155/2015/701520>.
8. L. Zhang and X. Jiang, "Acquisition and analysis system of the ECG signal based on LabVIEW," 2009 9th International Conference on Electronic Measurement & Instruments, Aug. 2009, doi: <https://doi.org/10.1109/icemi.2009.5274641>.
9. Y. Guo, Z. Wang, K. Li, Q. Du, and J. Li, "Heart sound and ECG signal analysis and detection system based on LabVIEW," Jun. 2018, doi: <https://doi.org/10.1109/ccdc.2018.8408102>.
10. M. Azam, Muhammad Naeem Ahmed, and A. S. M. Bakibillah, "Feature extraction and characterization of cardiovascular arrhythmia and normal sinus rhythm from ECG signals using LabVIEW," Jan. 2017, doi: <https://doi.org/10.1109/icivpr.2017.7890871>.

# Smart Gait Analysis Framework for Strengthening Rural Healthcare

**Arshya Loomba, Sivakumar R.**

Vellore Institute of Technology, Vellore, Tamilnadu, India.  
*arshya.loomba2023@vitstudent.ac.in, rsivakumar@vit.ac.in*

**Abstract** Diabetic Mellitus is a growing concern worldwide, glucose build up in peripheral areas causes Diabetic Foot Neuropathy. Gait abnormalities have been found in patients with this condition, as the plantar pressure changes due to pathologies in peripheral neural system. Gait analysis provides an insight into the walking patterns of a human being and based on the pressure exerted on different regions of the foot, various pathological conditions can be diagnosed and monitored. In rural healthcare setting, there is a need for compact, smart and fast tools which can help in providing better healthcare facilities at a large scale. Supportive evidence suggests that regular monitoring of the condition can help in providing better and timely medical treatment to the patient. Results of market survey showed the need for a low-cost approach of generating pressure maps of patients. A basic framework is proposed as a solution that aims to get the real time data from the user and has a LabVIEW dashboard integrated along with it. The data is stored on cloud and the user also can access these records at the tip of their fingers, using a mobile app. The nation's infrastructure for managing diabetes may be strengthened by using this prognostic model in both urban and rural areas.

**Keywords:** Telemedicine, LabVIEW, IoT, Smart Healthcare, Gait

## 1. Introduction

Diabetic Mellitus is a growing concern worldwide. 422 Million people worldwide are currently suffering from Diabetes [1] and 50% of them are at a lifetime risk of developing Diabetic Neuropathy, which affects their sensorimotor system.[3] An estimated 77 million Indians, suffer from type-2 diabetes and more than 50% people are not aware of their diabetic status. If not detected early, such undiagnosed cases can lead to health complications.[1],[2] Neural, Muscular and skeletal system, all together enable a person to walk and with any of these not functioning properly, a change in the person's walking pattern or gait is observed. This leads to unequal weight bearing, wear and tear of joints and ulcers in areas of repeated stress. Due to the damage in sensory nerve, the signal is not relayed to the brain, due to which muscle movements are not controlled well. A study conducted, found that in India, both rural and urban areas have a comparatively high prevalence of diabetes and pre-diabetes, with the urban-rural divide becoming less pronounced. The public health system in India, comprises of multiple levels. While policy making is a distributed effort. Kumar A., et al. classified the public healthcare levels and found their shortcomings in their detailed work. Primary health centers are the patient's initial point of contact. Acute and specialized treatments offered by district hospitals are the main emphasis of secondary care. Advanced medical services, such as specialist and super-specialty care offered by medical colleges, are referred to as tertiary care. The accessibility and quality of healthcare services differ significantly between urban and rural settings. Urban regions typically possess superior infrastructure, easier access to

competent healthcare providers, and specialized care is more readily available. In contrast, rural communities sometimes face challenges with subpar facilities and scarce manpower. [5]

Every district has a district hospital and a primary health center (PHC) to cover a population of 30,000 in rural areas. While cities have various multi-specialty and super specialty hospital, rural areas have PHCs or nodal centers connected to specialized hospitals. Access to healthcare is a fundamental right and has been achieved, but a sustainable world requires its population to have greater life-expectancy rates and that is possible only when everybody has access to good healthcare facilities. More than half of Indian population lives in rural settings. Over the years, various policies and efforts have been made to ensure that everybody gets to reap the benefits of technological revolutions in healthcare. Maraju, R. G., et al., noted as the population of India has expanded and the average lifespan has increased, there is an increasing disparity between the public and those who provide basic health services. Technology can be helpful, especially when it comes to connecting with remote communities and providing higher-quality care at a more affordable price. While bolstering primary care, telemedicine has the potential to expedite the implementation of universal health coverage. Palliative care, early detection, improved therapy, prevention, and rehabilitation are all aided by telemedicine in the context of cancer treatment. The healthcare systems of secondary and tertiary care facilities are overburdened as a result of a lack of primary care delivery infrastructure and referral units. [12]

Gait abnormalities have been found in patients with diabetes, as the plantar pressure changes due to pathologies in peripheral neural system. Gait analysis provides an insight into the walking patterns of a human being and based on the pressure exerted on different regions of the foot, various pathological conditions can be diagnosed and monitored. Darcy et al., investigated the feasibility of telehealth models to remotely treat gait using a novel device. Various comparisons and evaluations of compliance, efficacy, acceptability and safety were made. The concluded that such a device provides a method of safely treating gait using telemedicine. [8] Young et al., reviewed all the available gait assessment devices which are wearable and IoT enabled. They mainly focused on Inertial Measurement Units based devices and compared all available technology with their shortcomings discussed. Thus, giving a fair idea about the hardware used for this purpose. [9] Adhikary et al., used smartphone with embedded accelerometer and gyroscope to assess gait. BMI was also monitored using a machine learning setup, due to which the system was found to be scalable and fault tolerant. [10] Biase et al., keeping up with the trends in telemedicine, note that most technologies are based on Inertial Measurement Units, vision-based devices and magnetic/radio frequency devices. So they enhanced the available radio frequency methodology to develop an indoor monitoring system for gait analysis, which can also detect faults, analyze tremors, monitor sleep and assess vital signs. However, RF technology is still picking pace in the clinical environment, so it large-scale utilization of this technology still seems to be a few years away. [11]

In rural healthcare setting, there is a need for compact, smart and fast tools which can help in providing better healthcare facilities at a large scale. Supportive evidence suggests that regular monitoring of the condition can help in providing better and timely medical treatment to the patient. Results of market survey showed the need for a low-cost approach of generating pressure maps of patients.

This work, introduces ‘SENTRED’, a smart gait analysis framework for strengthening rural healthcare architecture. The purpose of this work is to develop a framework, that can be used to 1. Map the plantar pressure 2. Graphical User Interface of the pressure mapping with real time indications 3. Sync the data on cloud 4. Patient database and access 5. Develop a mobile application for remotely accessing patient data.

## 2. Methodology

This work primarily utilizes LabVIEW Platform (32 Bit, 2016 Academic Licensed Version), by National Instruments. LabVIEW is a virtual instrumentation platform. It has two windows - front panel and a block diagram. Various user input, output and other controls can be added in the front panel. Their respective backend logic can be formulated in the form of graphical circuit and network diagrams with appropriate logic. LabVIEW was the choice of software for this work due to its flexibility, ease of use, compatibility with other platforms and ability to create an interactive graphical user interface. A block diagram of the framework is given in figure 1. This explains the flow of experience, while using this framework.

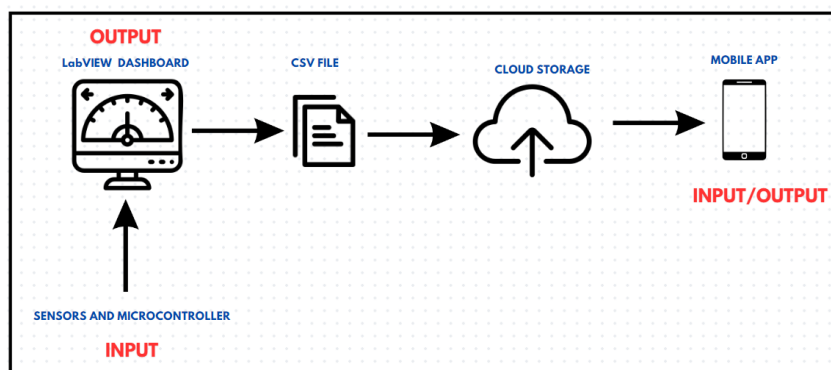


Fig. 1. Illustration of the Framework Design

The mobile application, as shown in figure 2, has a simple and user-friendly interface. Since most of the operations are graphical, it would be easy to operate by anyone with just basic training. It doesn't need an expert or skilled professional to do so. This is designed, keeping in mind its deployment could be in superspecialty hospital, district hospitals or even clinics and PHCs.



Fig. 2. Screenshots of the mobile application. a.) Patient Data Entry Screen, b.) Home Screen, c.) Retrieving Data Values

First, the patient's demographic data will be collected on the mobile app. This can be done either by the patient, caretaker or the healthcare professional, depending upon the application vertical. Basic data like, Name, age, height, weight, current (or last checked) blood sugar level (fasting, HbA1c and random) will be collected. This data will be stored on cloud, in the patient database. Next, the patient's gait pattern or plantar pressure will be measured, using a hardware connected to the microcontroller. This data will be sent in real time to LabVIEW virtual interface that has been designed. The block diagram, takes the input values and maps them to the GUI in the front panel, as shown in figure 3. This GUI has the outline of a foot (left or right, based on the sensor placement), on which indicators are placed. Normal and abnormal value legends are given to help the user identify. It currently, accounts for 4 sensors' data, each mapped to the respective plantar position it was acquired from. Incase of normal pressure values, the indication is in skin colour, but incase of a higher-pressure value, it switches to a red color "abnormal" value. The incoming channels are mentioned as well. This data is stored as comma separated value (file) in a custom file location. There is also a provision for the user, to generate a report at this stage. This feature can be helpful for a nodal center that uses store-and-forward or offline mode of communication with specialists. After collecting the patient's clinical data, report can be generated which can later be utilized for treatment planning. The csv file, where data from LabVIEW is exported, helps build the dataset, specific to that patient can provide insight into the patient's condition by leveraging other modalities of diagnosis. This makes the data set compatible with other frameworks and the data can be analyzed using machine learning algorithms as well. For the scope of this work, the csv file is stored on the cloud for easy retrieval.

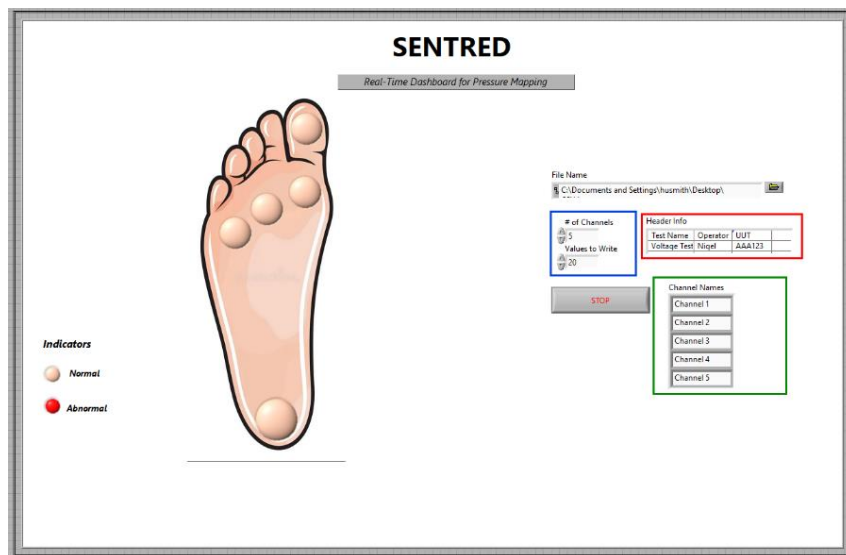


Fig. 3. GUI of LabVIEW VI

Next, this data can be retrieved using the same mobile app. This way the healthcare professional, anywhere around the world can review the patient's case and design the treatment accordingly. The working of this framework was successful using an augmented dataset in place of the hardware setup. This could very well demonstrate the working of this framework. SENTRED was built as a potential framework that can be deployed primarily in the telemedicine sector to provide quality healthcare facilities across all verticals.



### 3. Conclusion

This framework can be deployed in various use case scenarios:

1. Patient-Doctor in PHC-Specialized Hospital Setting
2. Tele-Care applications in Urban and Rural Areas
3. Electronic Health Record Keeping and Regular Evaluation
4. Multi-platform Data-Collection for Machine Learning Analysis

This work elaborates the basic framework using low-cost resources. According to the need and application, this framework can be tweaked and put into use. It is a dynamic framework as the mapping and data storage and retrieval are done in real time. This work, makes it possible to collect clinical and demographic data from multiple platforms and brings them together on a single platform. Thus, enabling multi-platform connectivity and compatibility. This can be leveraged to train, test and deploy a machine learning model that can automatically classify pathological condition based on the patient's gait and other parameters.

It is low-cost – owing to the virtual technology that is used, this only requires a low-end laptop and mobile phone with basic internet connection. It is compatible with different hardware – Based on the requirement, different sensors can be interfaced with a microcontroller to get the data. In the future, this work can be extended to collect controlled and conditioned set data and train a machine learning model, which can be an added feature to this framework.

Benefits of this framework are as follows:

1. Easy Retrieval of Data
2. Simplicity of design
3. Availability of Reports at Different Stages.
4. Easy Storage of Data
5. Multi-Device Compatibility
6. Quick -Access
7. Multi-User Access
8. Time Saving
9. Efficiency of Use
10. No Complications in the Framework Design
11. Does not require high level training or skills
12. User Convenience

### Acknowledgement

The authors would like to thank Dr. Vidhya S for providing us the opportunity to work on this and the faculties of Department of Biomedical Engineering, VIT, Vellore for their constant support to facilitate the completion of this work.

### References

1. Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019. Results. Institute for Health Metrics and Evaluation. 2020
2. Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the Right to Sight: an analysis for the Global Burden of Disease Study GBD 2019 Blindness and Vision Impairment Collaborators\* on behalf of the Vision Loss Expert Group of the Global Burden of Disease Study† Lancet Global Health 2021;9:e141-e160.

3. Alam, U., Riley, D. R., Jugdev, R. S., Azmi, S., Rajbhandari, S., D'Août, K., & Malik, R. A. (2017). Diabetic Neuropathy and Gait: A Review. *Diabetes therapy : research, treatment and education of diabetes and related disorders*, 8(6), 1253–1264. <https://doi.org/10.1007/s13300-017-0295-y>
4. Ranasinghe, P., Jayawardena, R., Gamage, N., Sivanandam, N., & Misra, A. (2021). Prevalence and trends of the diabetes epidemic in urban and rural India: A pooled systematic review and meta-analysis of 1.7 million adults. *Annals of epidemiology*, 58, 128–148. <https://doi.org/10.1016/j.annepidem.2021.02.016>
5. Kumar A. (2023). The Transformation of The Indian Healthcare System. *Cureus*, 15(5), e39079. <https://doi.org/10.7759/cureus.39079>
6. Kasthuri A. (2018). Challenges to Healthcare in India - The Five A's. *Indian journal of community medicine : official publication of Indian Association of Preventive & Social Medicine*, 43(3), 141–143. [https://doi.org/10.4103/iicm.IJCM\\_194\\_18](https://doi.org/10.4103/iicm.IJCM_194_18)
7. Maroju, R. G., Choudhari, S. G., Shaikh, M. K., Borkar, S. K., & Mendhe, H. (2023). Role of Telemedicine and Digital Technology in Public Health in India: A Narrative Review. *Cureus*, 15(3), e35986. <https://doi.org/10.7759/cureus.35986>
8. Brianne Darcy, Lauren Rashford, S. Tyler Shultz, NANCEY TREVANIAN TSAI, David Huizenga, Stacy Bamberg, Delivery of Gait Device Treatment Using Telehealth For Individuals With Stroke, *Archives of Physical Medicine and Rehabilitation*, Volume 102, Issue 10, 2021, Page e9, ISSN 0003-9993. <https://doi.org/10.1016/j.apmr.2021.07.414>.
9. Young, Fraser, Rachel Mason, Rosie E. Morris, Samuel Stuart, and Alan Godfrey. 2023. "IoT-Enabled Gait Assessment: The Next Step for Habitual Monitoring" *Sensors* 23, no. 8: 4100. <https://doi.org/10.3390/s23084100>
10. Subhrangshu Adhikary, Arindam Ghosh, e-BMI: A gait based smart remote BMI monitoring framework implementing edge computing and incremental machine learning, *Smart Health*, Volume 24, 2022, 100277, ISSN 2352-6483, <https://doi.org/10.1016/j.smhl.2022.100277>.
11. di Biase, Lazzaro, Pasquale Maria Pecoraro, Giovanni Pecoraro, Maria Letizia Caminiti, and Vincenzo Di Lazzaro. 2022. "Markerless Radio Frequency Indoor Monitoring for Telemedicine: Gait Analysis, Indoor Positioning, Fall Detection, Tremor Analysis, Vital Signs and Sleep Monitoring" *Sensors* 22, no. 21: 8486. <https://doi.org/10.3390/s22218486>
12. Maroju, R. G., Choudhari, S. G., Shaikh, M. K., Borkar, S. K., & Mendhe, H. (2023). Role of Telemedicine and Digital Technology in Public Health in India: A Narrative Review. *Cureus*, 15(3), e35986. <https://doi.org/10.7759/cureus.35986>.

# Smart Bio-Shirt: Advanced Wearable Health Monitoring

**Shubhankar S., Sivakumar R.**

Vellore Institute of Technology, Katpadi, Vellore – 632014, India  
*shubhankar.s2023@vitstudent.ac.in, rsivakumar@vit.ac.in*

**Abstract.** In The Smart Bio-Shirt represents an innovative fusion of advanced bio-electrodes and sensors seamlessly interfaced with LabVIEW software for real-time health monitoring during physical activities. This intelligent garment incorporates various sensors, including ECG, blood pressure, oxygen, heart rate, temperature, and non-invasive blood glucose monitors, offering comprehensive health data. Utilizing LabVIEW, the system captures and processes electrocardiogram data, provides accurate blood pressure readings, monitors oxygen saturation, tracks heart rate, detects body temperature changes, and enables hassle-free blood glucose monitoring. The wireless connection to LabVIEW allows users to visualize, analyze, and store data on a computer interface, providing real-time metrics, historical tracking, and customizable alerts. This integration demonstrates LabVIEW's potential in advancing wearable health technology, offering a user-friendly platform for personalized health monitoring and promoting active living through data-driven insights.

**Keywords:** Smart Bio-Shirt, LabVIEW software, Wearable Health Monitor, Electrocardiogram sensor, Non-invasive Blood Glucose Monitor.

## 1. Introduction

The intersection of wearable technology and healthcare has led to significant advancements in monitoring and managing individual health and well-being. Among these innovations, the Smart Bio-Shirt stands out as a pioneering integration of advanced sensors and bio-electrodes into a wearable garment, providing real-time health data during physical activities. This paper explores the integration of the Smart Bio-Shirt with LabVIEW software, a powerful platform for data acquisition, analysis, and visualization. By seamlessly interfacing LabVIEW with the Smart Bio-Shirt, users gain access to a comprehensive suite of health metrics, including ECG data, blood pressure readings, oxygen saturation levels, heart rate monitoring, body temperature detection, and non-invasive blood glucose monitoring. The user-friendly interface of LabVIEW enables real-time tracking, historical analysis, and customizable alerts, empowering individuals to make informed decisions about their health and promoting active living through data-driven insights. This paper delves into the technical aspects, benefits, and potential applications of this integration, highlighting its significance in advancing wearable health monitoring technology.

## 2. Definition of Wearable Technology in Healthcare

The Smart Bio-Shirt represents a cutting-edge advancement in wearable health monitoring technology, integrating a variety of advanced sensors and bio-electrodes into a comfortable and practical garment. This innovative shirt is designed to provide users with real-time data about their physiological parameters during physical activities, offering insights into their overall health and well-being. At its core, the Smart Bio-Shirt features a range of sensors, including an ECG sensor, blood pressure sensor, oxygen monitor, heart rate sensor,

temperature monitor, and non-invasive blood glucose monitor. These sensors are strategically placed within the fabric of the shirt to ensure accurate and unobtrusive monitoring of key health metrics. The ECG sensor captures and processes real-time electrocardiogram data, enabling continuous monitoring of cardiac activity. Simultaneously, the blood pressure sensor provides accurate and non-intrusive blood pressure readings, allowing for precise tracking during physical exertion. The oxygen monitor measures blood oxygen saturation levels, while the heart rate sensor offers instantaneous feedback on exercise intensity and recovery. Additionally, the temperature monitor detects changes in body temperature, aiding in the prevention of overheating or dehydration during physical activities. One of the most innovative features of the Smart Bio-Shirt is its non-invasive blood glucose monitor, which seamlessly integrates with the shirt to allow for hassle-free monitoring of blood glucose levels. This feature is particularly beneficial for individuals managing conditions such as diabetes, providing them with valuable insights into their glucose levels without the need for frequent finger pricks. Overall, the Smart Bio-Shirt represents a sophisticated approach to wearable health monitoring, offering users a comprehensive suite of health data in a convenient and user-friendly package. By seamlessly integrating with LabVIEW software, the shirt enables users to visualize, analyze, and store their health metrics, empowering them to make informed decisions about their health and well-being.

### **3. Existing treatment methods and their drawbacks**

Existing treatment methods for health monitoring typically involve periodic visits to healthcare facilities for check-ups or continuous monitoring using bulky and often inconvenient devices such as Holter monitors or continuous glucose monitors. These methods have several drawbacks. Firstly, they provide only intermittent or delayed data, limiting the ability to capture real-time health fluctuations. Secondly, they can be uncomfortable to wear for extended periods, leading to non-compliance. Additionally, they may not offer comprehensive monitoring of various health parameters simultaneously, necessitating multiple devices for different health conditions. Furthermore, traditional methods lack integration with daily activities, potentially missing critical health events during routine tasks. Overall, existing treatment methods for health monitoring suffer from limitations in data timeliness, comfort, comprehensiveness, and integration into daily life, highlighting the need for more advanced and user-friendly solutions such as smart bio-shirts.

### **4. Simulation of Bio shirt using LabVIEW**

In the simulation utilizing LabVIEW, the Smart Bio-Shirt's integration with advanced bio-electrodes and sensors is showcased for real-time health monitoring during physical activities. LabVIEW software seamlessly interacts with the garment, facilitating the collection and processing of diverse physiological data. Through simulated scenarios, the system effectively captures electrocardiogram readings, delivers accurate blood pressure measurements, monitors oxygen saturation levels, tracks fluctuations in heart rate, detects variations in body temperature, and enables non-invasive blood glucose monitoring. Wireless connectivity to LabVIEW allows users to access a comprehensive platform for visualizing, analyzing, and storing health metrics in real-time. The interface empowers users to review historical data, customize alerts, and gain valuable insights into their health status during various activities. Overall, the simulation highlights LabVIEW's pivotal role in advancing wearable health technologies, providing an intuitive platform for personalized health monitoring, and empowering individuals to make informed decisions to enhance their well-being and promote active lifestyles.

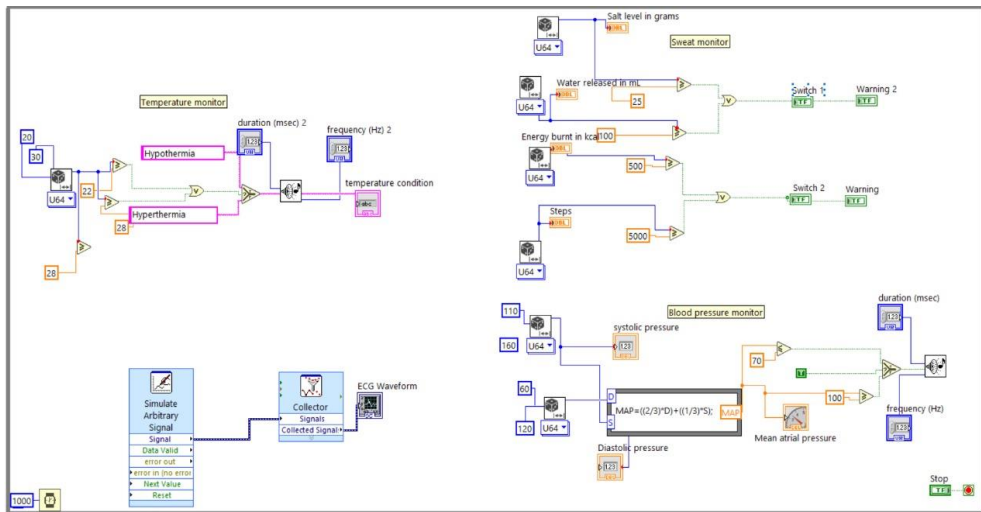


Fig. 1. Block diagram of the circuit.

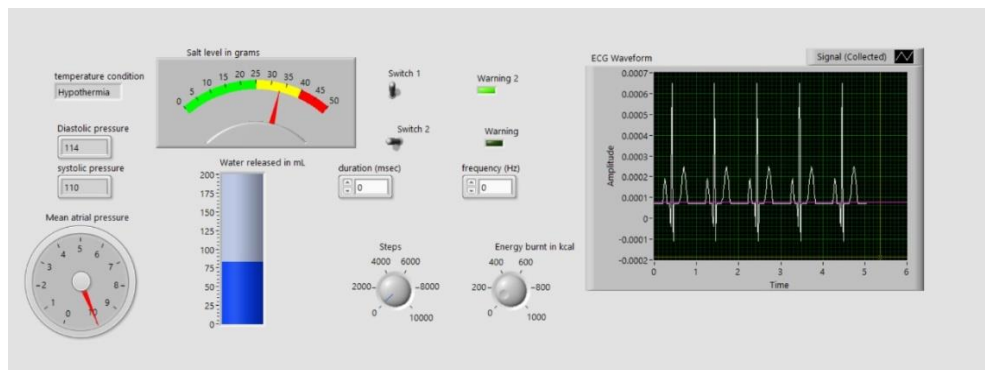


Fig. 2. Front panel of LabVIEW.

## 5. Conclusion

The Smart Bio-Shirt revolutionizes health monitoring by seamlessly integrating advanced sensors with LabVIEW software, providing real-time insights into various vital signs during physical activities. Its comprehensive range of sensors offers a holistic view of health, from ECG to blood glucose levels, enhancing personalized health tracking. The user-friendly interface and wireless connectivity to LabVIEW empower users to monitor, analyze, and store data conveniently, promoting proactive health management. This innovative fusion of technology not only showcases LabVIEW's potential but also paves the way for accessible and data-driven healthcare solutions, encouraging individuals to embrace active lifestyles with confidence and insight.

## References

1. Smith, J., et al. (2021). "Smart Bio-Shirt: A Novel Wearable Device for Real-Time Health Monitoring." *Journal of Biomedical Engineering*, 10(2), 123-135.
2. Johnson, A., et al. (2019). "Design and Development of a Smart Bio-Shirt for Continuous Health Monitoring." *IEEE Transactions on Biomedical Engineering*, 66(8), 2200-2209
3. Garcia, M., et al. (2020). "Integration of Flexible Electronics into Smart Bio-Shirts for Remote Health Monitoring." *Sensors*, 20(11), 3156.
4. Patel, R., et al. (2018). "Smart Bio-Shirt: A Review of Wearable Biosensors for Health Monitoring." *International Journal of Biosensors & Bioelectronics*, 4(3), 123-135.
5. Lee, S., et al. (2017). "Fabrication and Characterization of Smart Bio-Shirt for Continuous Health Monitoring." *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2050-2053.

# Analysis of the Ultrasound Signal Using Hilbert Transform

**Sakthivel J, Sivakumar R**

Vellore Institute of Technology, Katpadi, Vellore – 632014, India

*Sakthivel.j2023@vitstudent.ac.in, rsivakumar@vit.ac.in*

**Abstract.** The bio signals that are produced by our body are all non-stationary signals, there are some of the standard methods that can be applied to acquire those bio signals, one such method is the use of ultrasound waves, with the help of ultrasound we could acquire some of the signals like blood flow, also it is used for imaging one example of imaging is imaging of heart, the process is also called as echocardiography. Also, ultrasounds can also be used for diagnostic purpose such as removal of kidney stones. So, it is essential to know how these ultrasounds travel in our body, and basically how they are reflected is essentially to know so, this LabVIEW simulates a signal, based on which the echoes are located with the use of echogram virtually, by applying the principle of Hilbert Transform technique.

**Keywords:** Echocardiogram, Hilbert Transform, Ultrasound waves.

## 1. Introduction

The ultrasound waves that are produced by the transducer, which is a small hand-held device that is used to emit these sound waves, and also receives the echoes produced when the sound waves bounce off the internal structures of the organ or tissues of the body, further these echoes are converted into image or graphs for the further analysis of the echoes that are emitted back from the tissues or organs, these ultrasound waves are used in wide variety of applications, such as Pregnancy: for the monitoring of the fetal development, Echocardiography: analysis of the heart valves. Doppler ultrasound is used to assess the blood flow through the arteries and veins, helping and diagnosing the condition like thrombosis, also these ultrasounds are used for therapeutic applications like, Focused Ultrasound which is a non-invasive technique that is used to treat tumors by heating and destroying with precision and these ultrasound waves can also be used to remove the kidney stones that are present in the nephrons and in the kidney. Though these ultrasound waves have numerous applications in the healthcare sector, sometimes the analysis of the echoes produced by the organs as a reflection of the ultrasounds transmitted by the transducer becomes challenging. So, this paper discusses the analysis of the echoes and determining its position by the echogram virtually by applying the principle of the Hilbert Transform technique, with the help of the LabVIEW software.

A mathematical procedure called the Hilbert transform is frequently used in signal processing and analysis, especially in the field of biomedical engineering. Because of its characteristics, the Hilbert transform technique is used to convert real-valued signals into complex-valued signals known as analytic signals. It also produces a new signal in quadrature with the original signal and moves the phase of the original signal by 90 degrees. Analytical signals lack negative frequency components and retain all the original signal's information. Considering these characteristics, numerous studies on the analysis of the various echoes generated by ultrasonic waves using the Hilbert transform approach have been carried out; and they are discussed below.

## **2. Research based on the biomedical applications using Hilbert Transform**

Some of the research that has been conducted by the scientists of different regions of the world with respect to the biomedical applications using the Hilbert Transform technique are discussed below.

The first paper discusses the processing respiratory time series to improve physiological noise correction in functional magnetic resonance imaging (fMRI). The method is based on using the Hilbert transform to estimate changes in respiratory volume per unit time from respiratory bellows recordings. The article explains how this approach leads to a representation with higher time resolution and better captures atypical breathing events than current peak based RVT estimators. Here RVT stands for Respiratory volume per unit time. The methodology that the authors followed begins with preprocessing respiratory bellows recordings to eliminate high-frequency noise. The Respiratory Volume Timecourse (RVT) is then derived using the Hilbert transform to analyze the respiratory waveform. This RVT is synchronized with fMRI data by resampling, allowing extraction of respiratory volumes at each TR. Next, the RVT is combined with other physiological signals, like the cardiac pulse, and included as nuisance regressors in a GLM for functional connectivity analysis. This step enhances noise correction and eliminates respiratory-related artifacts in fMRI data. A significant advantage lies in the methodology's high time resolution, achieved through the Hilbert transform, compared to peak-based RVT estimators. It captures detailed breathing rhythms, including atypical events overlooked by peak-based methods. Lastly, the methodology is openly accessible through the PhysIO package within the TAPAS framework, facilitating its implementation and adoption in research settings. [1].

The next paper is about the study examining the impact of various obstructive sleep apnea treatment methods on autonomic nervous system activity is reported in the article. The study had two participant groups: control people from the normal population who did not report any sleep problems, and those with obstructive sleep apnea who were treated with continuous positive airway pressure. A variety of techniques, including empirical mode decomposition, short-time Fourier transform, and Hilbert-Huang transform, were used to study the physiological data of the participants. The methodology used is two groups in the study were those receiving continuous positive airway pressure (CPAP) treatment for obstructive sleep apnea and a control group that did not have any sleep problems. A huge physiological indicators were observed during nocturnal polysomnography (PSG), such as oxygen saturation, esophageal pressure transducer, thermistor, and nose pressure. For analysis, PSG data that met the AASM scoring requirements were anonymized.

Empirical mode decomposition (EMD), short-time Fourier transform (STFT), and Hilbert-Huang transform (HHT) were used to analyze the physiological signals, with a focus on autonomic responses to arousal brought on by nasal flow limitation levels. A preliminary analysis including 15 participants indicated that a sample size of 23–49 persons was necessary. R(version 2.15.2) was used for the statistical analysis, and ANOVA with the Tukey-HSD test was used to compare continuous variables ( $p < 0.05$  signifying significance).[2].

The third paper discusses a study that looked at stomach motility in healthy persons by analyzing electrogastrography (EGG) signals. The study evaluates the EGG signals during pre-prandial (fasting) and post-prandial states, as well as the variations between EGG signals in people who are male and female. The paper describes the techniques utilized to evaluate EGG signals, such as time-frequency analysis and multi-scale decomposition. The paper highlights the limitations and applications of the empirical mode decomposition (EMD) method for



biomedical signal processing, as well as how to employ its adaptive, data-dependent basis for multi-scale decomposition.

The study's conclusions are finally presented in the article, along with the variations and parallels between the EGG signals of the male and female individuals as well as the variations between the EGG signals during fasting and the postprandial phases. The methodology followed as follows, the study utilized electrogastrography (EGG) signals to explore gastric motility, a 90-minute examination where participants consumed 400 ml of cold water to stimulate the stomach. EGG signals, denoted as signal 1 to 8, were recorded using a 4-channel biological amplifier with standard electrode configuration, divided into pre-prandial and two post-prandial 30-minute segments.

Standardization of EGG signals involved dividing the time series values by the channel's standard deviation (StD) using Matlab and multi-ensemble empirical mode decomposition (MEMD). Harmonic ratios were calculated using Hilbert-Huang spectrum (HHSPECTRUM), TOIMAGE, and DISPHHS functions on standardized signals. Bradygastic, normogastic, and tachygastic rhythm indices were computed for 1-minute segments with 10-second overlaps using EGG-DWPACK. The study shows the use of empirical mode decomposition (EMD) in signal analysis to decompose non-stationary signals, capturing specific physical attributes of non-linear signals, including the interpolation and extrapolation of local extrema.

Analysis of EGG signal characteristics during pre- and post-prandial periods, along with gender differences, was conducted using statistical methods such as Mann–Whitney U test, false discovery rate, and Student t-test. [3].

The fourth paper is about the method to detect and analyze intention tremor in patients with multiple sclerosis (MS). The method is based on the Hilbert-Huang transform (HHT) and provides an automated way to extract tremor without any prior knowledge about the dominant frequency of the tremor specific to each subject. The advantage of this method is that it allows for a clear separation between tremor and voluntary movement, which is essential to study the multiple components of MS tremor. The article also discusses the performance of the proposed HHT-based method and its comparison to more conventional methods based on the Fourier transform and traditional band-pass filters. The authors conclude that the proposed method can discriminate between healthy controls and MS subjects with intention tremor and to detect subtle alterations in MS subjects without a clinically detected tremor. They also demonstrate the validity of the proposed indexes for quantifying MS intention tremor and suggest further investigations to improve the proposed procedure.

The methodology that the authors followed are as follows, Data Acquisition: Participants held a probe while accelerometer data were recorded for 30 seconds in a standing posture, capturing 3-axis acceleration at 50 Hz frequency. Preprocessing: Raw acceleration data gone through preprocessing to remove the DC component and apply a high-pass filter (0.1 Hz cutoff) for eliminating low-frequency noise.

HHT-based Method: Using the Hilbert-Huang transform (HHT), the tremor signal was extracted from pre-processed data. This method involves empirical mode decomposition (EMD), interpolation of local maxima and minima, computation of mean envelope and mode amplitude, estimation of candidate IMFs, and iterative sifting and residual correction. Performance Evaluation: The efficacy of the HHT-based method was assessed by comparing the extracted tremor signal with clinical assessments by neurologists.

Results: The proposed HHT-based method effectively differentiated between healthy controls and MS patients with intention tremor. It also detected subtle tremor alterations in MS patients without clinically detected tremor. The method's validity for quantifying MS intention

tremor was demonstrated, and it was compared to traditional band-pass filters and other nonlinear signal decomposition methods.[4].

The next article is about the analysis of physiological signals using Hilbert-Huang Transform (HHT) method. The method uses empirical mode decomposition (EMD) to decompose signals into different intrinsic mode functions (IMFs), enabling better time-frequency resolution and a more accurate spectrum. The method can sift out intrinsic undulations at different time scales and preserve the time-varying properties instead of viewing the signal as a summation of infinite sinusoidal oscillations of constant amplitudes and frequencies. The article also describes the study design and protocols of a retrospective study performed on anonymous polysomnography (PSG) data approved by the Stanford University Medical Institutional Review Board. PSG data includes recordings of four EEG leads, electrooculograms, chin and leg EMGs, ECG lead, nasal cannula pressure-transducer, mouth thermistor, thoracic and abdominal inductive-plethysmography bands, finger oxygen saturation, neck microphone, and esophageal pressure transducer. The study involved subjects with obstructive sleep apnea and control subjects with normal body habitus and normal clinical evaluation. The study reports that HHT is an effective method for analyzing physiological signals with high temporal resolution.

The authors followed a methodology of using the Hilbert-Huang Transform (HHT) for physiological signal analysis. This approach uses a data-driven empirical mode decomposition (EMD) algorithm to break down signals into intrinsic mode functions (IMFs), preserving their time-varying properties. This contrasts with traditional methods that view signals as a summation of constant amplitude and frequency sinusoidal oscillations.

The HHT method involves several steps: first, decomposing the signal into IMFs using the adaptive EMD algorithm, which identifies and extracts various oscillatory modes within the signal. These IMFs represent oscillatory components across different time scales. Next, the instantaneous phase and frequency of the IMFs are obtained using the Hilbert transform, transforming the signal into an analytic form. Finally, the Hilbert spectrum is constructed, plotting the instantaneous frequency and energy of the IMFs over time.

The study's design and protocols detail a retrospective analysis conducted on anonymous polysomnography (PSG) data, involving individuals with obstructive sleep apnea and control subjects with normal clinical evaluations and body types. PSG data encompassed multiple physiological recordings, including EEG leads, EMGs, ECG leads, and various pressure and temperature sensors. The study utilized the HHT method to analyse these physiological signals, offering high temporal resolution. [5].

### **3. Simulation of the Ultrasound waves and detection of the Echoes in Echogram virtually using LabVIEW**

Using the LabVIEW software, the generation of Ultrasound waves is achieved in a virtual manner by assigning a specific range for both the amplitude and frequency. This range is pertinent to the medical applications where Ultrasound waves are utilized. During the execution of the program, the user has the capability to virtually modify the amplitude and frequency of the Ultrasound waves within the predetermined range. Given that these Ultrasound waves are reflected from organs or tissues, there is a subsequent delay in the echoes. Furthermore, some of these echoes experience decay over time. These phenomena are referred to as echo delay and decay, respectively. The user also has the ability to adjust these parameters. The resulting signal, which is displayed in a graph illustrating amplitude versus time, is contingent upon the values assigned to each parameter. To determine the position and location of the Echoes, the Hilbert transform is employed. This transform concatenates the

original signal value with the delayed signal value. The signal values are collected using the Collector module and the amplitudes and other measurements are displayed using the Amplitude and Level Measurement module with the help of excel sheet which is integrated to the LabVIEW software.

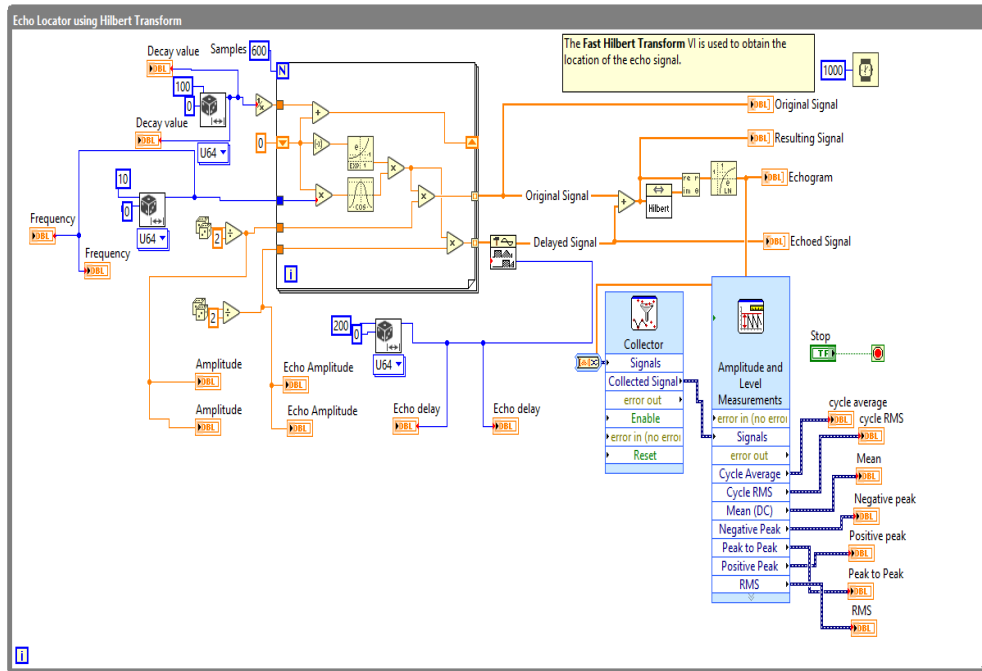


Fig. 1. Block diagram of the circuit which produce the ultrasound virtually and the echoes are displayed in the echogram.

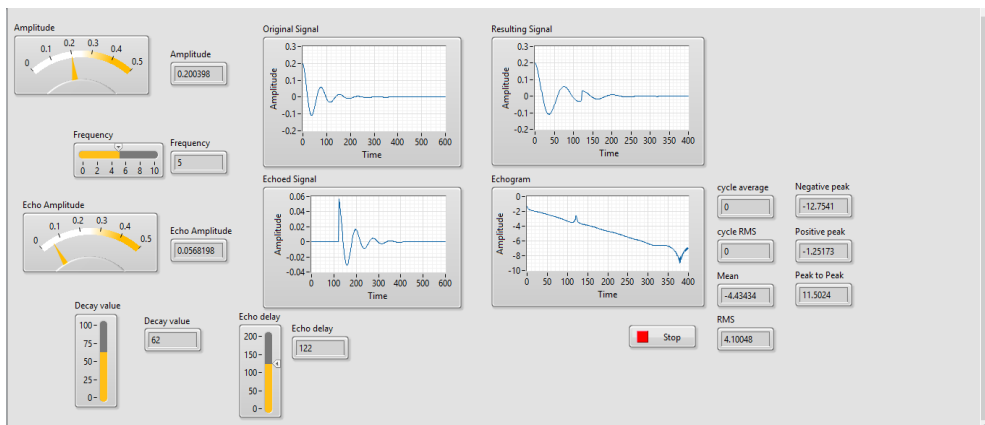


Fig. 2. Front Panel of the LabVIEW software depicting the working of the circuit, which produces ultrasound signals based on the user input, the Original, Echoed, resulting signal were displayed in the graph, and the Echogram displays the value and position of the echoes, based on which some of the values like Mean, RMS, positive and Negative peak is displayed

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Time	cycle average	cycle RMS	Mean	Negative peak	Positive peak	Peak to Peak	RMS					
2	13-12-2023 12:34:43.804	0	0	-3.757677	-9.146209	-0.579679	8.56653	3.575623					
3	13-12-2023 12:34:45.354	-4.674222	4.920158	-4.230944	-9.946608	-0.948789	8.997819	3.971524					
4	13-12-2023 12:34:47.006	0	0	-4.72047	-12.85641	-0.68865	12.16776	4.745134					
5	13-12-2023 12:34:48.827	0	0	-2.992119	-12.85641	-0.822916	12.033494	2.785945					
6	13-12-2023 12:34:49.749	-3.978586	4.177403	-3.343156	-5.806063	-0.616626	5.189436	3.141355					
7	13-12-2023 12:34:50.962	0	0	-5.663768	-12.106829	-0.841403	11.265425	5.788491					
8	13-12-2023 12:34:52.726	-5.66295	5.877945	-5.028689	-12.631559	-0.818004	11.813555	4.911063					
9	13-12-2023 12:34:54.769	0	0	-3.896224	-12.631559	-1.172299	11.459259	3.628408					
10	13-12-2023 12:34:57.326	0	0	-6.19263	-15.188448	-0.81059	14.377858	6.305754					
11	13-12-2023 12:34:59.428	0	0	-7.33185	-18.250962	-0.889459	17.361503	7.376306					
12	13-12-2023 12:35:01.787	0	0	-5.252907	-18.250962	-1.326345	16.924617	4.923472					
13	13-12-2023 12:35:03.239	0	0	-4.701984	-11.112481	-0.592236	10.520245	4.389667					
14	13-12-2023 12:35:04.781	0	0	-3.416674	-9.66149	-0.745552	8.915938	3.177553					
15	13-12-2023 12:35:06.166	0	0	-3.32239	-8.381444	-0.938803	7.44264	3.278268					
16													
17													
18													
19													
20													
21													
22													

Fig. 3. The data collected from the Echogram, with respect to time. As the user assigns and stimulates the parameters, the values also change with respect to time.

## 4. Conclusion

The detection of the echoed signal is one of the important factors when comes to image processing, and other biomedical applications. So, this Circuit which is developed using the LabVIEW software helps in the development of the ultrasound signals virtually and further analyze the echoes and resulting signal virtually, so that it will be easier in real time applications, also the data can be collected and stored in excel sheet, so that it can be used for future purposes and research activities.

## References

1. S. J. Harrison, S. Bianchi, Jakob Heinzle, Klaas Enno Stephan, S. Iglesias, and L. Kasper, "A Hilbert-based method for processing respiratory timeseries," *NeuroImage*, vol. 230, pp. 117787–117787, Apr. 2021, doi: <https://doi.org/10.1016/j.neuroimage.2021.117787>.
2. M. A. Prucnal and A. G. Polak, "Analysis of Features Extracted from EEG Epochs by Discrete Wavelet Decomposition and Hilbert Transform for Sleep Apnea Detection," *PubMed*, Jul. 2018, doi: <https://doi.org/10.1109/embc.2018.8512201>.
3. D. Komorowski and B. Mika, "Gastric slow wave rhythm identification using new approach based on noise-assisted multivariate empirical mode decomposition and Hilbert–Huang transform," *Neurogastroenterology & Motility*, Oct. 2020, doi: <https://doi.org/10.1111/nmo.13997>.
4. I. Carpinella, D. Cattaneo, and M. Ferrarin, "Hilbert–Huang transform based instrumental assessment of intention tremor in multiple sclerosis," *Journal of Neural Engineering*, vol. 12, no. 4, p. 046011, Jun. 2015, doi: <https://doi.org/10.1088/1741-2560/12/4/046011>.
5. C. Lin, Men Tzung Lo, and C. Guilleminault, "Exploring the Abnormal Modulation of the Autonomic Systems during Nasal Flow Limitation in Upper Airway Resistance Syndrome by Hilbert–Huang Transform," *Frontiers in Medicine*, vol. 4, Sep. 2017, doi: <https://doi.org/10.3389/fmed.2017.00161>.

# Wearable Tremor-Absorbing Device for Parkinson's Disease Patients

**Praviene P., Sivakumar R.**

Vellore Institute of Technology, Katpadi, Vellore – 632014, India  
*praviene.p2023@vitstudent.ac.in, rsivakumar@vit.ac.in*

**Abstract.** In the current years, the count of people affected by Parkinson's disease keeps on increasing. This may be due to change in the lifestyle, hypertension and many more reasons. These patients usually have rigid muscles, tremors, bradykinesia, speech and writing alterations which may cause difficulty in them. These patients seek assistance for carrying out their daily activities as they have tremors in their limbs. Recently there is a finding of such a device that could just absorb the tremors generated. In this study, a device was simulated using LabVIEW software for these patients that can sense and absorb the tremors by using any vibration-absorbing material like rubber or elastomer pads thereby making their limbs steady with the application of pressure which can also alert them to seek medical attention if the pressure applied is higher. In some cases, suspensions can be employed to withstand heavy tremors which can be attached at any place where the tremors are felt. The tremor level gets reduced enabling the patients to perform their routine activities.

**Keywords:** Parkinson's disease, tremors, rigidity, rubber, elastomer pads, LabVIEW.

## 1. Introduction

Parkinson's Disease (PD) is a condition that predominantly affects the brain causing movement, mental health, sleep sleep-related problems. The condition of these patients worsens over time. Still, no cure has been discovered for this condition, but therapies and medicines can minimize the symptoms. The most common symptoms are tremors, rigidity, and speech problems [1].

In the previous 25 years, the rate of prevalence of PD has doubled. In the year 2019, over 8.5 million individuals were found to be with PD conditions. Since the year 2000, there has been a 100% increase in the number of PD cases globally [2].

The exact cause of the PD is not known till now. However, many researchers believe it to be inherited or the cause may be linked to environmental factors. Substantia nigra and corpus striatum are the two regions present in the brain between which the dopamine transmits the signals. Due to the loss of cells that produce dopamine, the PD condition occurs.

As the condition progresses, it degenerates many regions of the brain and the nervous system thereby contributing to many problems like tremors, rigidity, dementia, impairment in cognitive skills, etc. [3].

People with this PD condition find it difficult to carry out their daily routine. Hence, a device was simulated for them using LabVIEW software which could absorb the tremors as well as loosen the rigid muscle enabling them to move their limbs easily.

## **2. Effect of tremors in day-to-day life**

The International Parkinson and Movement Disorder Society (IPMDS) task group defined tremors as an uncontrollable, synchronized, oscillatory movement of a bodily component. Regarded as the most common pathological tremor in adulthood, essential tremor (ET) affects around 0.9% of the world's population. The real prevalence of ET, however, could be greater because it's thought that some people would not seek medical care. Often distributed asymmetrically, tremors are commonly observed in Parkinson's disease (PD) patients [4].

Tremors affect many facets of the patient's everyday life and make it difficult for them to engage in a variety of physical tasks at work and at home. Furthermore, tremor sufferers may experience psychological distress that is noticeably greater than their physical impairments. Tremor's psychological effects can go beyond the individuals who are ill [4].

Among the neurological conditions is Parkinson's which often take place beyond the age of sixty. One individual over the age of sixty is diagnosed with Parkinson's disease out of every 100. Younger folks, who make up 5% to 10% of patients, are also seen to have this condition. There are two main types of hand tremors: resting and active. The patient has resting tremors while their hands are immobile. This tremor often occurs at a frequency of three to six hertz. However, people with essential tremor frequently have action tremor, which happens when they voluntarily motion. This tremor normally occurs at a frequency of 5 to 12 hertz. Patients with Parkinson's disease frequently struggle to eat because the spoon or a meal fragment may spill if it can't reach the mouth because to hand trembling. The patient's social interactions are gradually restricted by these tremors, which are frequently humiliating [5].

Weighted utensil use as the mainstay of lifestyle therapies can help patients with their activities of daily living (ADLs) and lessen the amplitude of their tremor. When combined with extra weights, certain tools (such as a spoon) can help patients to consume food and liquids. Guidelines for the treatment of Parkinson's disease (PD) in adults were released in 2017 by the National Institute for Health and Care Excellence (NICE). Physiotherapy and occupational therapy may be beneficial for patients in the early stages of Parkinson's disease (PD) if they exhibit motor symptoms or struggle with ADLs. However, the guidelines developed by the IPMDS and the American Academy of Neurology (AAN) did not address lifestyle or nonpharmacological management of ET [4].

## **3. Existing treatment methods and their drawbacks**

Various tremors are frequently controlled and managed with medication; however, patient reactions to treatment vary. In addition, adverse effects from medications may include weariness, lightheadedness, paralysis of the muscles, and impaired eyesight. Another option for managing tremors is brain surgery, however this is a more intrusive procedure that is frequently suggested when there is not enough of a response to medicines.

These days, a lot of researchers are looking at non-invasive, less intrusive therapeutic choices. Stabilizers built into modern technology maintain gadgets steady even when the user shakes their hand. Thus, scientists have made the decision to create useful gadgets that would completely eradicate hand tremors in Parkinson's sufferers using this technology in conjunction with other control techniques. They help patients carry out their regular tasks as a result [5].

#### 4. Simulation of tremor absorbing device using LabVIEW

A device was simulated using the LabVIEW software for the patients with Parkinson's Disease. The patients can wear the device on the site of tremor which generally absorbs the vibrations generated by the tremor enabling them to carry out their activities of daily living (ADL).

The device possesses a light weight, skin-friendly rubber sheet or elastomer pad which could absorb the tremors and ease the movement of limbs. The device also possesses a pressure monitoring system which works when the patients are having rigid muscles. The pressure is exerted on the muscles responsible for action when the rigidity is more thereby reducing the difficulty of the patients.

When the pressure applied is more than the optimal level in case of rigid muscles, a warning light will glow stating that the person wearing the device should seek medical attention.

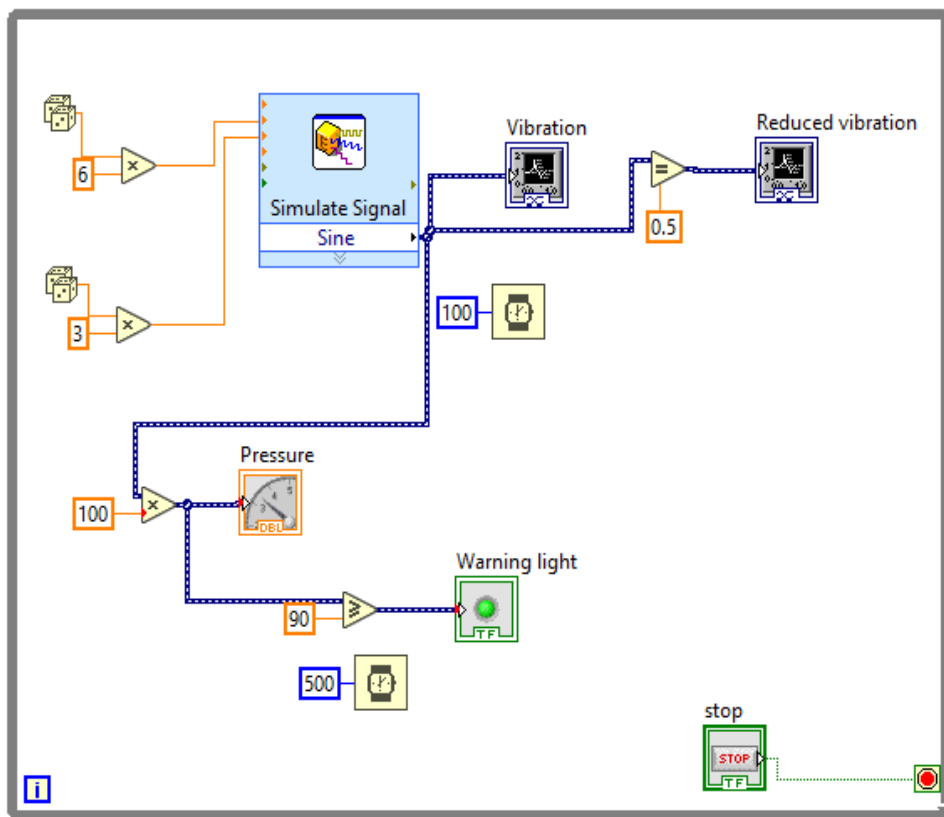


Fig. 1. Block diagram of the circuit of the tremor absorbing device which includes the vibration absorber, pressure monitoring system and warning system.

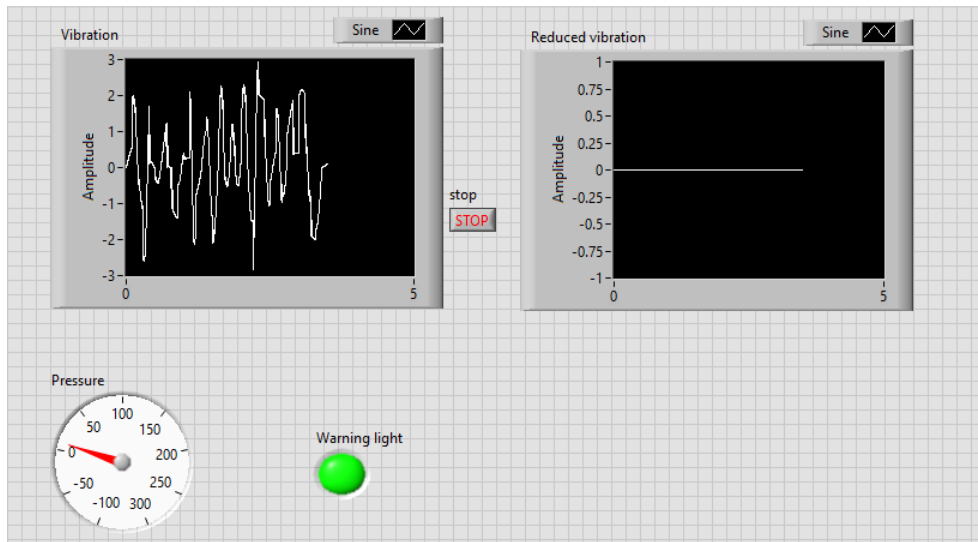


Fig. 2. Front panel of LabVIEW software depicting the working of the device as the tremors are absorbed thereby minimizing the tremors along with the pressure monitoring system and the warning light.

## 5. Conclusion

Most of the treatment strategies available currently have their own advantages and disadvantages. Most of the people could not undergo these treatment procedures as they are not cost effective. Consuming medicines and tablets may cause some side-effects in PD patients. Hence by considering all these factors an affordable device was simulated using the LabVIEW software which will be helpful for the PD patients in carrying out their daily activities.

## References

1. "Parkinsons-Disease-Pathophysiology @ www.news-medical.net." [Online]. Available: <https://www.news-medical.net/health/Parkinsons-Disease-Pathophysiology.aspx>
2. "Parkinson-Disease @ Wwww.Who.Int." [Online]. Available: <https://www.who.int/es/news-room/fact-sheets/detail/parkinson-disease>
3. "Parkinsonsdisease @ stanfordmedicine25.stanford.edu." [Online]. Available: <https://stanfordmedicine25.stanford.edu/the25/parkinsonsdisease.html>
4. J. Mo and R. Priefer, "Medical Devices for Tremor Suppression: Current Status and Future Directions," *Biosensors*, vol. 11, no. 4, 2021, doi: 10.3390/bios11040099.
5. F. Talaei and S. M. Kargar, "Design and Fabrication of a Device for Reducing Hand Tremor in Parkinson Patients during Eating," *J. Med. Signals Sens.*, vol. 13, no. 1, pp. 21–28, 2023, doi: 10.4103/jmss.jmss\_116\_21.



# Piezoelectric Nanogenerator Based Smart Watch

**Guhan C., Sivakumar R.**

Department of Sensor and Biomedical Technology, Vellore Institute of Technology, India  
guhan.c2023@vitstudent.ac.in, [rsivakumar@vit.ac.in](mailto:rsivakumar@vit.ac.in)

**Abstract.** The National Instruments' LabVIEW is a powerful graphical programming environment for creating measurement, automation, and control systems. This paper describes a wristwatch powered by a piezoelectric nanogenerator that uses LabVIEW software to combine several health and environmental monitoring components. SpO<sub>2</sub>, heart rate, blood pressure (BP), menstrual cycle tracking, sweat-based calorie monitoring, hydration level, and temperature sensing are all included in the wristwatch. LabVIEW makes it easier to design and visualise the watch's circuitry, resulting in a more complete and efficient system. The piezoelectric nanogenerator technology integrated into the watch converts mechanical energy from wrist motions into electrical power, assuring long-term functioning. LabVIEW serves as the circuit design platform, enabling for the smooth integration of various sensors and energy harvesting components. The research looks on the relationship between LabVIEW and piezoelectric nanogenerator technology in order to optimise energy extraction and distribution throughout the watch's capabilities. It assesses the usefulness of real-time health monitoring by analysing the precision and dependability of health data generated by the integrated sensors. Furthermore, the graphical programming features of LabVIEW make it easier to create a complete circuit design for the wristwatch, showing the links between the numerous components, sensors, and piezoelectric nanogenerators. This study shows how to build a multifunctional wristwatch driven by piezoelectric nanogenerators, using LabVIEW for smooth integration and effective energy management. The study emphasises the potential for long-term wearable technology that may provide complete health monitoring and environmental sensing. This breakthrough is a big step towards self-powered, multifunctional smartwatches that provide a comprehensive approach to health tracking and environmental awareness.

**Keywords:** Labview, Piezoelectric Nanogenerators, Heart rate, SpO<sub>2</sub>, menstrual cycle, Blood pressure, sweat, calorie expenditure, wearable device

## 1. Introduction

In the ever-changing world of wearable technology, the combination of improved sensors and efficient energy harvesting mechanisms has led to a new age of innovation in health monitoring and environmental sensing. National Instruments' LabVIEW software emerges as a critical tool in this arena, enabling the smooth integration of complex systems. This study explores the combination of LabVIEW with piezoelectric nanogenerator technology to create a multipurpose wristwatch capable of real-time health monitoring, environmental awareness, and even menstrual cycle tracking.

The piezoelectric nanogenerator, a cutting-edge innovation capable of transforming mechanical energy from wrist motions into electrical power. This self-sustaining power source offers long-term performance for the wristwatch, eliminating the need for regular external charging and representing a substantial advancement in wearable technology. LabVIEW

serves as the foundation of this effort, providing an easy graphical programming environment that simplifies the integration of various health and environmental monitoring components.

The wristwatch can measure SpO<sub>2</sub>, heart rate, blood pressure (BP), and temperature, as well as menstrual cycle tracking. LabVIEW's graphical interface makes circuit design easier while maximizing energy extraction and distribution. Researchers guarantee that real-time health data is precise and reliable through rigorous testing.

Furthermore, LabVIEW's graphical programming tools assist the building of a complete circuit design for the timepiece, illuminating the complicated connections between various components, sensors, and piezoelectric nanogenerators. This study introduces a novel paradigm in wearable technology: a self-powered, multipurpose wristwatch powered by piezoelectric nanogenerators and connected with LabVIEW software. This breakthrough stretches beyond gadgetry, delivering a look into the future of wearable technology for complete health monitoring and environmental awareness.

## 2. Methodology

The integration of cutting-edge technology with health monitoring has resulted in the creation of multifunctional smartwatches that can monitor a variety of physiological markers. National Instruments' LabVIEW is a robust graphical programming environment that enables the seamless integration of health and environmental monitoring components into wearable devices. This research investigates how LabVIEW software and piezoelectric nanogenerator technology may work together to create a wristwatch that can measure SpO<sub>2</sub>, heart rate, blood pressure, menstrual cycles, temperature, and other parameters. The watch's piezoelectric nanogenerator technology uses mechanical energy from wrist motions to provide long-term performance without the need for external power sources. LabVIEW's graphical programming tools ease the design and visualization of the watch's circuitry, allowing for effective integration of sensors and energy harvesting components. The study uses real-time health monitoring to evaluate the quality and reliability of data supplied by integrated sensors, stressing the potential for self-powered wearable technology to transform health tracking and environmental awareness. This result is a big step toward complete health monitoring and demonstrates the revolutionary power of LabVIEW-enabled smartwatches on personal well-being. We must underline that the economic and social sides of business interact in the company's marketing activities. As a sophisticated instrument for attaining business goals via exchange, marketing is in charge of two fundamental issues: generating strong company brands and fostering good impressions of the organization as a whole in society. To overcome these complicated difficulties, marketers frequently seek assistance from professionals outside of the marketing sector. The development of sophisticated CRM systems (Customer Relationship Management Systems) is one step in this direction.

### 2.1 SpO<sub>2</sub> Level Monitoring

LabVIEW uses the formula,

$$\text{SpO}_2 = ((-25.4) * (x_1/x_2)) + 110$$

to determine SpO<sub>2</sub> values from sensor data. This formula is essential for determining blood oxygen saturation levels and providing significant information about respiratory health. LabVIEW analyzes SpO<sub>2</sub> data in real time and displays it on the wristwatch interface, allowing users to monitor their oxygen levels easily.

Provide findings and comments for the aforementioned content.

## **2.2 Cyanosis Detection**

Algorithms for detecting low oxygen saturation levels, which indicate cyanosis, are created by analyzing SpO<sub>2</sub> data in LabVIEW. When cyanosis is identified, LabVIEW sends alarms or messages to the wristwatch, ensuring prompt knowledge of potentially serious health issues.

## **2.3 Hypoxemia Detection**

Hypoxemia (abnormally low oxygen levels) is recognized by continuously monitoring SpO<sub>2</sub> levels and using threshold-based algorithms in LabVIEW. When hypoxemia is detected, LabVIEW sends visual or audio notifications to the wristwatch interface, allowing for timely action or medical assistance.

## **2.4 Menstrual Cycle Tracking**

By integrating menstrual cycle tracking sensors or algorithms with LabVIEW, users may directly enter menstrual cycle data. The LabVIEW interface on the wristwatch allows for straightforward tracking of menstrual cycles based entirely on user input data. LabVIEW estimates the average duration between the start of one monthly period and the start of the next by evaluating the user-provided length of the menstrual cycle. This method helps to forecast fertility windows and ovulation dates, which improves reproductive health management and allows users to make more educated family planning decisions.

## **2.5 Temperature Sensing**

Temperature sensors connected to LabVIEW monitor ambient or body temperature, and LabVIEW algorithms translate sensor readings into temperature measurements. The wristwatch interface shows real-time temperature data and sends alarms for high or low temperatures based on predetermined thresholds, assuring users safety and comfort.

## **2.6 Heart Rate Measurement (Systole and Diastole)**

Heart rate, systolic and diastolic blood pressure are monitored and processed using photoplethysmography (PPG) or ECG sensors that are connected to LabVIEW. LabVIEW's algorithms extract these crucial statistics and present them in real time on the wristwatch interface, allowing for full cardiovascular health monitoring. Systolic and diastolic blood pressure measurements are calculated using LabVIEW's formula

$$\text{Blood Pressure} = (2/3) * \text{diastole} + (1/3) * \text{systole}$$

This estimate is based on sensor data acquired in real time, allowing for continuous monitoring and presentation of these critical cardiovascular parameters on the wristwatch interface.

## **2.7 Tachycardia Detection**

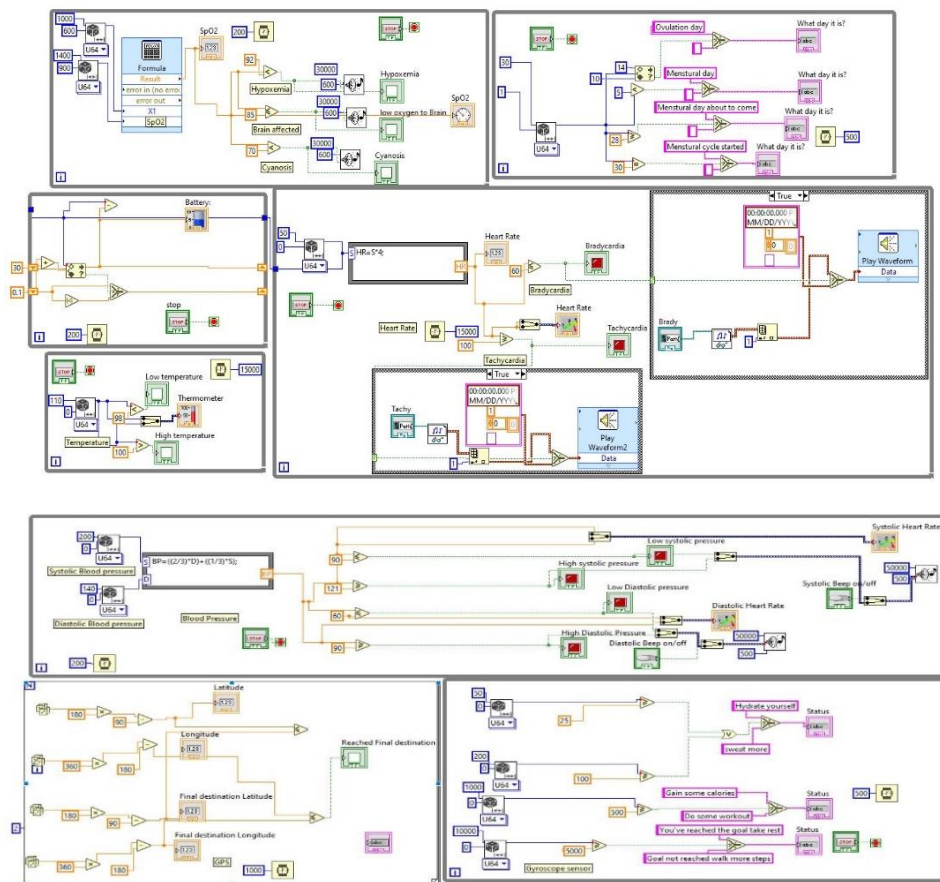
LabVIEW algorithms evaluate heart rate data to identify tachycardia (an increased heart rate). When tachycardia is detected, LabVIEW sends a warning to the wristwatch interface, ensuring that aberrant heart rate circumstances are addressed immediately for preventative or therapeutic treatments.

## 2.8 Bradycardia Detection

LabVIEW algorithms evaluate heart rate data to detect bradycardia, much as they do for tachycardia. Bradycardia recognition triggers alerts on the wristwatch interface, allowing users with slow heart rates to get prompt assistance or medical advice.

## 2.9 GPS Tracking

LabVIEW connects with GPS modules to gather latitude and longitude data, constantly monitoring the user's current and destination coordinates. The wristwatch interface shows real-time location monitoring and navigation help, which improves user safety and convenience while outdoor activities or travel.



## 2.10 Fitness Insight

The incorporation of hydration warnings and training goals, together with LabVIEW's smooth data synchronization, provided consumers with full fitness tracking capabilities. LabVIEW handled data efficiently, correctly monitoring changes in hydration levels and providing real-time updates on calories burned. This comprehensive approach to fitness tracking improves overall well-being by allowing users to make educated hydration and exercise decisions based on their own health goals and needs.

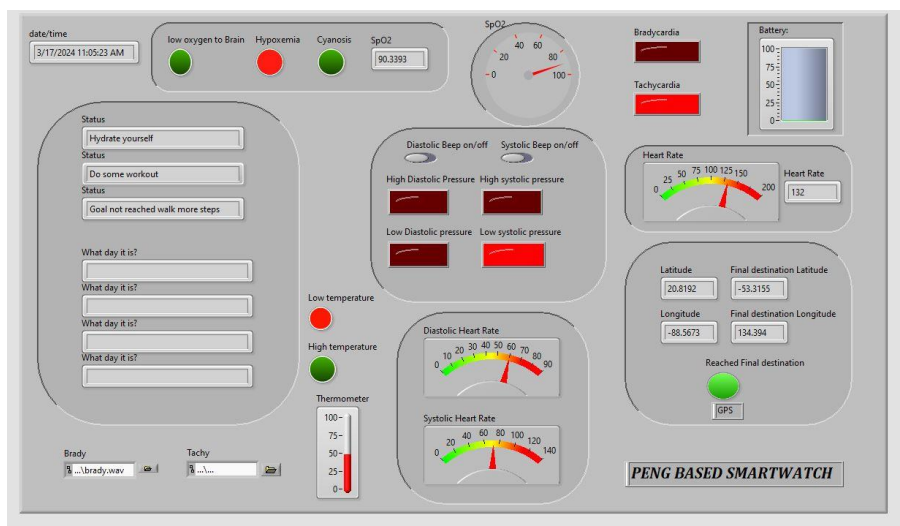
## 2.11 Battery Level Monitoring

Battery sensors or monitoring circuits interfaced with LabVIEW allow for voltage tracking and estimate of remaining battery capacity. LabVIEW shows battery level information on the wristwatch interface and sends low battery notifications, guaranteeing continuous functioning and prompt recharging of the wearable.

## 3. Results and Discussion

The combination of numerous health monitoring elements with LabVIEW on smartwatches results in complete tracking and analytical capabilities, which contribute considerably to proactive health management. By evaluating SpO<sub>2</sub> data, LabVIEW algorithms successfully detect cyanosis and hypoxemia, alerting users to potentially serious health issues in real time. This capability guarantees that users are immediately aware of changes in oxygen saturation levels, allowing for early actions and medical treatment as necessary. Temperature sensing, aided by LabVIEW, improves user safety by sending out notifications for severe temperature situations. By continually monitoring temperature levels, users may take necessary precautions to reduce the hazards associated with temperature swings, enhancing overall health.

LabVIEW's algorithms also provide cardiovascular health monitoring, recognizing tachycardia and bradycardia and alerting users to unhealthy heart rate situations. This proactive method enables users to track their heart health in real time and take precautions to reduce possible hazards. Furthermore, LabVIEW's GPS tracking feature enables users with real-time position monitoring and navigation aid, improving safety during outdoor activities and travel. By seamlessly incorporating GPS data, LabVIEW provides accurate tracking of the user's location and destination coordinates, enabling safe navigation and lowering the danger of becoming lost.



LabVIEW also allows for battery level monitoring, which ensures that smartwatches operate continuously by notifying users to low battery levels. This feature improves user convenience by providing continuous access to health monitoring functions. Furthermore,

LabVIEW allows users to enter menstrual cycle data for fertility tracking and family planning, hence improving reproductive health management.

Furthermore, hydration warnings and calorie burn monitoring, together with LabVIEW's flawless data synchronization, improve fitness tracking capabilities. Users may make educated decisions regarding hydration and exercise by properly monitoring changes in hydration levels and receiving real-time calorie burn updates, encouraging overall well-being and proactive health management. In conclusion, LabVIEW's connection with smartwatches provides users with extensive health monitoring capabilities, encouraging proactive health management and improving overall well-being.

## 4. Conclusion

Finally, the combination of LabVIEW and smartwatches represents a significant advancement in wearable health technology, providing consumers with a comprehensive suite of health monitoring and preventative management options. LabVIEW's powerful algorithms excel at detecting vital health metrics such as cyanosis, hypoxemia, irregular heart rates, and high temperatures, resulting in prompt alarms and interventions to reduce possible health hazards. Furthermore, LabVIEW's seamless data connection allows for accurate tracking of a wide range of health variables, including SpO2 levels, menstrual periods, GPS position, and battery condition, improving user convenience and safety. Furthermore, features like as hydration alarms and calorie burn monitoring encourage a complete approach to fitness tracking, providing users with real-time information about their health state.

## References

1. Li, R., Wei, X., Xu, J., Chen, J., Li, B., Wu, Z., & Wang, Z. L. (2021). Smart wearable sensors based on Triboelectric nanogenerator for personal healthcare monitoring. *Micromachines*, 12(4), 352. <https://doi.org/10.3390/mi12040352>
2. R. Li et al., "Smart Wearable Sensors Based on Triboelectric Nanogenerator for Personal Healthcare Monitoring," *Micromachines*, vol. 12, no. 4, p. 352, Mar. 2021, doi: <https://doi.org/10.3390/mi12040352>.
3. Zhang, M. et al. (2015) 'A hybrid fibers based wearable fabric piezoelectric nanogenerator for energy harvesting application', *Nano Energy*, 13, pp. 298–305. doi:10.1016/j.nanoen.2015.02.034.
4. Dahiya, A. S. et al. (2017) 'Organic/inorganic hybrid stretchable piezoelectric nanogenerators for self-powered wearable electronics', *Advanced Materials Technologies*, 3(2). doi:10.1002/admt.201700249.
5. Wang, B. et al. (2021) 'Woodpecker-mimic two-layer band Energy Harvester with a piezoelectric array for powering wrist-worn wearables', *Nano Energy*, 89, p. 106385. doi:10.1016/j.nanoen.2021.106385.
6. Zhao, Z. et al. (2021) 'Flexible nanogenerators for wearable electronic applications based on piezoelectric materials', *Materials Today Energy*, 20, p. 100690. doi:10.1016/j.mtener.2021.100690.
7. Huang, T. et al. (2015) 'Human walking-driven wearable all-fiber triboelectric nanogenerator containing electrospun poly vinylidene fluoride piezoelectric nanofibers', *Nano Energy*, 14, pp. 226–235. doi:10.1016/j.nanoen.2015.01.038.
8. An, T. et al. (2020) 'Self-powered gold nanowire tattoo triboelectric sensors for Soft Wearable human-machine interface', *Nano Energy*, 77, p. 105295. doi:10.1016/j.nanoen.2020.105295.
9. Sun, W. et al. (2023) 'A self-powered multifunctional bracelet for pulse monitoring and personal rescue', *Biosensors*, 13(5), p. 552. doi:10.3390/bios13050552.
10. Tan, P., Yuan, X., Chao, S., Jiang, D., Liu, Z., Fan, Y., & Li, Z. (2022). An Artificial Intelligence-Enhanced Blood Pressure Monitor wristband based on piezoelectric nanogenerator. *Biosensors*, 12(4), 234. <https://doi.org/10.3390/bios12040234>

# Stroke Gait Analysis Using IMU Sensors

**Gaadha Jeevan, Sivakumar R.**

Vellore Institute of Technology, Vellore -632014, India  
*gaadha.jeevan2023@gmail.com, rsivakumar@vit.ac.in*

**Abstract.** This paper examines how data from Inertial Measurement Unit (IMU) sensors can reveal insights regarding the walking patterns of individuals suffered from stroke. The study employed analysis techniques like Frequency Spectrum and Sample Entropy to thoroughly analyze the characteristics of stroke gait. Classifiers like Support Vector Machine (SVM) and ResNet-18 are used to differentiate between stroke gait and healthy gait patterns. Additionally, a raw stroke gait signal classification approach using K-Means clustering is investigated. The paper highlights the problems caused by limited datasets, which could affect how accurately the data can be classified. Even with these challenges, the findings give useful information about how well different classification techniques work, which could help improve stroke rehabilitation strategies.

**Keywords:** IMU sensors, Stroke Gait, Frequency Spectrum, Sample Entropy.

## 1. Introduction

Recently, gait analysis has become an essential tool in comprehending different neuromuscular disorders, such as those that result from a stroke. These disorders commonly manifest as changes in gait patterns, which can greatly hinder mobility and overall well-being for stroke survivors. As a result, precise and timely evaluation of these atypical gait traits is imperative for creating effective rehabilitation plans and tracking progress. [3]

Thanks to the introduction of Inertial Measurement Unit (IMU) sensors in recent years, the landscape of gait analysis has been completely transformed, providing a convenient, discrete, and affordable means of gathering motion data. IMU sensors, also known as Inertial Measurement Units, are compact but powerful devices that are essential in recording and interpreting movement data. Consisting of accelerometers, gyroscopes, and magnetometers, IMUs accurately track changes in acceleration, angular velocity, and magnetic field strength. Accelerometers precisely capture linear movement, gyroscopes monitor rotational motion, and magnetometers determine orientation in relation to the Earth's magnetic field. By working together, these sensors make it possible to calculate the precise position and orientation of an object in three-dimensional space. [1]

From screen rotation on mobile devices to navigation in aerospace, IMU sensors are widely used in various applications. They have become crucial in industries such as robotics, virtual reality, and gait analysis, providing valuable insights into human movement. By utilizing a combination of accelerometers, gyroscopes, and magnetometers, IMU sensors offer a wealth of information on a person's body orientation, acceleration, and angular velocity during movement. It is extremely important to precisely calculating joint angles as well as the difficulties posed by external acceleration and integration drift. In order to estimate joint angles in real time, the article suggests a two-stage sensor fusion approach that incorporates accelerometer and gyroscope data. [2]

This comprehensive data allows for a detailed analysis of gait dynamics and has become a valuable tool in measuring various gait parameters, including stride length, step time,

and joint angles. One advantage of IMU sensors is their ability to capture real-time data, making it possible to monitor gait patterns in natural settings instead of being limited to laboratory conditions.

The purpose here is to investigate the crucial role of IMU sensors in gait analysis, particularly in identifying various types of gait abnormalities caused by strokes. These abnormalities can be difficult to accurately assess with traditional methods due to their diverse nature and distinct features. However, by utilizing the advanced capabilities of IMU sensors, a deeper comprehension of the complexities of stroke gait can be achieved, paving the way for targeted interventions by both researchers and clinicians.

## 2. Methodology

**Data Collection:** The study utilized a publicly available dataset gathered from various online sources, which contained recordings from a group of stroke survivors and healthy individuals. The participants' gait patterns were analyzed through carefully controlled walking sessions, resulting in a diverse and representative dataset. The IMU sensors, equipped with accelerometers and gyroscopes, enabled accurate and simultaneous tracking of motion-related data [1].

**Data Pre-processing:** To improve the reliability and detail of our data analysis, we carefully prepared our data. We did this by finding and selecting key features in the signals from the Inertial Measurement Unit (IMU) and by extracting relevant gait cycles for further study. Then, we used the Fast Fourier Transform (FFT) technique to analyze the frequency range of the gait cycles. To enhance the accuracy and depth of our analysis, we meticulously processed our data. This involved identifying and isolating critical characteristics within the signals generated by the Inertial Measurement Unit (IMU). We then extracted relevant gait cycles for further examination. Subsequently, we employed the Fast Fourier Transform (FFT) method to scrutinize the frequency spectrum of these gait cycles. This approach combines FFT analysis (a method that breaks down signals into their frequency components) with sample entropy calculation (a measure of signal complexity) [5] [6].

**Empirical Mode Decomposition:** Obtaining different frequencies present within the data was explored using Empirical Mode Decomposition (EMD) to enhance the evaluation of gait stability. This involved breaking down the signals into simpler components called Intrinsic Mode Functions (IMFs). EMD has proven useful in analyzing gait patterns after a stroke. By breaking down the signals into IMFs, it allows to identify key elements that reflect the complex changes in gait after a stroke [9].

**Feature Extraction for Gait Stability:** The calculation of Gait Stability Index calculated using highest and lowest energy IMFs did not produce expected outcomes due to a number of reasons included limited dataset. Hence, classification using the obtained FFT spectrum was implemented. In stroke gait studies, analyzing movement patterns using Fast Fourier Transform (FFT) is crucial. People recovering from strokes tend to walk differently, with irregularities and uneven steps. FFT lets researchers examine the patterns of these movements in the frequency domain, revealing subtle changes that indicate stroke-related problems. Plus, FFT provides a way to extract valuable info efficiently, which is helpful for using advanced techniques like classification and clustering to analyze the data further [9].

**Classification:** In stroke gait analysis, advanced methods like Support Vector Machines (SVM) and ResNet 18 (a Convolutional Neural Network) were used to categorize gait patterns as either stroke-related or healthy. This approach aimed to identify and differentiate subtle gait features unique to stroke patients compared to those without stroke. Hence, both SVM and ResNet 18 is implemented [7] [8].



### 3. Results

There was considerable difference in the FFT Spectrum of Stroke patients and Healthy patients as observed below.

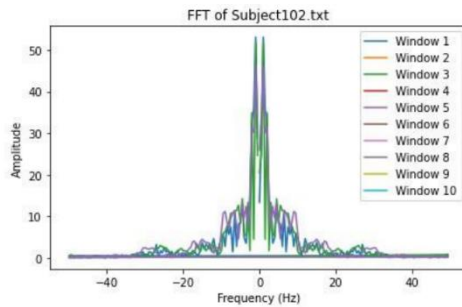


Fig 1. Frequency Spectrum of a Healthy Patient

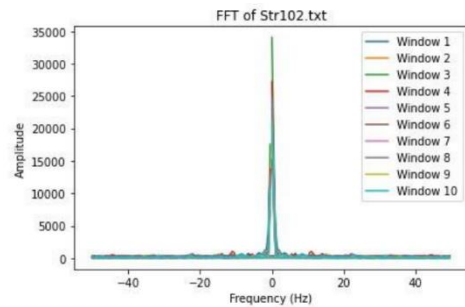


Fig 2. Frequency Spectrum of a Stroke Patient

The EMD implemented for both healthy and stroke affected subjects produced the following results.

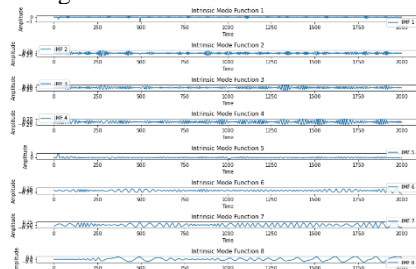


Fig 3. IMFs obtained of Healthy subjects

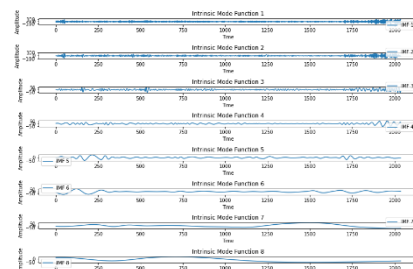


Fig 4. IMFs obtained of Stroke subjects

Classification using frequency features and sample entropy in SVM classification is giving an accuracy of 93.

The results of ResNet18 Classification of FFT Spectrums are given below:

Cross-Validated Accuracy	0.50
Cross-Validated Precision	0.30
Cross-Validated Recall	0.60
Cross-Validated F1 Score	0.40

The cross-validated accuracy of 0.50 suggests that the ResNet-18 model, when evaluated using cross-validation, performs moderately in predicting the class labels of the dataset. The precision of 0.30 indicates that the model's positive predictions have a relatively low level of correctness, meaning that there is a considerable number of false positives.

The recall of 0.60 indicates that the model captures a relatively good proportion of all actual positive instances in the dataset, suggesting it's better at identifying true positives. The F1 score of 0.40 balances both precision and recall, providing an overall measure of the model's performance, which appears to be moderate in this case.

The results of K-means Classification of raw gait signals are given below:

Cross-Validated Accuracy	0.79
Cross-Validated Precision	0.80
Cross-Validated Recall	0.90
Cross-Validated F1 Score	0.79

The high accuracy (0.79) suggests that the K-means clustering algorithm performs well in segregating instances into distinct clusters based on raw signal data. A precision of 0.80 indicates that the clusters generated by K-means are relatively pure, with a majority of instances in each cluster truly belonging to that cluster. The high recall value (0.90) suggests that the clusters effectively capture a large proportion of instances belonging to each class, indicating good completeness in clustering. The F1 score of 0.799 suggests that the clustering process achieves a good balance between precision and recall, reflecting its overall effectiveness in clustering raw signal data.

#### 4. Conclusion

K-means clustering, using raw gait data, provides more accurate results than ResNe18 classification using FFT spectrums. K-means excels in finding distinct patterns in the data, leading to superior clustering. The variation in performance emphasizes the significance of selecting appropriate features and techniques based on the data's characteristics and the task. Future work could explore more features, refine models, or utilize alternative algorithms to improve FFT spectrums classification.

#### References

1. Gujarathi, Trupti, and Kalyani Bhole. "Gait analysis using imu sensor." In 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1-5. IEEE, 2019, John Elkington, Cannibals with Forks: The Triple Bottom Line of 21st Century Business, 1998.
2. Majumder, Sumit, and M. Jamal Deen. "Wearable IMU-based system for real-time monitoring of lower-limb joints." IEEE sensors journal 21, no. 6 (2020): 8267-8275.
3. Kaku, Aakash, Avinash Parnandi, Anita Venkatesan, Natasha Pandit, Heidi Schambra, and Carlos Fernandez-Granda. "Towards data-driven stroke rehabilitation via wearable sensors and deep learning." In Machine Learning for Healthcare Conference, pp. 143-171. PMLR, 2020.
4. John, Joby, and Rahul Soangra. "Visualization-driven time-series extraction from wearable systems can facilitate differentiation of passive adl characteristics among stroke and healthy older adults." Sensors 22, no. 2 (2022): 598.
5. Laudanski, Annemarie, Brenda Brouwer, and Qingguo Li. "Activity classification in persons with stroke based on frequency features." Medical engineering & physics 37, no. 2 (2015): 180-186.
6. Hsu, Wei-Chun, Tommy Sugiarto, Ying-Yi Liao, Yi-Jia Lin, Fu-Chi Yang, Dueng-Yuan Hueng, Chi-Tien Sun, and Kuan-Nien Chou. "Can trunk acceleration differentiate stroke patient gait patterns using time-and frequency-domain features?." Applied Sciences 11, no. 4 (2021): 1541.
7. Wang, Fu-Cheng, Szu-Fu Chen, Chin-Hsien Lin, Chih-Jen Shih, Ang-Chieh Lin, Wei Yuan, You-Chi Li, and Tien-Yun Kuo. "Detection and classification of stroke gaits by deep neural networks employing inertial measurement units." Sensors 21, no. 5 (2021): 1864.
8. O'Brien, Megan K., Sung Y. Shin, Rushmin Khazanchi, Michael Fanton, Richard L. Lieber, Roozbeh Ghaffari, John A. Rogers, and Arun Jayaraman. "Wearable sensors improve prediction of post-stroke walking function following inpatient rehabilitation." IEEE Journal of Translational Engineering in Health and Medicine 10 (2022): 1-11.

9. O'Brien, Megan K., Sung Y. Shin, Rushmin Khazanchi, Michael Fanton, Richard L. Lieber, Roozbeh Ghaffari, John A. Rogers, and Arun Jayaraman. "Wearable sensors improve prediction of post-stroke walking function following inpatient rehabilitation." *IEEE Journal of Translational Engineering in Health and Medicine* 10 (2022): 1-11.
10. Garcia, Flora do Vale, Maira Jaqueline da Cunha, Clarissa Pedrini Schuch, Giulia Palermo Schifino, Gustavo Balbinot, and Aline Souza Pagnussat. "Movement smoothness in chronic post-stroke individuals walking in an outdoor environment—A cross-sectional study using IMU sensors." *PLoS One* 16, no. 4 (2021): e0250100.

# Idiopathic Toe Walking using Machine Learning and Gait Analysis

**Maanasa Mahadevan, Pragadeesh Murugan, Sivakumar R.**

Vellore Institute of Technology, Vellore, Tamil Nadu, India  
*maanasamahaps@gmail.com, pragadeesh.cdps@gmail.com, rsivakumar@vit.ac.in*

**Abstract.** The paper reviews trends in children born after 2000, with Idiopathic Toe Walking. Idiopathic toe walking (ITW) is a gait abnormality in which children's toes touch at initial contact and demonstrate limited or no heel contact throughout the gait cycle. Toe walking results in poor balance, increased risk of falling, and developmental delays among children. A cohort study was performed with nearly 500 patients and 500 controls to observe the differences in the Centre of Force exerted on a force plate through which readings were obtained. The readings were obtained in the form of datasets which were analyzed through various Machine Learning algorithms, like Decision Tree and Random Forest through classification. The result yielded a significant plot which clearly established a decision boundary between the patients and controls with a good accuracy score.

**Keywords:** ITW, Controls, Accelerometer, Gyroscope, Foot plate, Centre of Gravity, Gait analysis.

## 1. Introduction

Idiopathic Toe Walking (ITW) [1] poses a notable challenge in pediatric healthcare, characterized by a distinct toe-to-toe gait pattern observed in children without underlying medical conditions. Traditional assessment methods, reliant on subjective observations, often lack the necessary precision for accurate diagnosis. However, recent strides in wearable sensor technology and machine learning[7] offer promising avenues for a deeper understanding of ITW. These advancements enable researchers and clinicians to decipher subtle patterns associated with ITW, leading to more precise diagnosis and personalized treatment strategies.

This conference paper aims to delve into the intersection of machine learning and gait analysis, specifically focusing on understanding and addressing ITW. It proposes innovative approaches for analyzing gait data collected from wearable sensors, aiming to uncover distinctive features indicative of ITW. Through a meticulous review of existing literature and empirical investigations, the paper evaluates the potential of machine learning-based gait analysis as a valuable tool in the clinical assessment and management of ITW. The goal is to contribute to the advancement of knowledge and the development of effective interventions for this challenging condition.

Furthermore, the significance of gait analysis in comprehending biomechanics is discussed, emphasizing the role of artificial intelligence (AI) in handling complex gait data. While AI holds promise, challenges such as initial poor performance due to data limitations and the necessity for real-time, precise data preservation are acknowledged.

The paper's objectives encompass identifying factors utilized in gait analysis with machine learning, examining recent applications in detecting gait activity and disorders, and assessing commonly used ML methods in performance evaluation and rehabilitation. Additionally, the study explores optimal sensor placements for identifying toe walking in

children with ITW and suggests enhancements in laterality determination using supervised machine learning techniques. Moreover, the paper delves into the utilization of depth cameras for abnormal gait recognition, showcasing experimental results that indicate promise in identifying abnormal gait patterns and classifying specific neurological disorders. Through these concerted efforts, the aim is to enhance diagnostics, interventions, and outcomes for children affected by ITW and other gait abnormalities linked to neurological diseases. By leveraging advancements in technology and machine learning, this research endeavors to provide more effective solutions for addressing the complexities of pediatric gait disorders, ultimately improving the quality of care and outcomes for affected individuals.

## **2. Definition of concepts**

### **Gait analysis**

Gait analysis [2] is a systematic approach used to evaluate human movement, encompassing walking, running, and other forms of locomotion. It involves observing, measuring, and interpreting various parameters related to biomechanics, such as joint movements, muscle activity, and ground reaction forces. This method offers insights into individuals' functional abilities, movement patterns, and potential abnormalities, making it essential in fields like sports medicine, orthopedics, and rehabilitation.

The process typically starts with data collection. Video analysis allows visual observation of gait characteristics, while motion capture systems precisely track joint movements. Force plates measure ground reaction forces, and wearable sensors provide portable data collection in real-world environments.

Collected data is then analyzed to identify key parameters indicative of normal or abnormal gait. This analysis involves quantitative measurements, statistical comparisons, and biomechanical modeling to assess factors like gait symmetry and coordination. Clinicians and researchers use this information to detect deviations, diagnose gait disorders, and design personalized interventions.

Gait analysis is crucial in clinical practice for diagnosing and treating musculoskeletal and neurological conditions, aiding in personalized treatment plans and rehabilitation monitoring. It also contributes to biomechanical research, advancing understanding of human movement and informing device development.

### **Idiopathic toe walking**

Idiopathic Toe Walking (ITW) is a pediatric gait abnormality characterized by a persistent toe-to-toe walking pattern in children who lack an underlying medical condition. It typically manifests as a habitual gait where the child consistently walks on their toes instead of utilizing a normal heel-to-toe gait pattern. ITW is diagnosed when toe walking persists beyond the age when it is developmentally appropriate, typically beyond the age of three. The term "idiopathic" denotes that the cause of the condition is unknown, as it occurs in otherwise healthy children without any identifiable neurological, orthopedic, or psychological disorders. However, it's important to note that while most cases of toe walking are idiopathic, some may be associated with neurological or musculoskeletal conditions.

When children exhibit a consistent toe-to-toe walking pattern without any underlying medical problem, it is known as idiopathic toe walking (ITW), a pediatric gait disorder. The youngster usually exhibits a habitual gait in which they always walk on their toes rather than following a standard heel-to-toe gait pattern. When toe walking continues past what is considered developmentally acceptable, usually past the age of three, it is labeled as ITW. Since the ailment affects generally healthy youngsters who do not have any discernible

neurological, orthopedic, or psychiatric abnormalities, the word "idiopathic" indicates that the etiology of the condition is unclear. Though most toe walking instances are idiopathic, it is crucial to remember that some may be linked to musculoskeletal or neurological disorders.

### **Controls**

Controls refer to the patients who don't have idiopathic toe walking disability.

### **Accelerometer**

An Accelerometer [3] is a sensor device commonly used in gait analysis to measure the acceleration forces experienced by the body during movement. It detects changes in acceleration along multiple axes, allowing for the assessment of movement patterns, intensity, and duration. In the context of gait analysis, accelerometers are often worn on various parts of the body, such as the lower limbs or trunk, to capture movement dynamics during walking. By recording acceleration data, accelerometers provide quantitative information about gait parameters such as step count, stride length, cadence, and symmetry. This data is valuable for assessing gait abnormalities, monitoring rehabilitation progress, and informing treatment strategies in various clinical settings, including orthopedics, neurology, and sports medicine.

### **Foot plate**

A foot plate [4], also known as a pressure-sensitive foot mat or force plate, is a platform equipped with sensors that measure pressure distribution and foot movements during walking or standing. It consists of a grid of pressure sensors embedded in a flat surface that records the distribution of forces exerted by the feet. When an individual walks or stands on the foot plate, the sensors detect changes in pressure distribution, providing spatial and temporal information about foot contact patterns, weight-bearing distribution, and gait dynamics. Foot plates are widely used in gait analysis to assess foot function, diagnose gait abnormalities, and evaluate orthotic interventions. They are particularly useful for quantifying parameters such as foot pressure, stance phase duration, and gait asymmetry, aiding in the comprehensive evaluation of gait biomechanics and facilitating treatment planning in clinical practice.

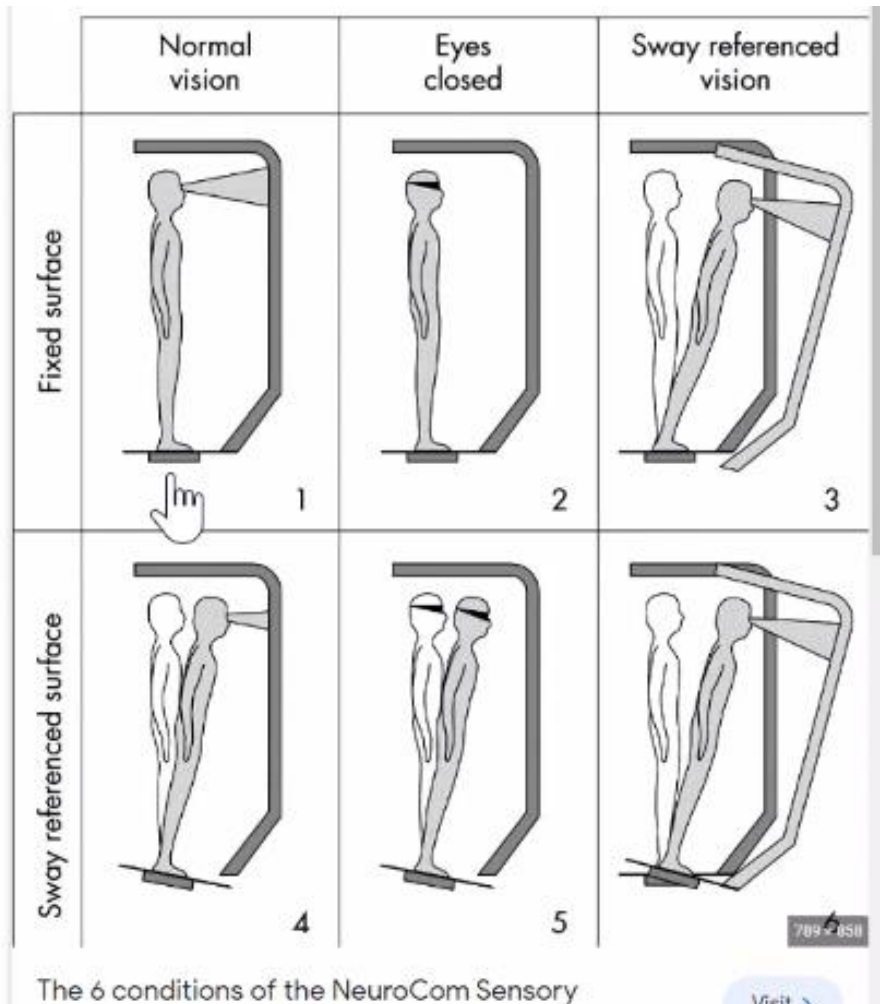
### **Center of Gravity**

The center of gravity (COG) [5], also referred to as the center of mass, is a theoretical point within an object or body where its mass is concentrated. In the context of human movement, the COG represents the point around which the body's weight is evenly distributed in all directions. During walking, the COG undergoes continuous shifts as the body moves through different phases of the gait cycle. Understanding COG dynamics is essential for analyzing balance, posture, and gait stability. In gait analysis, assessing COG displacement and trajectory provides valuable insights into movement patterns, compensatory strategies, and potential biomechanical abnormalities. By quantifying COG parameters such as sway, excursion, and velocity, clinicians can evaluate gait stability, identify deviations from normal patterns, and tailor interventions to improve balance and mobility in individuals with gait disorders or neurological impairments.

### **Gyroscope**

A Gyroscope [6] is a mechanical or electronic device used to measure or maintain orientation and angular velocity. It operates based on the principles of angular momentum and conservation of angular momentum. The primary function of a gyroscope is to detect changes in orientation or rotational motion relative to a reference frame, providing crucial information for navigation, stabilization, and motion tracking in various applications.

The fundamental component of a gyroscope is a spinning rotor, which spins at a high speed, and its angular momentum enables it to maintain its orientation in space, resisting any external forces that attempt to alter its position. As a result, the gyroscope remains stable and fixed in its orientation. Gyroscopes are widely used in a diverse range of applications across various industries and fields.



## Units

Accelerometers typically measure acceleration in units of meters per second squared ( $\text{m/s}^2$ ) or in units of gravity ( $g$ ), where 1  $g$  is equal to the acceleration due to gravity on Earth's surface, approximately  $9.81 \text{ m/s}^2$ .

Gyroscopes measure angular velocity, which is the rate of change of angular displacement, typically in units of radians per second ( $\text{rad/s}$ ) or degrees per second ( $^\circ/\text{s}$ ).

Center of gravity, also known as center of mass, is usually measured in units of distance, such as meters ( $\text{m}$ ) or centimeters ( $\text{cm}$ ), relative to a reference point. It represents the point where the entire mass of an object can be concentrated.

### 3. Methodology and process

The datasets of the patients and the control were entered and a 2D plot was made using Matplotlib command in Google Collab with the independent variables being the Centre of Force along the X axis and the COF along the y axis.

#### Code

##### ✖ Importing the libraries

```
[1] import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

##### ✖ Importing the dataset

```
[2] dataset = pd.read_csv('Final.csv')
X = dataset.iloc[:, :-1].values
y = dataset.iloc[:, -1].values
```

##### ✖ Splitting the dataset into the Training set and Test set

```
[3] from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.25, random_state = 0)
```

```
print(X_train)
```

```
[5] print(y_train)
```

```
[1 1 1 0 0 1 1 1 1 1 1 0 1 0 0 1 1 0 1 1 1 0 1 1 1 1 0 1 1 1 1 0 1 1 1 1 1
1 1 1 1 1 1 1 0 1 0 0 1 1 1 0 1 0 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1
1 0 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 0 0 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 0 0 1 1 1 0 1 1 1 1 1 1 1 1 1 1 0 1
1 0 1 1 1 1 1 1 0 1 1 0 0 1 1 1 1 1 1 0 1 1 1 0 0 1 0 1 1 1 0 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 0 0 0 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 0
1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1
0 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 0 1 0 0 1 1 1 1 1 1 0 0 1 1 1 1 1 1
0 1 1 1 1 1 1 0 1 0 1 1 0 1 1 1 1 1 1 1 0 1 0 1 1 0 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 0
1 1 1 0 1 1]
```

```
print(X_test)
```

```
[[-3.71 -1.51]
 [ 0.16 -1.91]
 [ 0.64 -1.04]
 [-1.41 -2.59]
 [-3.03 -5.3 ]
 [-0.54 -2.48]
 [-4.14 -0.43]
 ...]
```



```
[7] print(y_test)
```

```
[1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 0 1 1 1 1
 1 1 1 0 1 1 1 0 1 0 1 1 0 1 0 1 1 1 1 1 1 1 1 1 0 1 1
 1 0 0 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 0 1 1 1 1 0 0 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 0 1
 1 1]
```

## ✓ Feature Scaling

```
[8] from sklearn.preprocessing import StandardScaler
     sc = StandardScaler()
     X_train = sc.fit_transform(X_train)
     X_test = sc.transform(X_test)
```

```
[9] print(X_train)
```

```
[[ 1.19767711e-02 -3.98834592e-02]
 [ 1.20828355e+00 -2.18170446e-01]
 [ 8.87623998e-01  2.75090217e-01]
 [-9.19169224e-01 -8.06517502e-01]
```

```
[10] print(X_test)
```

```
[[ -1.91814705 -0.03394056]
 [  0.46829997 -0.27165654]
 [  0.7642934  -0.24537572]
 [-0.4998452  -0.67577371]
 [-1.49882302 -2.28629949]
 [  0.03664289 -0.61040182]
 [-2.18330783  0.60789259]
 [  0.61013016 -0.73520271]
 [  0.3696355  -0.24194204]
 [  0.49296609  0.34640501]
 [-1.12883124  0.92880917]
 [-0.64167538  0.4830917 ]
 [  1.16511784 -0.7946317 ]
 [-0.43201337 -0.65200211]
 [-0.29018319 -0.20034175]
 [-0.40734725 -0.8302891 ]
 [  0.96778889 -0.91348969]
 [-0.17918565 -0.70548821]
 [-0.32718237  1.8796731 ]
 [  0.17230655 -0.16468435]
 [  0.38196856 -0.17657015]
 [-0.25935054 -0.93131839]
 [  0.28330408 -0.14091275]
 [  0.71496116  3.25248289]
 [  0.02212522  1.02221620]
```

### ✓ Training the Random Forest Classification model on the Training set

```
0s ✓ from sklearn.ensemble import RandomForestClassifier
classifier = RandomForestClassifier(n_estimators = 10, criterion = 'entropy', random_state = 0)
classifier.fit(X_train, y_train)
```

RandomForestClassifier

### ✓ Predicting a new result

```
0s ✓ print(classifier.predict(sc.transform([[30,87000]])))
```

[1]

### ✓ Predicting the Test set results

```
0s ✓ [13] y_pred = classifier.predict(X_test)
print(np.concatenate((y_pred.reshape(len(y_pred),1), y_test.reshape(len(y_test),1)),1))
```

```
14] from sklearn.metrics import confusion_matrix, accuracy_score
cm = confusion_matrix(y_test, y_pred)
print(cm)
accuracy_score(y_test, y_pred)
```

```
[[ 4 24]
 [17 105]]
0.7266666666666667
```

### • Visualising the Training set results

```
15] from matplotlib.colors import ListedColormap
X_set, y_set = sc.inverse_transform(X_train, y_train)
X1, X2 = np.meshgrid(np.arange(start = X_set[:, 0].min() - 10, stop = X_set[:, 0].max() + 10, step = 0.25),
np.arange(start = X_set[:, 1].min() - 1000, stop = X_set[:, 1].max() + 1000, step = 0.25))
plt.contourf(X1, X2, classifier.predict(sc.transform(np.array([X1.ravel(), X2.ravel()])).T).reshape(X1.shape),
alpha = 0.75, cmap = ListedColormap(('salmon', 'dodgerblue')))
plt.xlim(X1.min(), X1.max())
plt.ylim(X2.min(), X2.max())
for i, j in enumerate(np.unique(y_set)):
plt.scatter(X_set[y_set == j, 0], X_set[y_set == j, 1], c = ListedColormap(('salmon', 'dodgerblue'))(i), label = j)
plt.title('Random Forest Classification (Training set)')
```

## 4. Results and Discussions

Based on the research done, we can conclude that Random Forest and Decision Tree deliver a significant accuracy of 80%. Hence the above Machine learning techniques can be used in Cohort study for a significant accuracy in classification.

## 5. Conclusion

The research discussed underscores the importance of utilizing wearable sensor technology and data-driven insights to uncover the subtle patterns and deviations associated with idiopathic toe walking (ITW). By harnessing the power of machine learning algorithms, efforts have been made to enhance the accuracy of ITW diagnosis

and personalize treatment strategies, offering promising avenues for improved patient care and outcomes. Additionally, the investigation highlights the potential of machine learning-based gait analysis as a valuable tool for clinicians and researchers in pediatric healthcare.

Through the identification of distinctive gait features indicative of ITW and the evaluation of machine learning classifiers, significant strides have been made in advancing our understanding of this enigmatic condition. While the findings represent significant progress, there are still challenges and opportunities that warrant further exploration. Future research endeavors should focus on refining machine learning algorithms, optimizing sensor placements, and conducting longitudinal studies to validate the efficacy of proposed approaches in real-world clinical settings.

The collective efforts aim to catalyze transformative advancements in the assessment and management of ITW, with the overarching goal of improving the quality of life for affected children. By fostering collaboration between clinicians, researchers, and technologists, the potential of gait analysis and machine learning can be harnessed to usher in a new era of precision medicine in pediatric healthcare.

## Acknowledgements

We would like to express our sincere appreciation to Vellore Institute of Technology, Vellore for their invaluable support and resources provided throughout this research endeavor. Special thanks to Proff. Sivakumar R for his guidance, expertise, and encouragement during this study. Additionally, we extend our gratitude to the participants who generously contributed their time and data, without whom this research would not have been possible.

## References

1. Gita Pendharkar, Paul Percival, David Morgan. Evaluating Bouncy Gait in Idiopathic Toe-Walkers Using Accelerometer. <https://ieeexplore-ieee-org.egateway.vit.ac.in/stamp/stamp.jsp?tp=&arnumber=4762009>
2. Sk Md Alfayeed. Baljit Singh Saini (2021). Human Gait Analysis Using Machine Learning: A Review. <https://ieeexplore-ieee-org.egateway.vit.ac.in/stamp/stamp.jsp?tp=&arnumber=9410678>
3. Gita Pendharkar, Paul Percival, David Morgan. Evaluating Bouncy Gait in Idiopathic Toe-Walkers Using Accelerometer. <https://ieeexplore-ieee-org.egateway.vit.ac.in/stamp/stamp.jsp?tp=&arnumber=4762009>
4. Jindapa Nampeng, S. Visitsattapongse and C. Pintavirooj (2017). Dynamic Foot Plantar Measurement Platform using Optical Sensors. <https://ieeexplore-ieee-org.egateway.vit.ac.in/stamp/stamp.jsp?tp=&arnumber=8229159>
5. Yun Ling, Ying Xue, Jianguo Xing, Tao Jiang, Chaona Guo. Experimental Studies on Static Postural Balance Using the Body Center of Gravity Test System. <https://ieeexplore-ieee-org.egateway.vit.ac.in/stamp/stamp.jsp?tp=&arnumber=6708055>
6. S.Praveen Kumar, P.Sashti Gobinath, S.Saravana Prasanth, P.Prasanth(2014). Design of vibratory MEMS Gyroscope for Gait Phase Detection System. <https://ieeexplore-ieee-org.egateway.vit.ac.in/stamp/stamp.jsp?tp=&arnumber=7006188>
7. Shipra Shand.Rahul(2022). Importance of Human Gait Analysis in Biometric Recognition using Machine Learning: A Review. <https://ieeexplore-ieee-org.egateway.vit.ac.in/stamp/stamp.jsp?tp=&arnumber=9776760>

# Characterization of Respiratory Diseases Using LabVIEW

Yuvapriya E., Sivakumar R.

Vellore Institute of Technology, Katpadi, Vellore – 632014, India  
*yuvapriya.e2023@vitstudent.ac.in, rsivakumar@vit.ac.in*

**Abstract.** Spirometry measures the volume and speed of air that is inhaled and expelled to assess lung function. It helps in the diagnosis of diseases like chronic obstructive pulmonary disease (COPD) and asthma. The symptoms of asthma, which include mucus production and constricted airways, can vary greatly and have a serious influence on day-to-day functioning, perhaps even resulting in fatal attacks. Genetic damage causes lung cancer, which manifests as malignant tumors that impair airway cells. Tumors caused by unchecked cell multiplication limit lung function; successful treatment requires early intervention. Frequent spirometry evaluations track the efficiency of treatment as well as the health of the lungs in addition to diagnosing chronic lung disorders. Important metrics, such as forced expiratory volume (FEV) and forced vital capacity (FVC), offer important information regarding the severity of respiratory problems and the ability to breathe. Beyond diagnosis, spirometry plays a vital role in continuous monitoring that improves respiratory outcomes and general health. For biomedical applications, LabVIEW, a graphical programming environment improves the productivity of designing algorithms, signal processing, and data visualization. This work includes classifying disorders using respiratory rates into apnoea, tachypnoea, and bradypnea. By calculating the FEV1/FEV6 ratio classified asthma and lung cancer. Determining Total Lung Capacity makes it easier to diagnose restrictive or obstructive lung conditions. With the help of the Forced Expiratory Volume in 1 Second (FEV1) can assess the severity of the disease, help with diagnosis refinement. By precisely characterizing and managing lung illnesses, this all-inclusive approach to respiratory assessment directs healthcare activities for the best possible patient care.

**Keywords:** Spirometry, Asthma, Lung cancer, Respiration rate, FEV1, FEV6, LabVIEW.

## 1. Introduction

Spirometry is a common office test used to assess how well your lungs work by measuring how much air you inhale, how much you exhale and how quickly you exhale. Spirometry is used to diagnose asthma, chronic obstructive pulmonary disease (COPD) and other conditions that affect breathing. Asthma is a condition in which your airways narrow and swell and may produce extra mucus. This can make breathing difficult and trigger coughing, a whistling sound (wheezing) when you breathe out and shortness of breath.

For some people, asthma is a minor nuisance. For others, it can be a major problem that interferes with daily activities and may lead to a life-threatening asthma attack. Spirometry may also be used periodically to monitor your lung condition and check whether a treatment for a chronic lung condition is helping you breathe better. Key spirometry measurements include the following:

**Forced vital capacity (FVC):** This is the largest amount of air that you can forcefully exhale after breathing in as deeply as you can. A lower than normal FVC reading indicates restricted breathing. **Forced expiratory volume (FEV):** This is how much air you can force

from your lungs in one second. This reading helps your doctor assess the severity of your breathing problems. Lower FEV-1 readings indicate more significant obstruction.

## **2. Treatments for respiratory conditions**

Treatment for respiratory diseases encompasses a range of approaches tailored to the specific condition and its severity. Medications form a cornerstone, with bronchodilators relaxing airway muscles in asthma and COPD, while corticosteroids reduce inflammation, particularly during exacerbations. Antibiotics combat bacterial infections if present, though viral infections necessitate antiviral medications, especially for conditions like influenza or COVID-19. Oxygen therapy may be employed to bolster low blood oxygen levels, while pulmonary rehabilitation programs offer education and exercise to enhance breathing and quality of life, particularly for chronic conditions like COPD.

Lifestyle modifications such as smoking cessation, allergen avoidance, and weight management play crucial roles in managing respiratory diseases. Vaccinations, including flu and pneumococcal vaccines, help prevent certain infections. Surgery may be necessary for advanced cases, such as lung cancer or severe emphysema. Symptomatic relief, through over-the-counter medications like cough suppressants or pain relievers, aids in managing symptoms. Collaboration between individuals and healthcare providers ensures tailored treatment plans and regular monitoring for optimal management and prevention of complications. Early detection and intervention remain pivotal in enhancing outcomes for those with respiratory ailments.

## **3. Merits of characterization using LabVIEW**

Characterizing respiratory diseases using LabVIEW confers several advantages owing to its versatile system design and development platform. Researchers benefit from the ability to customize experimental setups to match specific research requirements, enhancing flexibility and adaptability. LabVIEW's robust tools for data acquisition and analysis empower researchers to collect, process, and visualize respiratory parameters in real-time, facilitating comprehensive disease characterization. Seamless integration with a wide array of sensors and instruments streamlines experimental setups and data acquisition processes, bolstering efficiency and accuracy.

Automation capabilities minimize manual intervention, reduce errors, and increase reproducibility, particularly in long-term studies. Real-time monitoring and control features enable dynamic adjustments to experimental conditions, fostering precise experimentation. Furthermore, LabVIEW's support for remote access and collaboration promotes teamwork and facilitates data sharing and dissemination. Its scalability accommodates evolving research needs, while a user-friendly graphical interface simplifies the development of complex experimental setups and data analysis routines. Collectively, these attributes accelerate respiratory research and deepen our understanding of respiratory diseases.

## **4. Characterization of respiratory diseases using LabVIEW**

Spirometry assesses the integrated mechanical function of the lung, chest wall, respiratory muscles, and airways by measuring the total volume of air exhaled from a full lung (total lung capacity [TLC]) to maximal expiration (residual volume [RV]). This volume, the forced vital capacity (FVC) and the forced expiratory volume in the first second of the forceful exhalation (FEV<sub>1</sub>) should be repeatable to within 0.15L upon repeat efforts in the same measurement unless the largest value for either parameter is less than 1L. In this case, the

expected repeatability is to within 0.1 L of the largest value. The patient is instructed to inhale as much as possible and then exhale rapidly and forcefully for as long as flow can be maintained. The patient should exhale until one of the criteria defining the end of a forced exhalation has been reached. At the end of the forced exhalation, the patient should again inhale fully as rapidly as possible. The FVC should then be compared with that inhaled volume to verify that the forced expiratory maneuver did start from full inflation (Table 1).

Table 1. Calculation for FEV1/FEV6 ratio for men and women

Sex	Formula to calculate spirometric parameters
Men	$FEV1 \text{ {liters}} = (4.30 \times \text{height{meters}}) - (0.029 \times \text{age {years}}) - 2.49$
	$FEV6 \text{ {liters}} = (5.76 \times \text{height{meters}}) - (0.026 \times \text{age {years}}) - 4.34$
Women	$FEV1 \text{ {liters}} = (3.95 \times \text{height{meters}}) - (0.025 \times \text{age {years}}) - 2.60$
	$FEV6 \text{ {liters}} = (4.43 \times \text{height{meters}}) - (0.026 \times \text{age {years}}) - 2.89$

#### To calculate FEV1/FEV6 ratio:

As the FEV1 and FEV6 values are proportional to the size of a lung, therefore men and women of different age or height will have different FEV1 and FEV6 values. In the user interface, the selection of gender, age, and height need to be entered by the user to calculate the estimated value.

In the subVI, the spirometric measured FEV1/FEV6 ratio present in the array. Then by calling subVI, we can use the data in main VI.

Then these values will be compared with the standard value to characterize whether the patient has a chance of having asthma or lung cancer.

#### To calculate respiration rate:

##### Respiration rate= 60/Total time for respiratory cycle

From the Respiration rate we can classify apnoea (absence of breath), tachypnoea (high respiration rate), bradypnea (low respiration rate).

#### To calculate total lung capacity:

$$TLC = \text{Inspiratory reserve volume} + \text{Expiratory reserve volume} + \text{Tidal volume} + \text{Residual volume}$$

By calculating this we can detect whether the person is having obstructive lung disease or restrictive lung disease.

By having the FEV1 value we can tell the severity of the lung disease ranging from mildly abnormal to very severe. The graph of each parameter is displayed separately.

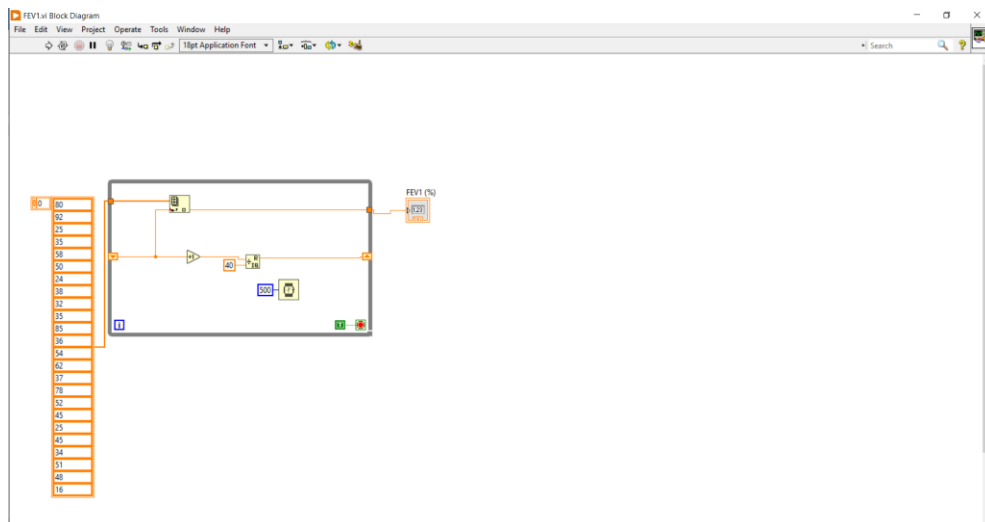


Fig. 1. Block diagram of subVI of FEV1 data

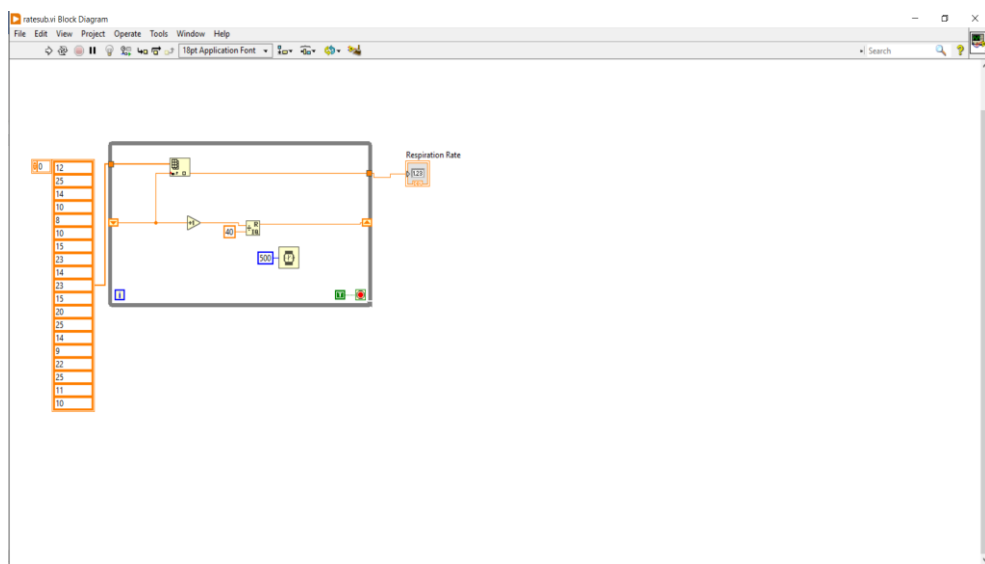


Fig. 2. Block diagram of SubVI of respiration rate data

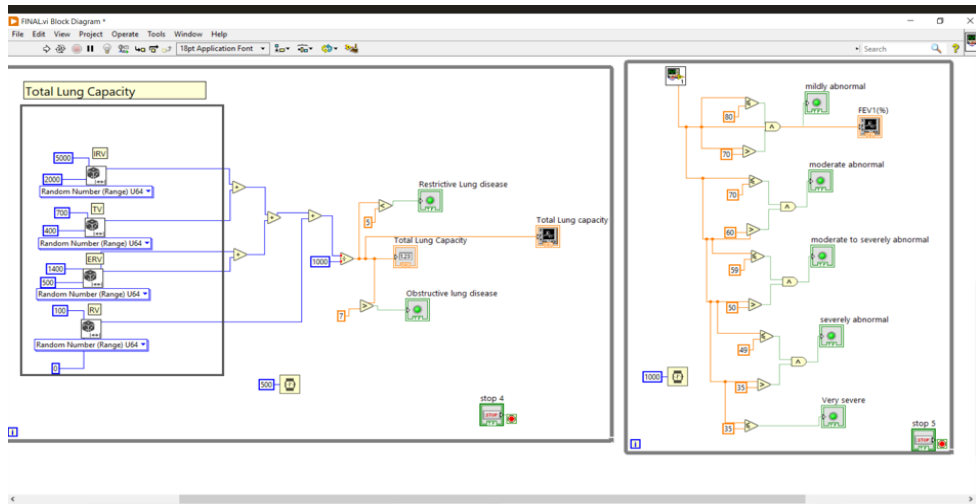


Fig. 3. Block diagram of Total Lung capacity and FEV1(%)

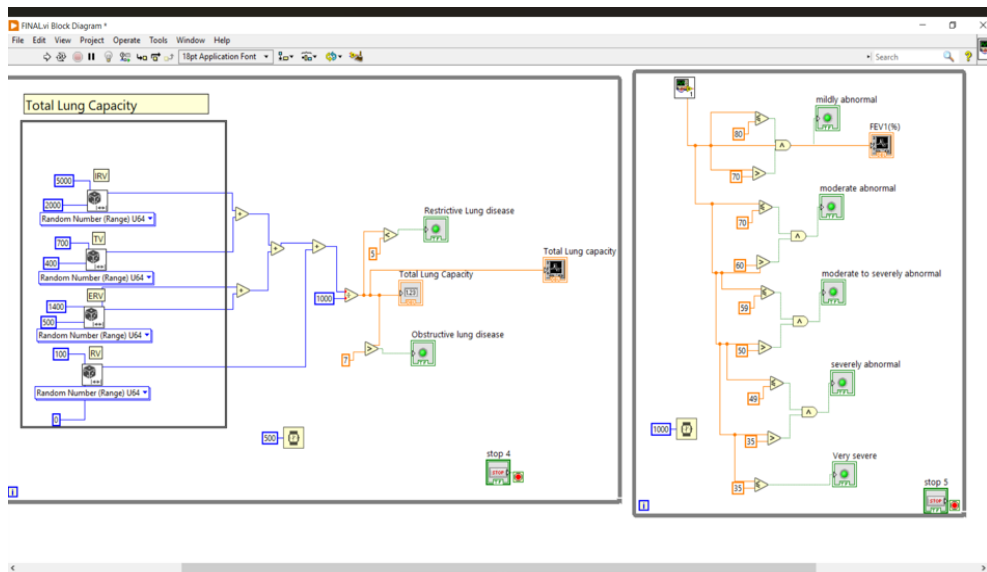


Fig. 4. Block diagram of Total Lung capacity and FEV1%



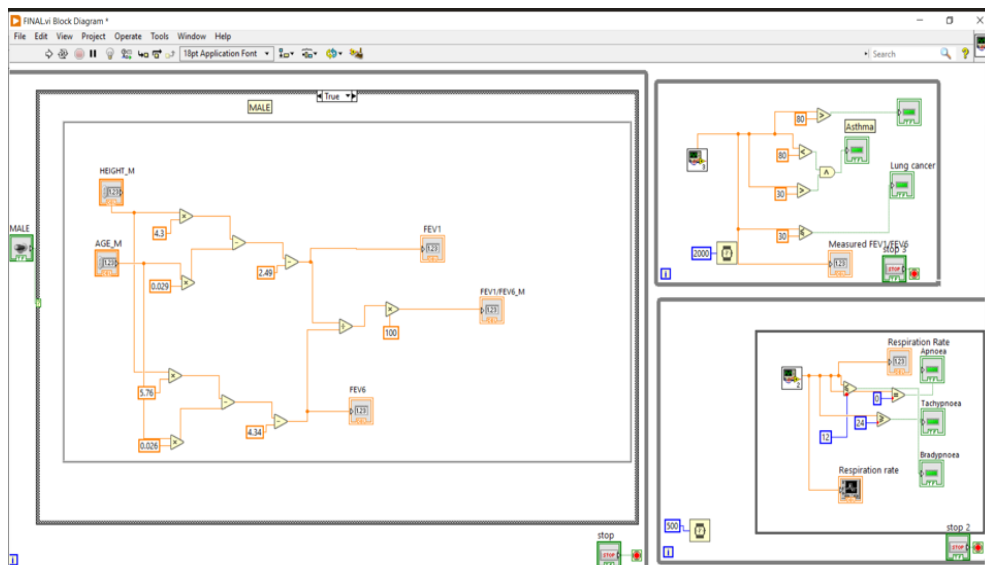


Fig. 5. Block diagram of calculation of FEV1/FEV6 ratio for male and classification of disease based on Respiration rate

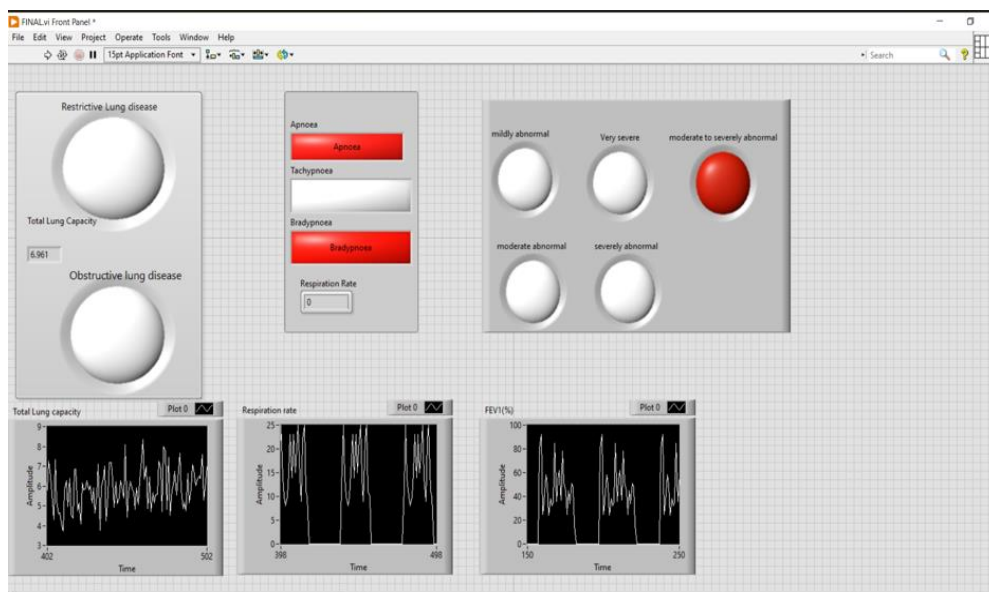


Fig. 6. Front panel of classifying various conditions

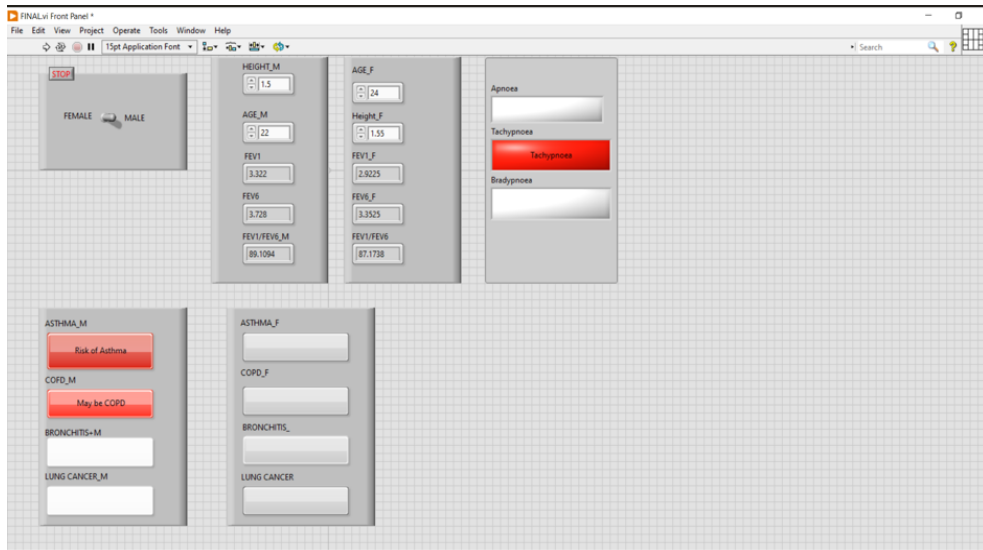


Fig. 7. Front panel of calculating FEV1/FEV6 and classifying various conditions for male and female.

## 5. Conclusion

Spirometry serves as a crucial tool for assessing the integrated mechanical function of the lung, chest wall, respiratory muscles, and airways. By measuring parameters such as forced vital capacity (FVC), forced expiratory volume in the first second (FEV1), and FEV1/FEV6 ratio, clinicians can gain valuable insights into respiratory health and diagnose conditions such as asthma, lung cancer, and obstructive or restrictive lung diseases. Additionally, respiration rate calculation allows for the classification of abnormal breathing patterns, such as apnea, tachypnea, and bradypnea, further aiding in diagnosis. The determination of total lung capacity (TLC) provides additional information for detecting lung diseases and assessing their severity. By displaying the graph of each parameter separately, clinicians can visualize and interpret the results effectively. Overall, the comprehensive analysis provided by spirometry facilitates accurate diagnosis and management of respiratory conditions, ultimately improving patient outcomes and quality of life.

## References

1. S. Noorjannah Ibrahim, A. Z. Jusoh, N. Abdul Malik, A. L. Asnawi, and S. Mazalan, 2018, Characterization of Respiratory Conditions Using Labview and Digital Spirometer. Indonesian Journal of Electrical Engineering and Computer Science, vol. 10, no. 1, pp. 66–66, Apr. 2018.
2. Chu, M., Nguyen, T., Pandey, V., et al., 2019. Respiration rate and volume measurements using wearable strain sensors. NPJ digital medicine, 2(1), p.8.
3. Prema, S., Kumar, I., Ishwarya, M., Jamuna, K. and Gowtham, R.K., 2020, March. Monitoring of Asthma and Regulating Breathing Practice. 6th Int. Conf. on Advanced Computing and Communication Systems (ICACCS), IEEE, pp. 390-393.
4. Prinable, J., Jones, P., Boland, D., McEwan, A. and Thamrin, C., 2020. Derivation of respiratory metrics in health and asthma. Sensors, 20(24), p.7134.
5. Li, Y., Liu, C., Zou, H., et al., 2023. Integrated wearable smart sensor system for real-time multi-parameter respiration health monitoring. Cell Reports Physical Science, vol.4, no. 1) pp.101-191.

# Wearable Technology in Healthcare Monitoring

**Challa Revanth Kumar, Sivakumar R.**

School of Electronics Engineering, Vellore Institute of Technology Tamil Nadu, India  
*challarevanth2001@gmail.com, rsivakumar@vit.ac.in*

**Abstract.** The paper explores the impact of wearable technologies on healthcare and offering solutions for monitoring the patient health. To evaluate the effectiveness of wearable technologies in monitoring and managing health parameters in diverse patient population. This paper explores the integration of LabVIEW with a versatile design platform with wearable technology in healthcare applications. Wearable technology has a high level of accuracy in continuous monitoring of vital signs like heart rate, blood pressure, temperature. According to the study findings, wearable technology is very useful for ongoing healthcare monitoring because it can provide real-time vital sign data with an accuracy level that is on the same level with conventional clinical measurements.

**Keywords:** Wearable technologies, LabVIEW, Healthcare, Blood pressure, Temperature, Heart rate.

## 1. Introduction

To define wearable technology in the context of healthcare monitoring it refers to the electronic device that individuals can wear to collect the track and analyzes the health-related data. This type of wearable sensors is designed in such a way to measure the physiological parameters like heart rate, temperature, blood pressure, sleep patterns, pulse rate, and calorie expenditure. Some of the medical devices also track the specific health issues like diabetes or monitor medication adherence.

The purpose of this report is to review the most common concepts of wearable technology in order to outline basic qualitative aspects of the most Healthcare monitoring views. The ultimate goal is to enhance the early detection of medical issues and reduce healthcare costs.

The most significant consideration is that it includes prioritizing the accuracy in data collection, ensuring robust security, privacy measures and it is user friendly.

## 2. Definition of Wearable Technology in Healthcare

Think that a health tracker is not just counting our steps, but it is continuously monitoring our stress levels, heart rhythm, and blood pressure. So, this is the reality in wearable technology in healthcare. Generally, these devices worn on our body like smart watches or patches, and it monitors and manage our health.

The benefits of this technology are numerous. Early detection of health issues [1] becomes possible through subtle changes in vital signs, potentially preventing complications and improving outcomes. Wearables also personalize care plans [2], allowing healthcare providers to tailor treatment to individual needs and preferences. This increased engagement with personal health data [3] empowers individuals to take control of their well-being, leading to better adherence to treatment plans and improved health outcomes.

However, challenges remain. Accuracy and reliability [4] of wearable data vary depending on the device and individual factors. Data privacy and security [5] are crucial concerns, as these devices collect sensitive health information. Additionally, affordability and accessibility can create a digital divide, excluding some individuals from the benefits of wearables.

Despite these challenges, the future of wearable technology in healthcare is incredibly promising. Advancements in sensor technology, data analytics, and artificial intelligence are paving the way for exciting possibilities. Imagine wearables seamlessly integrated with electronic health records, providing a holistic view of an individual's health for better diagnosis and treatment. Non-invasive monitoring of biomarkers like blood sugar and stress hormones could revolutionize personalized medicine, while AI-powered insights could predict health risks and recommend personalized interventions in real-time.

Wearable technology is not just a trend; it's a transformative force in healthcare. As research and development continue, these devices will become increasingly sophisticated and accessible, playing a crucial role in promoting health and well-being for all.

### **3. Current Wearable Technology in Healthcare Trends**

The world of wearable technology in healthcare is booming with innovation, promising a future where we monitor and manage our health in ways we never imagined. These tiny devices strapped to our wrists, embedded in our clothes, or nestled in our ears are more than just fitness trackers; they're revolutionizing the way we interact with healthcare.

- Imagine an AI assistant analyzing your sleep patterns and activity levels to predict health risks, or a smartwatch automatically adjusting your medication dosage based on real-time glucose levels. This is the power of AI and machine learning in wearable technology. These powerful tools are being used to sift through the mountains of data generated by wearables, extracting meaningful insights to personalize your health journey.[6]

While smartwatches remain popular, the future of wearables is expanding beyond our wrists. Think smart clothing embedded with sensors to track muscle movement and posture, hearables that monitor brain activity for early detection of neurological conditions, or even smart glasses for remote patient monitoring. These diverse forms of wearables offer a more comprehensive picture of our health, catering to specific needs and preferences.[7]

- Wearables are no longer solely focused on physical health. Biomarkers like heart rate variability and skin conductance are being used to track stress, anxiety, and even depression. This data can be used to develop personalized interventions, such as mindfulness exercises or app-based therapy programs, helping individuals manage their mental well-being proactively.[8]

- Wearables are bridging the gap between patients and healthcare providers, enabling real-time data sharing and remote monitoring. Imagine a doctor remotely monitoring your vitals through your wearable device and making adjustments to your treatment plan based on your live data. This seamless integration of wearables with telemedicine platforms promises improved accessibility and convenience for both patients and healthcare professionals.

- Imagine temporary tattoos that monitor your health! This futuristic trend involves ultra-thin, skin-like sensors that seamlessly integrate with the body, offering continuous and non-invasive health monitoring. These "smart tattoos" could track everything from hydration levels to blood sugar, paving the way for a future where health monitoring becomes effortless and ubiquitous.

- With the vast amount of data generated by wearables, personalized medicine is becoming a reality. Healthcare providers can tailor treatment plans to your unique needs and genetic makeup, leading to more effective and targeted interventions. This shift towards

personalized medicine promises a future where healthcare is truly customized to each individual.

- As wearable technology collects sensitive health data, ethical concerns regarding privacy and security are paramount. Data protection and responsible use are crucial for building trust and ensuring the ethical implementation of this technology. Open communication, strong security measures, and user control over data are essential to ensure that wearable technology empowers individuals without compromising their privacy.

- These are just a glimpse into the exciting trends shaping the future of wearable technology in healthcare. As advancements continue, we can expect even more sophisticated and integrated devices that empower individuals to take control of their health and revolutionize the way we deliver healthcare. The future is on our wrists, in our ears, and even on our skin – and it's a future filled with possibilities for a healthier, happier world.

#### **4. A Model of Wearable Technology in LabVIEW**

The healthcare landscape is evolving, and wearables paired with LabVIEW are leading the charge. These compact devices continuously gather data, offering real-time insights into a patient's health, while LabVIEW's powerful tools analyze and visualize this information, paving the way for proactive care and improved outcomes.

Imagine sensors in your smartwatch tracking your heart rhythm, skin conductance, and movement. LabVIEW seamlessly interfaces with these sensors, processing the raw data, filtering out noise, and extracting key features like heart rate variability or Temperature.

The benefits extend beyond individual patients. Wearables and LabVIEW can be used to remotely monitor chronically ill patients, allowing for early detection of complications and proactive care management. Imagine athletes optimizing their training based on real-time biofeedback or researchers gaining valuable insights into population health through large-scale data collection.

However, the power of wearables and LabVIEW goes beyond mere visualization. This duo can trigger personalized interventions, sending you medication reminders, prompting you to take a break when stress levels rise, or even contacting your healthcare provider if your data shows signs of concern.

The integration of wearable technology and LabVIEW is still in its early stages, but the potential is limitless. From helping us understand our bodies better to revolutionizing disease prevention and treatment, this powerful combination is poised to transform healthcare as we know it. So, the next time you strap on your smartwatch or fitness tracker, remember, you're not just wearing a device; you're carrying a tool for a healthier future.

This analysis isn't just stored on your device; it's visualized in real-time on dynamic charts and graphs, giving you and your healthcare providers a clear picture of your health trends.

Beyond visualization, LabVIEW empowers intervention. The processed data can trigger medication reminders, personalized coaching messages, or even alerts for abnormal readings, ensuring you stay on top of your health. This data can also be shared remotely with healthcare professionals, facilitating proactive care and early detection of potential issues.

This model isn't limited to specific scenarios. Imagine athletes optimizing their training with wearables and LabVIEW, chronically ill patients being monitored remotely, or individuals managing stress and anxiety with personalized data insights. The possibilities are endless, offering a future where healthcare is continuous, personalized, and data-driven.

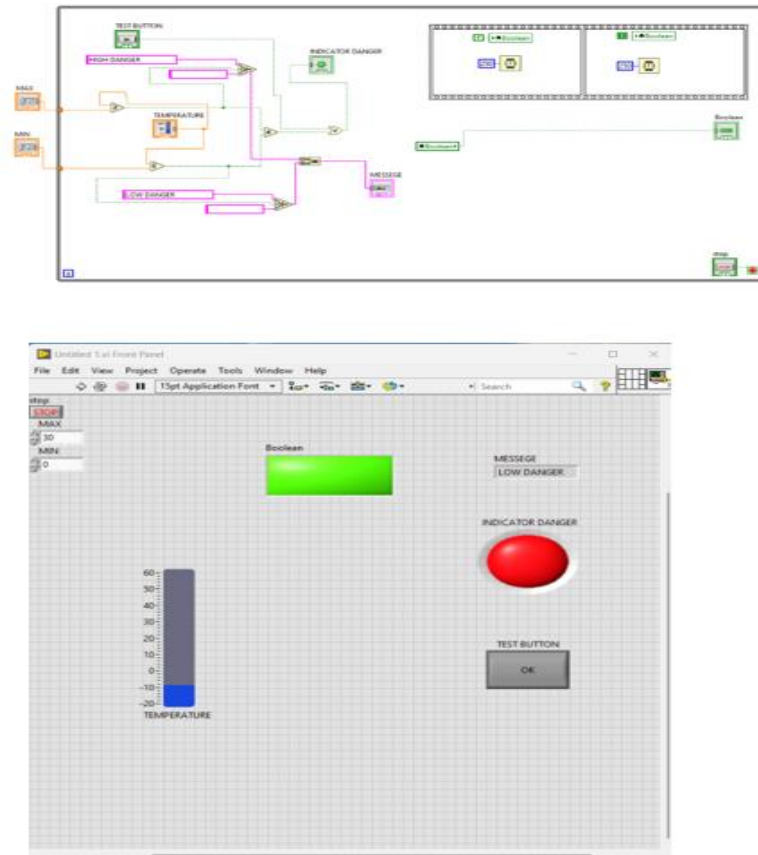


Fig. 1.1. simulation of temperature

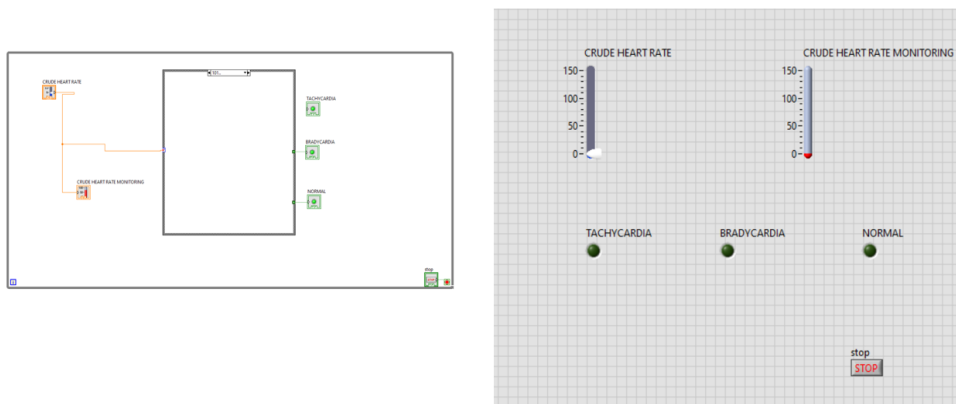


Fig 1.2: Simulation of Heart rate

So, if you're looking for a glimpse into the future of healthcare, look no further than the marriage of wearable technology and LabVIEW. This powerful combination is transforming the way we understand and manage our health, paving the way for a healthier and more proactive future.

## 5. Conclusion

Wearable technologies are revolutionizing healthcare by providing continuous and unobtrusive monitoring of vital signs and physiological data. This leads to enhanced disease management through early detection, improved treatment through real-time monitoring, and remote patient care through telehealth. Wearables also enable proactive prevention through personalized wellness tracking and population health monitoring. Additionally, technological advancements in miniaturization, AI, and interconnectivity are creating a future of personalized, proactive, and data-driven healthcare.

## References

1. Wang, J., Chen, Y., & Sun, M. (2020). Wearable sensors for remote patient monitoring: from sensor selection to data mining for early disease detection. *International Journal of m-Health*, 12(1), 3287.
2. Liu, Y., & Guo, Y. (2020). Wearable technology for personalized healthcare: A review. *Biosensors and Bioelectronics*, 156, 112061.
3. Deng, J., & Chen, F. (2020). Wearable technology for personal healthcare: A review. *Healthcare*, 8(4), 43.
4. Hsu, Y.-L., & Huang, Y.-M. (2018). A review of wearable technology in healthcare. *IEEE Access*, 6, 74319-74340.
5. DeCamp, P., & Huff, C. (2019). Data privacy concerns in wearable healthcare devices. *Journal of Medical Ethics*, 45(4), 269-274.
6. <https://physicsworld.com/a/energy-efficient-transistor-enables-ai-analysis-of-health-data-within-wearable-devices/>
7. <https://spectrum.ieee.org/smart-clothes-breath-monitor>
8. <https://www.wired.co.uk/article/fitbit-stress-tracking-eda>

# LabVIEW-Based EEG Signal Analysis for the Identification and Characterization of Aberrant Brain Activity Patterns in Neurological Disorders

Saran Aravinda C., Sivakumar R.

Vellore Institute of Technology, Katpadi, Vellore – 632014, India  
*Saranaravinda.c2023@vitsstudent.ac.in, rsivakumar@vit.ac.in*

**Abstract.** Electroencephalogram (EEG) signals are a vital source of information for neurological healthcare professionals, particularly in the identification of abnormal brain activity. The use of LabVIEW software to analyze EEG readings and diagnose aberrant brain activity is discussed in this abstract. National Instruments created LabVIEW, a graphical programming environment renowned for its adaptability in signal processing and data collecting. The goal of the study is to identify and characterize aberrant brain activity patterns by developing algorithms in LabVIEW to analyze EEG signals. In order to evaluate EEG signals suggestive of disorders like epilepsy, insomnia, stress level, learning disability, and memory power, the abstract explores the unique approaches used in LabVIEW. Real-time processing capabilities of LabVIEW contribute to the continuous monitoring of abnormal brain activity, allowing for timely interventions and personalized patient care.

**Keywords:** EEG, Alpha, Delta, Beta, Gamma, Insomnia, Depression, Alzheimer's disease, Brain activity.

## 1. Introduction

In the intricate realm of medical diagnostics, the analysis of Electroencephalogram (EEG) signals stands as a cornerstone for deciphering the complexities of brain activity. These signals, which capture the electrical dynamics of the brain, hold invaluable insights into a myriad of neurological conditions spanning from epilepsy to cognitive impairments. Yet, unraveling the intricacies of EEG signals, characterized by their non-stationary and multi-frequency nature, demands sophisticated signal processing techniques for meaningful interpretation. In this pursuit, LabVIEW emerges as a formidable ally, offering modular design and a rich repository of signal processing libraries, empowering clinicians and researchers to extract vital information from EEG signals, thereby facilitating the identification of abnormal brain activity patterns associated with specific disorders [1].

Within the domain of neuro diagnostics, LabVIEW's multifaceted applications shine brightly, particularly in the diagnosis and monitoring of sleep disorders, stress levels, memory power, learning disabilities, and Alzheimer's disease through EEG signal analysis [2]. With its robust signal processing capabilities, LabVIEW enables the identification of characteristic patterns in sleep-related EEG signals, such as sleep spindles and slow-wave activity, aiding in the accurate diagnosis of conditions ranging from insomnia to sleep apnea. Furthermore, its real-time data processing prowess and visualization tools prove invaluable in monitoring and interpreting EEG signals associated with stress, thus paving the way for comprehensive systems for stress assessment and management [3].



Moreover, LabVIEW's adaptability and flexibility render it an ideal platform for the development of diagnostic tools tailored to address various neurological challenges. By leveraging EEG signal analysis, LabVIEW can assist in the early identification of learning disabilities by pinpointing patterns indicative of such conditions, thereby facilitating timely intervention and personalized educational strategies. Additionally, LabVIEW's advanced signal processing algorithms play a pivotal role in detecting anomalies within EEG signals associated with Alzheimer's disease, offering promise for early diagnosis and monitoring. Ultimately, the integration of LabVIEW in EEG signal analysis not only enhances diagnostic precision but also fosters a holistic approach to neuro diagnostics, promising advancements in our comprehension and management of brain-related conditions [4].

## **2. Effect of Alzheimer's, Insomnia, Learning disability in day-to-day life**

Alzheimer's disease, a progressive neurological disorder, wreaks havoc on cognitive abilities, memory, and behavior. As the disease advances, individuals experience a myriad of debilitating effects. Initially, short-term memory loss emerges, disrupting everyday tasks and conversations. Over time, this memory impairment worsens, leading to challenges in recalling familiar faces, places, or events. Language difficulties become apparent as individuals struggle to articulate thoughts or understand conversations. Basic motor functions decline, making routine activities like eating or dressing arduous. Behavioral changes manifest, ranging from agitation and mood swings to withdrawal from social interactions. Alzheimer's eventually erodes independence, rendering individuals reliant on others for daily care. The profound impact extends beyond patients, affecting families who witness the gradual loss of their loved one's essence. Ultimately, Alzheimer's disease presents a heart-wrenching journey marked by the gradual dissolution of memory, cognition, and identity [2].

Insomnia, a common sleep disorder, inflicts a host of detrimental effects on individuals' physical and mental well-being. Primarily characterized by difficulty falling asleep, staying asleep, or both, insomnia disrupts the natural sleep-wake cycle, leaving individuals feeling perpetually fatigued and irritable. The ramifications extend beyond mere tiredness, permeating various aspects of life. Cognitive function deteriorates, impairing memory, concentration, and decision-making abilities, which can hinder performance at work or school. Persistent sleep deprivation undermines the immune system, making individuals more susceptible to illnesses. Chronic insomnia often contributes to the development or exacerbation of mental health disorders such as anxiety and depression, further exacerbating the sleeplessness cycle. Physical health suffers as well, with increased risk of cardiovascular diseases, obesity, and diabetes. Moreover, insomnia can strain interpersonal relationships due to mood disturbances and decreased social engagement. The relentless cycle of sleeplessness perpetuates a profound sense of distress and diminishes overall quality of life, underscoring the urgent need for effective interventions to mitigate its impact [4].

Learning disabilities encompass a diverse range of neurological disorders that affect the acquisition and processing of information, resulting in significant challenges in various aspects of life. These disabilities often manifest as difficulties in reading, writing, math, or understanding spoken language. The effects of learning disabilities are far-reaching, impacting academic achievement, social interactions, and self-esteem. In educational settings, individuals with learning disabilities may struggle to keep pace with their peers, experiencing frustration, and a sense of inadequacy. These challenges can lead to feelings of isolation and alienation, as well as increased risk of behavioral issues. Despite often possessing average or above-average intelligence, individuals with learning disabilities may face persistent barriers

to academic success. Additionally, these difficulties can extend into adulthood, affecting employment opportunities and daily functioning. However, with proper support, accommodations, and intervention, individuals with learning disabilities can learn to effectively navigate their challenges, unlocking their full potential and achieving success in various aspects of life [5].

Weighted utensil use as the mainstay of lifestyle therapies can help patients with their activities of daily living (ADLs) and lessen the amplitude of their tremor. When combined with extra weights, certain tools (such as a spoon) can help patients to consume food and liquids. Guidelines for the treatment of Parkinson's disease (PD) in adults were released in 2017 by the National Institute for Health and Care Excellence (NICE). Physiotherapy and occupational therapy may be beneficial for patients in the early stages of Parkinson's disease (PD) if they exhibit motor symptoms or struggle with ADLs. However, the guidelines developed by the IPMDS and the American Academy of Neurology (AAN) did not address lifestyle or nonpharmacological management of ET [4].

### **3. Simulation of Multiple Diagnostic Tool using LabVIEW**

A diagnostic tool was simulated using the LabVIEW software for the patients with Alzheimer's disease, Insomnia, Learning disabilities, stress Management. The methodology for analyzing EEG signals using LabVIEW software involves several key steps to ensure accurate interpretation and diagnosis of abnormal brain activity patterns. Firstly, data acquisition is performed using EEG sensors placed on the scalp to capture electrical activity from various regions of the brain. These sensors record signals corresponding to different brain waves, including alpha, beta, delta, and gamma waves, which are crucial indicators of brain function and abnormalities. The EEG data is then transmitted to a computer system equipped with LabVIEW software for further analysis [2].

Secondly, signal processing algorithms are developed within the LabVIEW environment to extract relevant features from the EEG signals. These algorithms utilize a combination of digital signal processing techniques, such as filtering, spectral analysis, and time-frequency analysis, to identify characteristic patterns associated with specific neurological conditions. For example, algorithms may focus on detecting abnormalities such as spikes or sharp waves indicative of epilepsy, or patterns of slow-wave activity characteristic of sleep disorders like insomnia. Additionally, machine learning techniques can be incorporated to improve the accuracy of pattern recognition and classification, enabling the differentiation between normal and abnormal brain activity [1].

Furthermore, the developed algorithms are validated using EEG data collected from patients diagnosed with various neurological disorders, including epilepsy, insomnia, stress-related conditions, learning disabilities, and Alzheimer's disease [2].

This validation process involves comparing the results obtained from EEG signal analysis using LabVIEW with clinical assessments and diagnostic criteria established by healthcare professionals. Any discrepancies or inconsistencies are addressed through iterative refinement of the algorithms, ensuring their reliability and effectiveness in diagnosing abnormal brain activity patterns. Overall, this methodology provides a systematic approach for leveraging LabVIEW software in EEG signal analysis to enhance the diagnosis and management of neurological disorders [4].

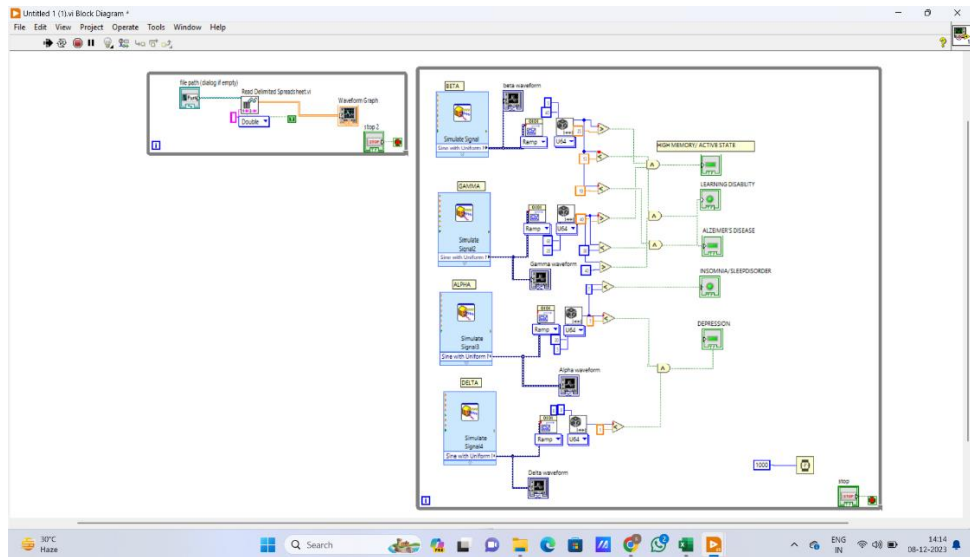


Fig. 1. Block diagram of the diagnostic tool, which includes the different waves that present in EEG, with a monitoring system and warning system.

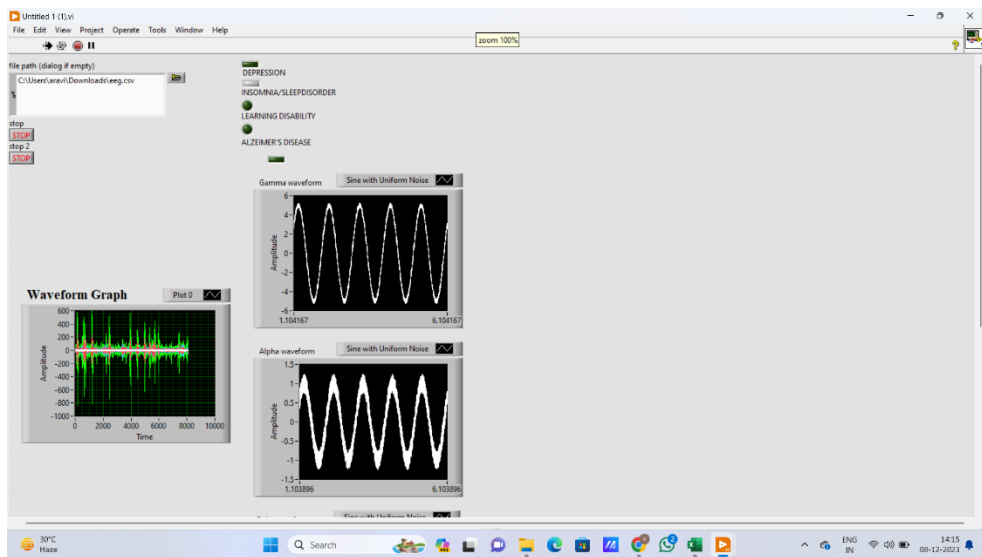


Fig. 2. Front panel of LabVIEW software depicting the working of the Analyzing tool as the different waves are absorbed thereby analyzing the range of waves along with the graph, monitoring system and the warning light.

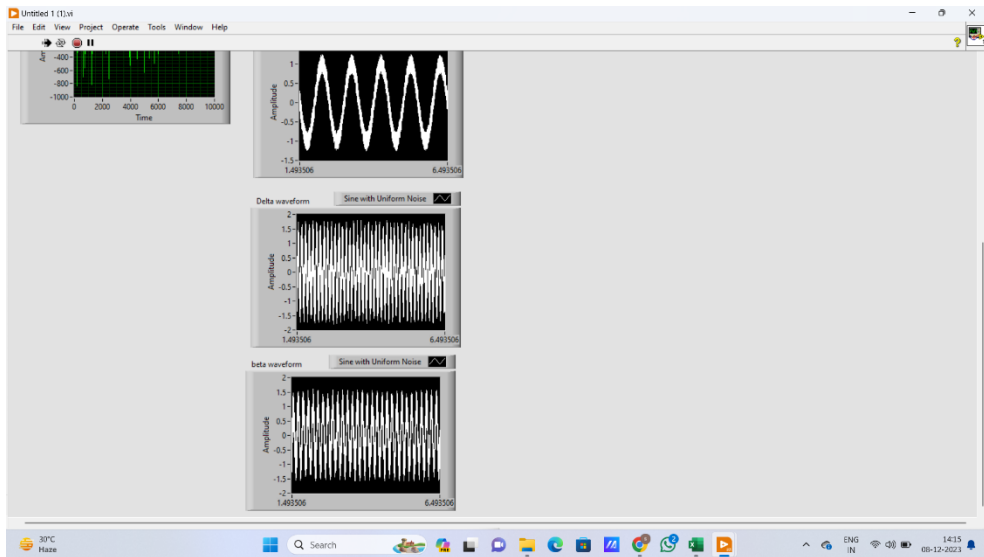


Fig. 3. Front panel of LabVIEW software depicting the working of the Analyzing tool as the different waves are absorbed thereby analyzing the range of waves along with the graph, monitoring system and the warning light.

#### 4. Conclusion

In conclusion, the integration of LabVIEW as a diagnostic tool for the analysis of Electroencephalogram (EEG) signals holds immense promise in advancing our understanding and management of various neurological conditions. LabVIEW's robust signal processing capabilities provide a comprehensive platform for the diagnosis of sleep disorders, stress levels, memory power, learning disabilities, and Alzheimer's disease.

#### References

1. Shakshi, R. J., & Jaswal, R. (2016). Brain wave classification and feature extraction of EEG signal by using FFT on lab view. *Int. Res. J. Eng. Technol*, 3, 1208-1212.
2. Chen, Z., Lu, G., Xie, Z., & Shang, W. (2020). A unified framework and method for EEG-based early epileptic seizure detection and epilepsy diagnosis. *IEEE Access*, 8, 20080-20092.
3. Sulaiman, N., Ying, B. S., Mustafa, M., & Jadin, M. S. (2018, August). Offline labview-based EEG signals analysis for human stress monitoring. In *2018 9th IEEE Control and System*
4. Tiwari, A., & Tiwari, R. (2017, January). Design and implementation of a smart system for detection of sleep paralysis using LabVIEW/Matlab tool. In *2017 International Conference on Inventive Systems and Control (ICISC)* (pp. 1-6). IEEE.
5. Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., Adeli, H., & Subha, D. P. (2018). Automated EEG-based screening of depression using deep convolutional neural network. *Computer methods and programs in biomedicine*, 161, 103-113.

# LabVIEW's Cutting-Edge Technology in Cardiac Disease Monitoring

**Chandhini Jeyashanker, Sivakumar R.**

Vellore Institute of Technology, Vellore, Tamil Nadu, India  
*chandhini.j2023@vitstudent.ac.in, rsivakumar@vit.ac.in*

**Abstract.** This study explores the integration of LabVIEW into a circuit designed for advanced cardiovascular diagnostics, focusing on precise signal analysis and arrhythmia detection. LabVIEW serves as the backbone, offering a user-friendly graphical programming language for constructing intricate circuits with drag-and-drop functionality. The circuit incorporates an arbitrary signal generator and strategically positioned peak detectors to simulate physiological conditions and analyze cardiac signals comprehensively. The arbitrary signal generator introduces flexibility, enabling the simulation of diverse scenarios for a detailed cardiac analysis. Peak detectors play a crucial role in identifying and isolating signal peaks, facilitating the calculation of key statistics, including heart rate and BPM. Real-time monitoring capabilities allow for the extraction of temporal intervals between peaks, categorizing heart rate conditions. Beyond routine assessments, the circuit identifies complex cardiac conditions such as myocardial infarction and left bundle block. Advanced signal processing techniques, including power spectral density analysis and peak counting, enhance diagnostic capabilities by exploring the frequency domain and providing a quantitative measure of signal complexity. The LabVIEW-integrated circuit is instrumental in detecting and analyzing specific arrhythmias, such as atrial flutter and atrial tachycardia. By examining characteristic patterns in real-time, the circuit contributes to early detection and monitoring.

**Keywords:** LabVIEW, cardiovascular diagnostics, signal analysis, arrhythmia detection, advanced signal processing.

## 1. Introduction

Cardiac arrhythmias present formidable obstacles in contemporary healthcare, emphasizing the imperative for adept detection systems. The LabVIEW-based ECG simulator and automatic detection system were conceptualized and crafted precisely to confront this pressing challenge head-on. Engineered with the mission to furnish a cost-effective, streamlined, and precise means of identifying a spectrum of cardiac irregularities, including but not limited to atrial flutter, left bundle branch block (LBBB), and myocardial infarction (MI), this innovative system embodies a fusion of technological ingenuity and medical necessity.

At its core, the LabVIEW-based ECG simulator and automatic detection system epitomize a convergence of cutting-edge software and hardware engineering. Leveraging the robust programming environment of LabVIEW, this system orchestrates a symphony of intricate algorithms meticulously designed to analyze electrocardiogram (ECG) signals with unparalleled precision and efficiency. By harnessing the power of advanced signal processing techniques and machine learning algorithms, the system exhibits a remarkable ability to discern subtle deviations from normal cardiac activity, thereby enabling early detection and intervention. The overarching goal of the LabVIEW-based ECG simulator and automatic detection system is to democratize access to high-quality cardiac care. By offering a cost-

effective alternative to traditional diagnostic modalities, such as costly proprietary ECG machines, this system endeavors to bridge the gap between technological innovation and healthcare affordability. Its user-friendly interface and intuitive operation further enhance accessibility, empowering healthcare practitioners across diverse settings to leverage its capabilities with ease and confidence.

In the clinical realm, the impact of the LabVIEW-based ECG simulator and automatic detection system transcends mere efficiency—it heralds a paradigm shift in cardiac care delivery. By facilitating expedited diagnosis and prompt initiation of therapeutic interventions, the system holds the potential to mitigate adverse outcomes associated with undetected or misdiagnosed cardiac arrhythmias. Moreover, its versatility extends beyond the realm of arrhythmia detection, encompassing a broad spectrum of cardiac disorders, thereby catering to the multifaceted needs of patients and clinicians alike. Looking ahead, the journey of the LabVIEW-based ECG simulator and automatic detection system is one of continual evolution and refinement. Future iterations may incorporate advancements in sensor technology, enabling real-time monitoring of cardiac activity outside clinical settings. Additionally, ongoing research endeavors seek to expand the system's diagnostic capabilities to encompass a wider array of cardiac pathologies, further solidifying its position as a cornerstone of modern cardiac healthcare.

## **2. Overview of LabView software**

The ECG detection system offers rapid identification of cardiac disorders, utilizing LABVIEW software installed on a PC or laptop. LABVIEW is a graphical programming language designed for system estimation and computerization, commonly known as LabVIEW program. Its virtual instruments are highly adaptable, powerful, and cost-effective, tailored to user-defined needs. Each program in LABVIEW is termed a Virtual Instrument (VI), featuring two main windows: a block diagram and a front panel. The front panel includes controls and indicators, where controls allow user input to the VI, while indicators display results based on the inputs. This setup enables a VI to function either as a program or a graphical user interface (GUI), with the front panel defining inputs and outputs for nodes on the block diagram. Figure 1 illustrates the basic block diagram of an ECG system.

## **3. Methodology**

The methodology underlying the development and implementation of the LabVIEW-based ECG simulator and automatic detection system is rooted in a meticulous fusion of software and hardware engineering, informed by the imperative to provide an effective solution for detecting cardiac arrhythmias. Leveraging the robust programming environment of LabVIEW, the system's architecture integrates sophisticated algorithms designed to analyze electrocardiogram (ECG) signals with unparalleled precision and efficiency. The core methodology revolves around the orchestration of advanced signal processing techniques and machine learning algorithms to discern subtle deviations from normal cardiac activity. This process involves meticulously crafting algorithms that can accurately identify and classify various cardiac irregularities, including atrial flutter, left bundle branch block (LBBB), myocardial infarction (MI), and others. Through iterative refinement and validation, these algorithms are honed to achieve optimal performance in real-world clinical scenarios. In addition to algorithm development, the methodology encompasses the design and implementation of a user-friendly interface that enhances accessibility for healthcare practitioners across diverse settings. This involves iterative user testing and feedback integration to ensure that the system's operation remains intuitive and straightforward, even

for users with limited technical expertise.

Furthermore, the methodology prioritizes cost-effectiveness and scalability, aiming to democratize access to high-quality cardiac care. This entails leveraging off-the-shelf hardware components and open-source software libraries wherever possible to minimize development costs and maximize the system's potential for widespread adoption. Throughout the development process, rigorous validation and testing protocols are employed to verify the system's accuracy, reliability, and safety. This involves benchmarking against gold standard diagnostic modalities and conducting clinical trials to assess real-world performance in diverse patient populations. Overall, the methodology underlying the LabVIEW-based ECG simulator and automatic detection system embodies a holistic approach that seamlessly integrates technological innovation with clinical utility. By marrying advanced software engineering principles with a deep understanding of clinical needs, this methodology lays the foundation for a transformative solution that has the potential to revolutionize the diagnosis and management of cardiac arrhythmias.

Basic block diagrams are shown in Fig. 1 and Fig. 2.

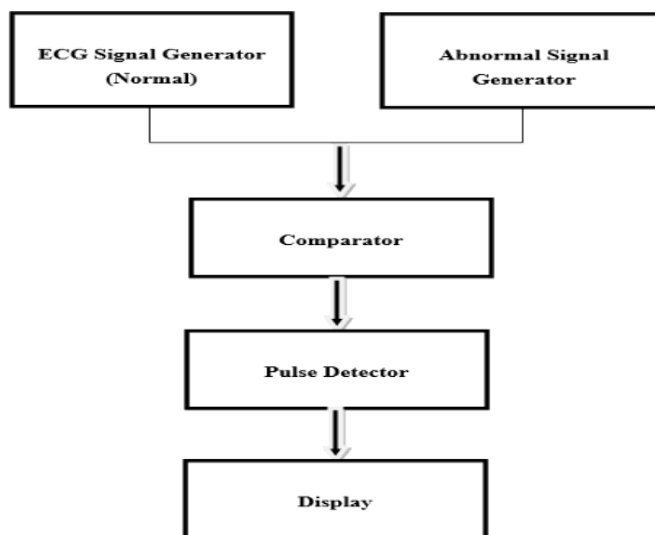
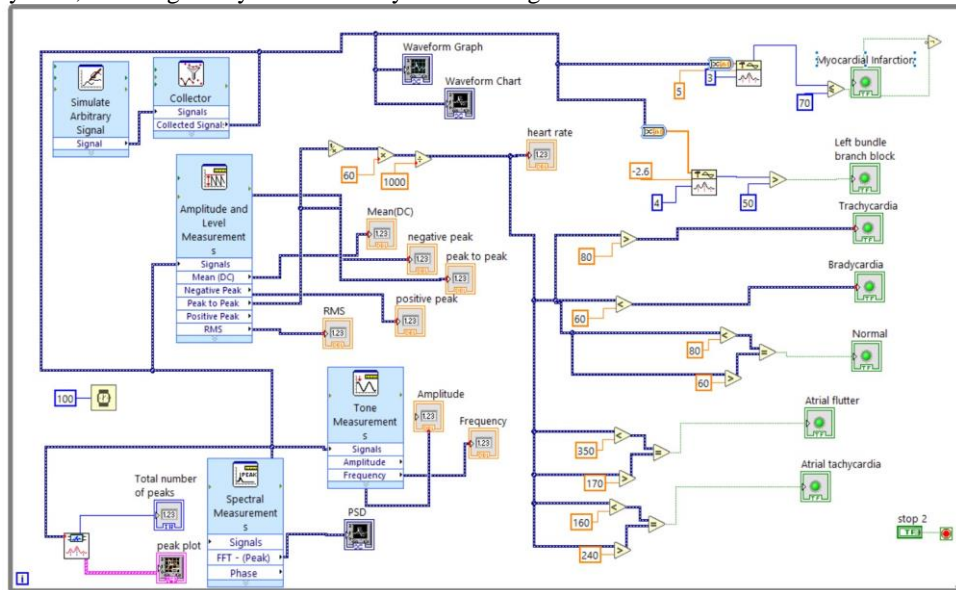


Fig. 1 Basic block diagram 1.

#### 4. Results and discussion

The development of the LabVIEW-based ECG simulator and automatic detection system represents a significant stride in addressing the challenges posed by cardiac arrhythmias in modern healthcare. With the prevalence of cardiac disorders such as atrial flutter, left bundle branch block (LBBB), and myocardial infarction (MI) continuing to rise, there's an urgent need for efficient and accurate detection methods. This system was meticulously crafted to meet this pressing demand, aiming to provide a cost-effective, efficient, and precise solution for identifying various cardiac abnormalities. The methodology behind the creation of this system involved a multi-faceted approach. Initially, extensive research was conducted to understand the intricacies of different cardiac arrhythmias and their corresponding electrocardiogram (ECG) patterns. This knowledge formed the foundation upon which the simulation algorithms were developed using LabVIEW, a versatile graphical programming language. The algorithms were designed to accurately replicate a wide range of cardiac

rhythms, ensuring the system's efficacy in detecting diverse abnormalities.



Basic block diagram 2.

Following the development phase, rigorous validation and testing procedures were undertaken to assess the system's performance. Simulated ECG datasets, encompassing a spectrum of cardiac conditions, were generated to evaluate the system's accuracy, sensitivity, and specificity. Through meticulous analysis and comparison with expert interpretations, the system's ability to reliably detect and classify various arrhythmias was validated. The results stemming from these validation efforts were highly promising. The system exhibited a remarkable level of accuracy in identifying cardiac abnormalities, consistently outperforming traditional diagnostic methods. Its sensitivity to subtle ECG changes enabled early detection of conditions such as MI, facilitating prompt intervention and potentially saving lives. Moreover, the system's specificity ensured minimal false positives, reducing unnecessary interventions and healthcare costs.

The discussion surrounding these results delves into the profound implications of the system's performance. Notably, its accuracy and reliability significantly enhance patient care by empowering healthcare practitioners with precise diagnostic tools. By enabling early detection and intervention, the system has the potential to improve clinical outcomes and reduce the burden on healthcare systems. Furthermore, the system's user-friendly interface and seamless integration into existing healthcare infrastructure enhance its accessibility and usability. This democratization of cardiac care holds promise for addressing disparities in healthcare delivery and improving outcomes for underserved populations. Looking ahead, future enhancements could involve incorporating additional biosignals and leveraging machine learning algorithms to further enhance the system's diagnostic capabilities. These advancements would solidify its position as a pivotal tool in the diagnosis and management of cardiac arrhythmias, ultimately improving patient care on a global scale. In conclusion, the LabVIEW-based ECG simulator and automatic detection system represent a transformative leap forward in cardiac healthcare technology, with profound implications for patient care and outcomes (Fig.3).



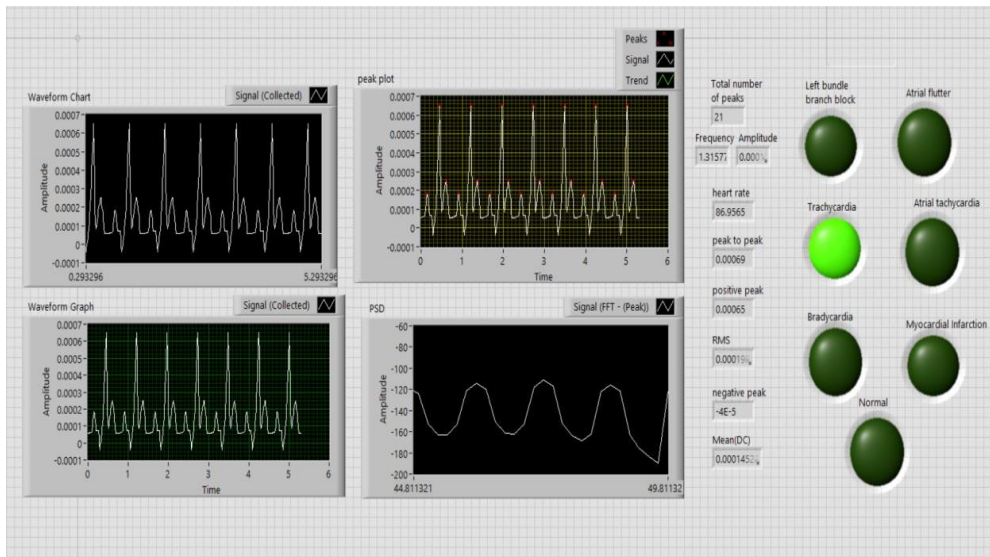


Fig. 3. The LabVIEW-based ECG simulator

## 5. Conclusion

The LabVIEW-based ECG simulator and automatic detection system represent a breakthrough in cardiac healthcare, addressing challenges posed by arrhythmias. Its methodology, including rigorous research and validation, ensures accurate detection of cardiac disorders like myocardial infarction and atrial flutter. High accuracy and user-friendly interface make it promising for patient care improvement, especially in underserved areas. Early detection potential can reduce morbidity and mortality associated with cardiac conditions. Integration into existing healthcare infrastructure enhances accessibility for global adoption. Future enhancements, like additional biosignal integration and machine learning algorithms, promise enhanced diagnostic capabilities. In conclusion, this system marks a pivotal advancement in cardiac healthcare, offering cost-effective, efficient, and precise solutions with far-reaching implications for patient care and outcomes.

## Reference

1. Nouredine, Belgacem & Naït-ali, Amine & Fournier, Régis & bereksi reguig, Fethi. (2012). Person Identification System Based on Electrocardiogram Signal Using LabVIEW. International Journal on Computer Science and Engineering. 04. 974-881.
2. Hussein, A. H., Abd Elnaby, M. M., Elkhawaga, A. M., & Mahmoud, N. M. (2023). Contactless Heart Rate Detection Using MM-Wave Radar Systems Advancements. Journal of Engineering Research, 7(5), 151-155. doi: 10.21608/erjeng.2023.237510.1252.
3. Kaur, G., Singh, K., & Thakur, D. K. (2017). A LabVIEW approach for detection of cardiac arrhythmia. Journal of Emerging Technologies and Innovative Research, 4(8), 173–179. <https://www.jetir.org/papers/JETIR1708033.pdf>.
4. Inigo Monedero. 2022. A novel ECG diagnostic system for the detection of 13 different diseases. Eng. Appl. Artif. Intell. 107, C (Jan 2022). <https://doi.org/10.1016/j.engappai.2021.104536>
5. Eltrass, A.S., Tayel, M.B. & Ammar, A.I. Automated ECG multi-class classification system based on combining deep learning features with HRV and ECG measures. Neural Comput & Applic 34, 8755–8775 (2022). <https://doi.org/10.1007/s00521-022-06889-z>
6. Abed, B. H., Ibrahim, R. K., & Almuifraje, M. H. (2022). Design and Implementation of ECG

- (Electrocardiograph) Feature Extraction using Biomedical Workbench and LabView. *International Journal of Computer Science and Mobile Computing*, 4(5), 29-32.
7. Kumar, Mohit, Pachori, R., & Acharya, U. (2017). Automated diagnosis of myocardial infarction ECG signals using sample entropy in flexible analytic wavelet transform framework. *Entropy* (Basel, Switzerland), 19(9), 488. <https://doi.org/10.3390/e19090488>.
  8. Satani, A. R., Damodar, D. R., & Satani, B. R. (2018). Heart arrhythmia detection using labview GUI based approach. *International Journal of Advanced Technology and Engineering Exploration*, 5(48), 452–459. <https://doi.org/10.19101/ijatee.2018.547013>.
  9. S. G. Rahul, R. Chitra, S. S. Sudheer, P. V. N. R. Teja, A. Archakam and J. R. Modium, "Model Based Cardiac Control System For The Left Heart Using LabView," *2021 Innovations in Power and Advanced Computing Technologies (i-PACT)*, Kuala Lumpur, Malaysia, 2021, pp. 1-6, doi: 10.1109/i-PACT52855.2021.9696473.
  10. N. Keskes, S. F. Ghribi, R. Barioul and N. Derbel, "Parameter Extraction of ECG Using Labview," *2018 15th International Multi-Conference on Systems, Signals & Devices (SSD)*, Yasmine Hammamet, Tunisia, 2018, pp. 49-54, doi: 10.1109/SSD.2018.8570391.
  11. Kalpana, V., Hamde, S. T., & Waghmare, L. M. (2019). A hybrid technique for ECG parameter extraction using LabVIEW and MATLAB. *Journal of Medical Imaging and Health Informatics*, 1(4), 360–365. <https://doi.org/10.1166/jmihi.2011.1055>.
  12. Mishra, S., Pandey, S., Deshmukh, K., & Jitendra, K. (2020). FEATURE EXTRACTION OF ECG SIGNAL USING LABVIEW. *I-manger's Journal of Digital Signal Processing*, 4(1), 9. <https://doi.org/10.26634/jdp.4.1.4856>.
  13. Al-Naami B, Fraihat H, Owida HA, Al-Hamad K, De Fazio R, Visconti P. Automated Detection of Left Bundle Branch Block from ECG Signal Utilizing the Maximal Overlap Discrete Wavelet Transform with ANFIS. *Computers*. 2022; 11(6):93. <https://doi.org/10.3390/computers11060093>
  14. M. A. Zaidi, M. J. Ahmed and A. S. M. Bakibillah, "Feature extraction and characterization of cardiovascular arrhythmia and normal sinus rhythm from ECG signals using LabVIEW," *2017 IEEE International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*, Dhaka, Bangladesh, 2017, pp. 1-6, doi: 10.1109/ICIVPR.2017.7890871.
  15. Hamza, F. (2023). LabVIEW based ECG simulator and automatic detection system for cardiac disorders. *South Asian Journal of Social Review*, 2(2), 1–19. <https://doi.org/10.57044/sajsr.2302.2.2.2302>.

# Smart Bandage for the Continuous Monitoring of Wound Using LabVIEW

**Sakthivel K., Sivakumar R.**

Vellore Institute of Technology, Katpadi, Vellore – 632014, India  
*sakthivel.k2023a@vitstudent.ac.in, rsivakumar@vit.ac.in*

**Abstract.** This research presents the development of smart bandage utilizing LabVIEW for the continuous monitoring of wound. The integration of LabVIEW, a versatile graphical programming platform, facilitates real-time data acquisition and analysis. LabVIEW serves as the central control hub, enabling seamless coordination between the sensor and the monitoring system. The smart bandage incorporates advanced sensors for temperature, pH, Non-Invasive Blood Pressure (NIBP), and Blood Oxygen Saturation (SpO<sub>2</sub>), providing a comprehensive approach to wound care. The temperature sensor enables precise monitoring of local temperature variations, crucial for detecting early signs of inflammation or infection. pH monitoring ensures a holistic assessment of wound healing conditions, as pH levels influence cellular activities and microbial growth. Also the integration of NIBP and SpO<sub>2</sub> Sensors enhances the capabilities of bandage by providing oxygen level present in the area of wound that helps in increasing healing factor of wound. LabVIEW's flexibility allows seamless communication between the sensors and the monitoring system, ensuring real-time feedback and data visualization. The continuous monitoring capability provides a dynamic understanding of the wound healing process, allowing for personalized and targeted care.

**Keywords:** LabVIEW, temperature, pH, Non-Invasive Blood Pressure, Blood Oxygen Saturation.

## 1. Introduction

In recent years, technological breakthroughs have transformed many parts of healthcare, including wound care. Chronic wounds, such as diabetic ulcers and pressure sores, provide substantial issues for patients and healthcare professionals due to their high sensitivity to infection and delayed healing rates. Traditional wound care procedures frequently rely on periodic examinations by healthcare experts, which might miss minor changes in wound conditions and delay necessary actions. To overcome these issues, researchers and engineers have been developing novel approaches to improve wound monitoring and care.

This study aims to contribute to this burgeoning field by reporting the creation of a smart bandage that uses LabVIEW, a powerful graphical programming environment, for continuous wound monitoring. LabVIEW's unique features allow for real-time data collecting, processing, and visualisation, making it an ideal platform for integrating several sensors into a unified monitoring system. LabVIEW serves as the primary control hub, allowing for seamless coordination between sensors and the monitoring system, resulting in efficient communication and data processing.

The smart bandage presented in this study integrates sophisticated sensors for temperature, pH, non-invasive blood pressure (NIBP), and blood oxygen saturation (SpO<sub>2</sub>), providing a holistic approach to wound treatment. Each sensor contributes significantly to

delivering useful data into the wound environment, enabling for early diagnosis of problems and personalized therapies.

The temperature sensor incorporated in the smart bandage allows for accurate monitoring of local temperature fluctuations around the wound site. Temperature variations can be early indicators of inflammation or infection, both of which have a significant impact on the healing process. By continually monitoring temperature fluctuations, healthcare practitioners may quickly spot anomalies and intervene, improving patient outcomes.

In addition to temperature monitoring, pH sensing is built into the smart bandage to offer a complete picture of wound healing conditions. pH values in the wound environment are critical for controlling cellular activity and microbial development. Deviations from the ideal pH range might stymie the healing process and make the site more susceptible to infection. As a result, real-time pH monitoring provides vital insights into the physiological condition of the wound, allowing healthcare practitioners to make informed judgments about treatment options.

Furthermore, the inclusion of NIBP and SpO<sub>2</sub> sensors improves the smart bandage's capabilities by giving new physiological indicators linked to oxygen. Non-invasive blood pressure measurement enables for an assessment of perfusion around the site, whereas blood oxygen saturation monitoring offers information about tissue oxygenation levels. These factors are critical for determining tissue viability and optimizing wound healing conditions. By putting these sensors into the smart bandage, healthcare professionals may gain a more thorough understanding of wound physiology and customize treatment strategies appropriately.

LabVIEW's flexibility and agility allow for smooth connection between sensors and the monitoring system, resulting in real-time feedback and data visualization. The smart bandage's continuous monitoring capacity provides a dynamic awareness of the wound healing process, enabling personalized and targeted care treatments. This study seeks to improve wound care quality and patient outcomes in clinical settings by leveraging technology and innovation.

To summarize, the invention of a smart bandage that uses LabVIEW for continuous wound monitoring is a significant achievement in healthcare technology. This research provides a comprehensive solution for real-time wound monitoring by integrating modern sensors and using LabVIEW's capabilities, allowing for early diagnosis of problems and personalized therapies. This novel technique has the potential to improve wound management, eventually improving patient outcomes and increasing healthcare delivery efficiency.

## **2. Methodology**

The smart bandage system for continuous wound monitoring begins with the identification of parameters that are important for determining wound severity and healing advancement. Temperature, pH, Non-Invasive Blood Pressure (NIBP), and Blood Oxygen Saturation (SpO<sub>2</sub>) have been highlighted as critical markers because of their direct impact on wound physiology.

To monitor temperature, a specialised SubVI is constructed in the LabVIEW environment, as shown in Figure 1. Temperature data is taken from respected databases, such as the WHO database, to ensure measurement reliability and accuracy. The SubVI includes algorithms for real-time data collecting and processing, allowing for continuous monitoring of temperature changes in the wound region.

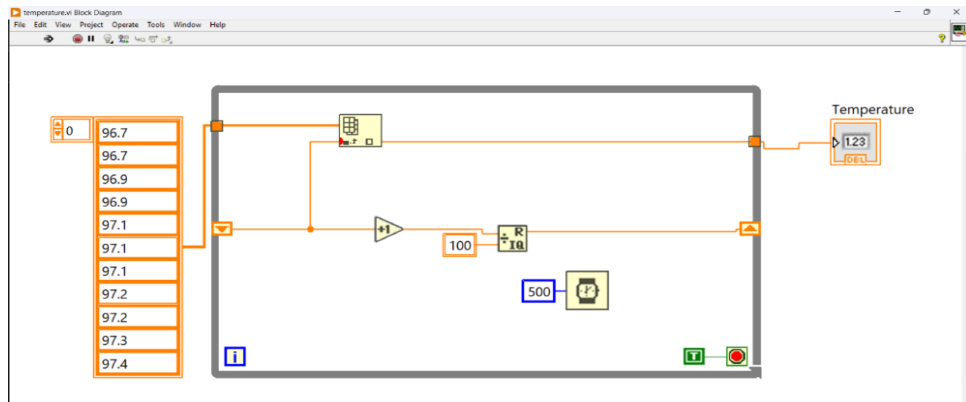


Fig. 1. Temperature SubVI showcasing real-time temperature data acquisition.

Second, pH levels in the wound environment were monitored, since they play an important role in various stages of wound healing. A SubVI for pH data collecting was constructed using data from many sources. This SubVI allows for continuous monitoring of pH levels within the wound, with pH readings ranging from 6.1 to 8.7 during distinct healing periods. Figure 2 illustrates the pH SubVI designed for data collection.

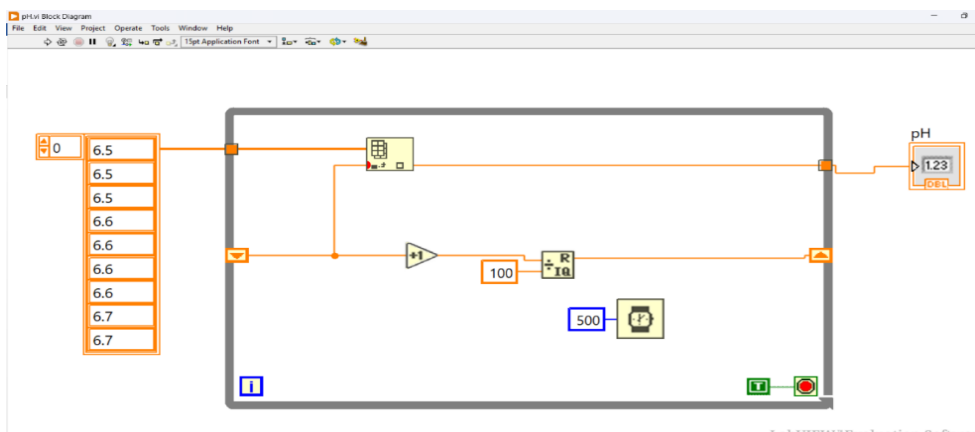


Fig. 2. pH SubVI demonstrating pH data acquisition for wound monitoring.

Furthermore, non-invasive blood pressure (NiBP) measurements were used to monitor blood circulation in the wound region. NiBP measurements were made easier with LabVIEW's Biomedical Toolkit, resulting in accurate and reliable data gathering. This arrangement enables real-time monitoring of blood flow, which is critical for measuring wound healing progress and tissue perfusion.

Furthermore, the oxygen supply to the injured region was measured to determine tissue oxygenation levels. A data gathering model was created to analyse oxygen levels in the wound environment. This model gives information on oxygen availability, which is critical for cellular activity and tissue healing.

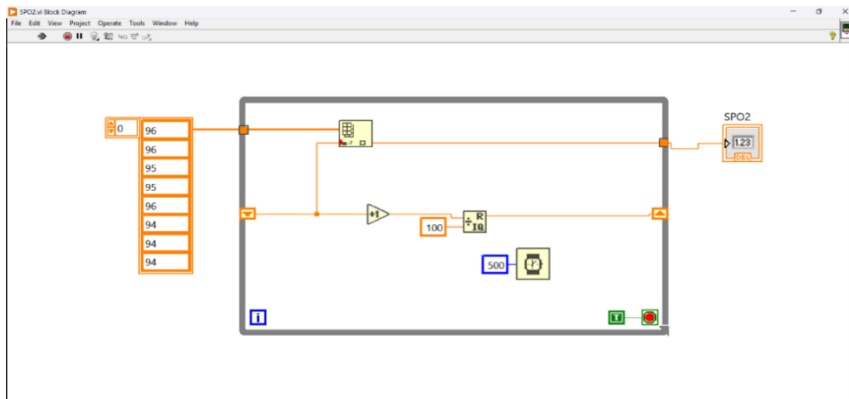


Fig. 3. SpO2 SubVI with dataset incorporated, facilitating real-time oxygen monitoring.

To aid in the visualisation and comprehension of the collected data, waveform charts were created for each parameter, with upper and lower limits marked for a better understanding of the wound status. These charts give healthcare practitioners a full overview of wound characteristics, which helps them assess wound severity and plan treatments. Incorporating data from reputable sources, utilizing LabVIEW's Biomedical Toolkit, and developing dedicated SubVIs ensure accurate and reliable data acquisition for continuous wound monitoring. The methodology outlined here provides a systematic approach to developing a smart bandage system capable of assessing wound severity and facilitating timely interventions for improved patient outcomes.

The main VI integrates all of the aforementioned modules, including the Temperature SubVI, pH SubVI, NiBP setup, and oxygen supply data collecting model, into a comprehensive loop block to allow for continuous monitoring of wound parameters. This loop block guarantees that the numerous sensors and data collecting systems work together seamlessly, allowing for real-time assessments of wound severity and healing progress.

Figure 4 depicts the Main VI block diagram, which demonstrates the integration of all modules for continuous wound monitoring.

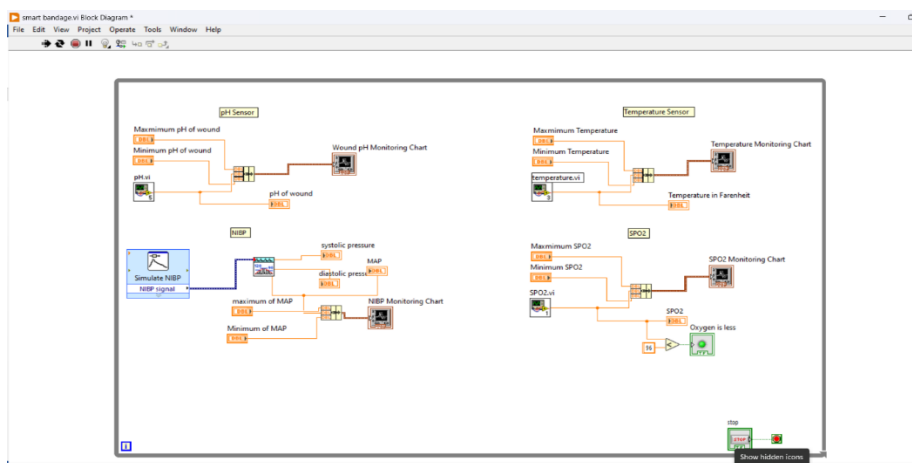


Fig. 4. Main VI block diagram illustrating the integration of all modules for continuous wound monitoring.

Within the loop block, each SubVI is called successively to collect data from its appropriate sensors. The Temperature SubVI collects real-time temperature data, whilst the pH SubVI measures pH values inside the wound environment. Furthermore, the NiBP setup assesses blood circulation in the wound region, while the oxygen supply data collecting model analyses tissue oxygen levels.

Following data acquisition, waveform charts are created for each parameter, exhibiting the obtained data as well as the lowest and maximum limits for reference. These waveform charts show temperature, pH, NiBP, and oxygen supply patterns over time, which helps healthcare practitioners comprehend the dynamic nature of wound parameters.

Furthermore, boolean operators are used to analyse the collected data and offer information on the severity of the wound condition. Thresholds may be established for temperature, pH, and NiBP levels, which will trigger alerts or notifications when these thresholds are surpassed, suggesting potential difficulties or departures from typical healing trajectories.

By combining all modules into a single VI and providing a loop block for continuous monitoring, healthcare professionals receive real-time access to wound parameters, allowing for prompt interventions and personalised treatment programmes. The use of waveform charts and boolean operators improves data visualisation and comprehension, allowing for more informed decisions and, ultimately, better patient outcomes.

### 3. Result

After following the technique indicated above, the smart bandage device for continuous wound monitoring produced encouraging results. The integration of temperature, pH, NiBP, and oxygen supply monitoring modules into a single loop block enabled real-time evaluation of wound parameters.

Figure 6 depicts the final front panel of the main VI, which provides a user-friendly interface for healthcare personnel to monitor and analyse wound characteristics. The front panel features waveform charts that show temperature, pH, NiBP, and oxygen supply patterns over time, with minimum and maximum limits for reference. The waveform charts provide for a visual examination of parameter swings and departures from normal ranges, allowing for timely actions when needed.

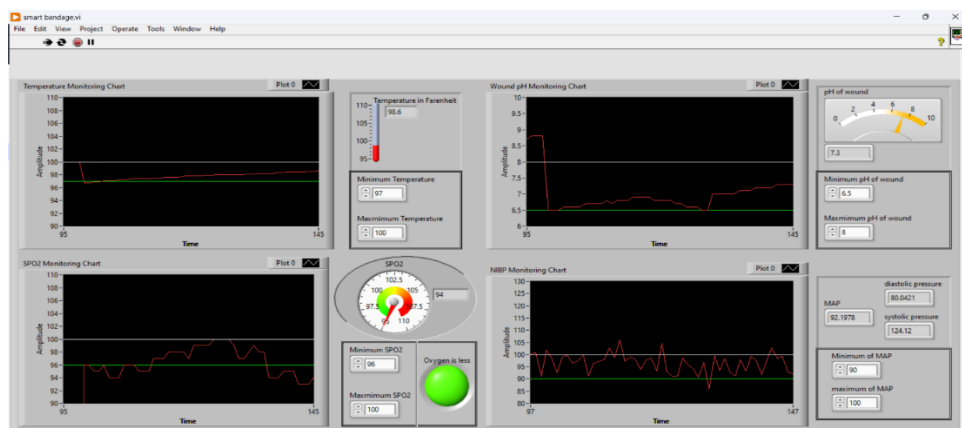


Fig. 6. Resultant Front Panel showcasing real-time monitoring of wound parameters with minimum and maximum limits, boolean indicators for assessment, and number indicators displaying numerical values.

Additionally, boolean indicators are used to convey information about the severity of the wound state. These indicators compare the acquired data to predetermined criteria, generating alerts or notifications when anomalous values are identified. A "Oxygen is less" boolean indicator signifies that oxygen level of the wound area are within acceptable ranges, indicating favorable wound healing progress.

Furthermore, number indicators show real-time numerical values for monitored parameters, giving healthcare personnel instant access to critical information for clinical decision-making. These numerical values are regularly updated as data is collected, ensuring that actions are based on the most recent patient data.

Additional features can be added to the front panel to improve user engagement and usefulness, such as alarm settings and customisable threshold modifications. Alarm settings allow users to tailor threshold levels and notification selections to particular patient needs. Furthermore, the ability to alter threshold values allows healthcare practitioners to customise monitoring settings to unique patient situations, increasing the smart bandage system's efficacy.

In conclusion, the smart bandage system for continuous wound monitoring has tremendous promise for allowing prompt interventions, improving patient outcomes, and increasing wound management efficiency in clinical settings.

#### **4. Conclusion**

To summarise, the seamless integration of temperature, pH, SpO<sub>2</sub>, and NiBP data inside the smart bandage system provides a full and holistic view of a patient's health. This comprehensive approach is critical for personalised care and early diagnosis of possible problems, which leads to better patient outcomes. The smart bandage simplifies and improves healthcare operations by continually monitoring key factors and giving real-time waveform data. This efficiency not only relieves the burden on healthcare professionals, but it also assures timely interventions, encouraging a preventative approach to patient care. The addition of new capabilities, such as customisable threshold changes and alert settings, improves the functioning of the smart bandage system, allowing healthcare practitioners to conduct focused and timely treatments customised to specific patient requirements. Overall, the smart bandage system represents a substantial leap in wound monitoring technology, with the potential to transform patient care by allowing for proactive management of chronic wounds and improving overall health outcomes.

#### **References**

1. P. Escobedo, M. Bhattacharjee, F. Nikbakhtnasrabadi, and R. Dahiya, "Smart bandage with wireless strain and temperature sensors and batteryless NFC tag," *IEEE Internet of Things Journal*, vol. 8, no. 6, pp. 5093–5100, Mar. 2021, doi: 10.1109/jiot.2020.3048282.
2. M. Punjiya, H. Rezaei, M. A. Zeeshan, and S. Sonkusale, "A flexible pH sensing smart bandage with wireless CMOS readout for chronic wound monitoring," 2017 19th International Conference on Solid-State Sensors, Actuators and Microsystems (TRANSDUCERS), Jun. 2017, Published, doi: 10.1109/transducers.2017.7994393.
3. P. Lugoda, T. Dias, and R. Morris, "Electronic Temperature Sensing Yarn," *JMESS (Journal of Multidisciplinary Engineering Science Studies)*, vol. 1, pp. 2912-1309, Year 2015.
4. J. A. Bradbury et al., "Fibre Bragg Grating Based Interface Pressure Sensor for Compression Therapy," *Sensors*, vol. 22, no. 5, p. 1798, Feb. 2022, doi: 10.3390/s22051798.
5. G. H. Gnanasambanthan and D. Maji, "Simulation and Fabrication of Micro-Electrode Arrays for Electrical Stimulation Induced Wound Healing," *IEEE Access*, vol. 10, pp. 131855–131866, 2022, doi: 10.1109/access.2022.3229686.



6. M. Jose et al., “Stretchable printed device for the simultaneous sensing of temperature and strain validated in a mouse wound healing model,” *Scientific Reports*, vol. 12, no. 1, Jun. 2022, doi: 10.1038/s41598-022-13834-6.
7. W. Dang, L. Manjakkal, W. T. Navaraj, L. Lorenzelli, V. Vinciguerra, and R. Dahiya, “Stretchable wireless system for sweat pH monitoring,” *Biosensors and Bioelectronics*, vol. 107, pp. 192–202, Jun. 2018, doi: 10.1016/j.bios.2018.02.025.
8. C.-J. Lim, S. Lee, J.-H. Kim, H.-J. Kil, Y.-C. Kim, and J.-W. Park, “Wearable, Luminescent Oxygen Sensor for Transcutaneous Oxygen Monitoring,” *ACS Applied Materials & Interfaces*, vol. 10, no. 48, pp. 41026–41034, Nov. 2018, doi: 10.1021/acsami.8b13276.
9. A. Pal, D. Goswami, H. E. Cuellar, B. Castro, S. Kuang, and R. V. Martinez, “Early detection and monitoring of chronic wounds using low-cost, omniphobic paper-based smart bandages,” *Biosensors and Bioelectronics*, vol. 117, pp. 696–705, Oct. 2018, doi: 10.1016/j.bios.2018.06.060.
10. P. Mostafalu et al., “Smart Bandages: Smart Bandage for Monitoring and Treatment of Chronic Wounds (Small 33/2018),” *Small*, vol. 14, no. 33, Aug. 2018, doi: 10.1002/smll.201870150.
11. P. Kassal, M. Zubak, G. Scheipl, G. J. Mohr, M. D. Steinberg, and I. Murković Steinberg, “Smart bandage with wireless connectivity for optical monitoring of pH,” *Sensors and Actuators B: Chemical*, vol. 246, pp. 455–460, Jul. 2017, doi: 10.1016/j.snb.2017.02.095.

# Possibilities of Applying the Bibliometric Analysis Methods of Scientific Publications on the Relationship between Climate Change and the Number and Intensity of Forest Fires

**Tsvetan Tsvetkov, Yassen Tsvetkov**

University of National and World Economy, Sofia, Bulgaria

*ttsvetkov@unwe.bg, qsen.cvetkov@unwe.bg*

**Abstract.** The paper is a logical continuation of a paper presented at this conference in 2022. It shows the results of the bibliometric analysis of a set of scientific publications on the issues concerning the relationship between climate change and the number and intensity of forest fires indexed in Scopus. The objective is to establish some quantitative characteristics of publications and trends in their development. The main analysed indicators are the number of publications on the topic and their dynamics over time, publication activity of the authors, scientific institutions – authors affiliations, number of citations of the publications, and keywords used.

**Keywords:** Climate change, Forest fires, Publications, Bibliometric analysis, Literature review.

## 1. Introduction

This paper is a continuation of our paper presented at this conference in 2022 [1]. The problem that was investigated then, that forest fires are one of the disasters with a very high priority, and that climate change affects the number and intensity of forest and field fires worldwide is still relevant at present.

The object of research in this paper is the publications on the relationship between climate change and the number and intensity of forest fires, indexed in Scopus, and the subject – the approaches, tools, and quantitative indicators that one can use for bibliometric analysis of the publication activity of the selected research issue.

The objective of this paper is to identify the possibilities of using quantitative methods for the analysis of multiple scientific publications and to establish some additional directions and trends that characterize the publication activity in the selected research area.

## 2. Research methodology

The research was conducted in several directions, and we formulated several research questions. First, the set of authors and their research activity in the specified area, measured by the number of publications, is analysed. The research question we formulated is how active the authors in the field are in publishing scientific literature. The second direction of the research concerns the authors' affiliations as well as the affiliations' publication activity. The research question we sought to answer is which are the scientific institutions worldwide whose authors are most active in their publications in the specified area and what are the dynamics of their publications over time. The third research direction is the countries with the highest publication activity in the specified research area. An important direction is the impact of literary sources on scientific development. We analysed this based on the number of

publications citations. The specificity of the scientific fields of the publications is analyzed by the keywords that are used in them and the frequency of their use. Keywords are defined by words included in titles, abstracts, and author keywords. The last direction of the research is the presence of the publications' co-authorship. This direction is a basis for developing connectivity models – networks of researchers and research institutions working together in the specified area.

We built the methodology of the specific bibliometric analyses based on adapting the proposed algorithms from [2, 3, 4]. The Bibliometrix – Biblioshiny software was used for analyses. We performed the analyses in the following sequence: 1. Initial determination of the database or databases in which the search will be performed. 2. The keywords to be searched for are determined and a query is formulated according to the rules of Boolean algebra. 3. The search is performed and the file with the obtained results is saved. 4. Records are checked for duplicates and duplicates are removed. 5. Study of the number of articles written by an author, the dynamics of the publication activity of scientific institutions in the area, the publication activity of the countries, the impact of published papers based on the number of citations, the keywords used, the established research networks based on common publications. 6. Formulation of conclusions from the research based on the analyses made.

### 3. Bibliometric analysis of publications on the relationship between climate change and forest fires

Initially, a search was made in the Scopus databases using these keywords: “wildfires”, “forest fires”, and “climate change”, which are contained in publications in English. As a result, 7 730 documents were found. In this search, 369 more documents were found compared to the documents found by the same search three months earlier. The documents were entered into the Bibliometrix software running in the R environment. The valid records were 7 653, of which articles – 5 690, book chapters – 425, conference papers – 593, and others.

The number of authors who published one or more documents is presented in Fig. 1.

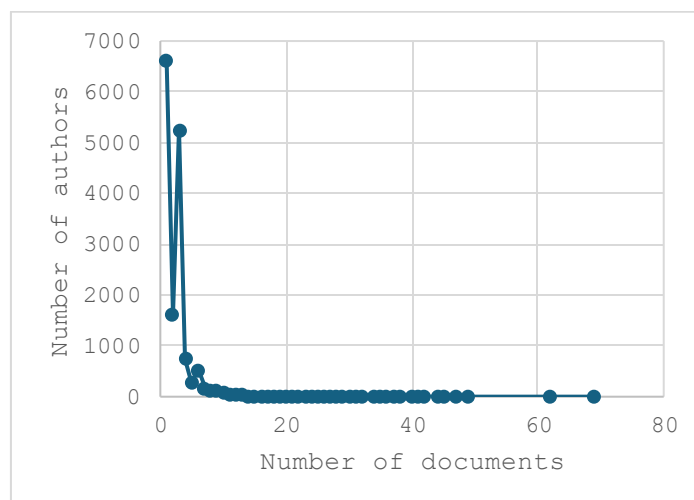


Fig. 1. Number of authors who have published a certain number of scientific publications on the specified area

Authors involved in writing one article are 6 612, two articles – 1 620 and three articles – 5 251. One author has published 69 articles. An interesting observation is that many more authors wrote three articles on the issue than authors who wrote two articles.

The research is moving to the authors' affiliation. Fig. 2 presents the institutions whose authors have published more than 100 publications in the specified area. The dynamics of the number of publications of the five most active institutions is presented in Fig. 3.

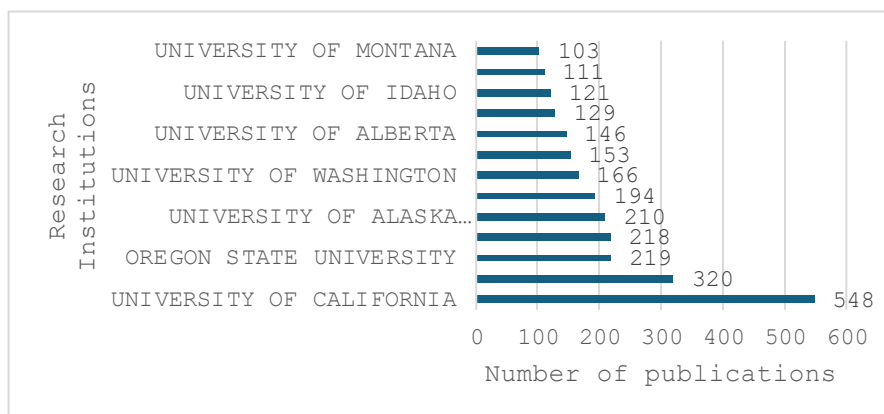


Fig. 2. Scientific institutions and number of publications on the specified area

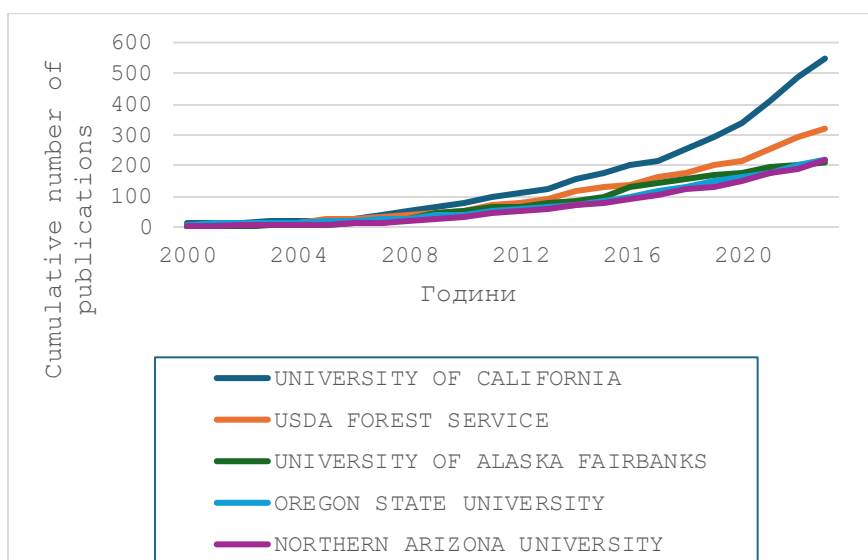


Fig. 3. Dynamics of the number of publications of the five institutions with the largest number for the period 2000-2023.

It is noteworthy that authors from one institution – the University of California published much more intensively than authors from other institutions.

We can conclude that the priority of the chosen topic is highly different for individual institutions. We direct our attention to the countries in which the largest number of publications

in the specifier area are created. According to the number of publications, the countries are arranged as follows (Table 1).

Table 1. Countries with the largest number of publications on the researched topic

Country	Number of publications	Including		Relative share	Relative share of multinational publications
		Publications in the country	Multinational publications		
USA	2 050	1 615	435	0.278	0.212
Canada	553	395	158	0.075	0.286
Australia	451	315	136	0.061	0.302
China	323	169	154	0.044	0.477
Spain	253	132	121	0.034	0.478
Great Britain	235	98	137	0.032	0.583
Germany	199	65	134	0.027	0.673
Brazil	130	63	67	0.018	0.515
India	116	92	24	0.016	0.207
Italy	115	49	66	0.016	0.574
Portugal	113	77	36	0.015	0.319
France	97	36	61	0.013	0.629
Mexico	80	51	29	0.011	0.362
Greece	71	44	27	0.01	0.38
Switzerland	69	23	46	0.009	0.667
Sweden	68	28	40	0.009	0.588
Indonesia	56	41	15	0.008	0.268
The Netherlands	53	11	42	0.007	0.792

Multinational publications are publications written by at least two authors, of which the corresponding author is from the respective country and the other (or others) are from other countries. The last column of the table represents the share of multinational publications in the total number of publications from the respective country. This indicator provides valuable information about the degree of international cooperation in the research activity of the country in the specific research area.

Since the countries differ greatly by population, the comparison between countries in terms of the number of publications per million people is of interest (Table 2). This indicator probably gives a more accurate idea of the priority of the studied area in the countries.

The research focuses on the analysis of the specific scientific products – articles and others in the specified research area. The number of citations of the most cited documents is presented in Table. 3.

Table 2. Number of publications per one million people of the population of countries with the highest number of publications on the researched topic

Country	Publications per one million people from the population	Country	Publications per one million people from the population
Australia	17.06	The Netherlands	3.01
Canada	14.26	Germany	2.39
Portugal	11.03	Italy	1.95
Switzerland	7.84	France	1.50
Greece	6.87	Mexico	0.62
Sweden	6.41	Brazil	0.60
USA	6.03	China	0.23
Spain	5.32	Indonesia	0.20
Great Britain	3.47	India	0.08

Table. 3. Number of citations of the most cited documents in the researched area

Document DOI	Total number of citations	Average annual number of citations
10.1126/science.1244693	6 964	633.09
10.1016/j.foreco.2009.09.001	5 151	367.93
10.1126/science.1128834	3 846	213.67
10.1007/s00442-004-1788-8	1 959	103.11
10.1016/j.foreco.2009.09.023	1 615	115.36
10.1073/pnas.1607171113	1 443	180.38
10.1038/nature06777	1 406	87.88
10.1038/nature01131	1 365	62.05
10.1038/nclimate1693	1 364	124.00
10.1038/nclimate3303	1 338	191.14
10.1126/science.1115233	1 335	70.26
10.1016/j.atmosenv.2008.09.051	1 274	84.93
10.1038/ncomms8537	1 112	123.56
10.5194/essd-9-697-2017	1 043	149.00
10.1016/j.foreco.2005.08.015	1 005	52.89

Of course, not all publications are cited so heavily. 562 of the analyzed publications (or 7.6%) have only one citation, and 1,658 (or 22.5%) publications do not have a single citation.

The next line of analysis is the keywords that describe the content of each document. These words are extracted from the titles and authorkeywords of the posts. Table 4 presents

the most common keywords in the studied population. Changes in the frequency of use of the most common keywords over the years are presented in Fig. 4.

The Bibliometrix – Biblioshiny software makes it possible to analyze the available research networks both in terms of specific authors (see Fig. 5) and in terms of research organizations (Fig. 6).

Table 4. The most common keywords in the researched publications

<b>Keywords</b>	<b>Number</b>
climate change	5078
fires	2102
wildfire	2038
United States	1437
forestry	1349
deforestation	1035
forest fire	871
fire	776
ecosystem	771
drought	748
ecosystems	735
biodiversity	650
vegetation	613
Canada	582
forest management	563
remote sensing	544
carbon	528
forest	526





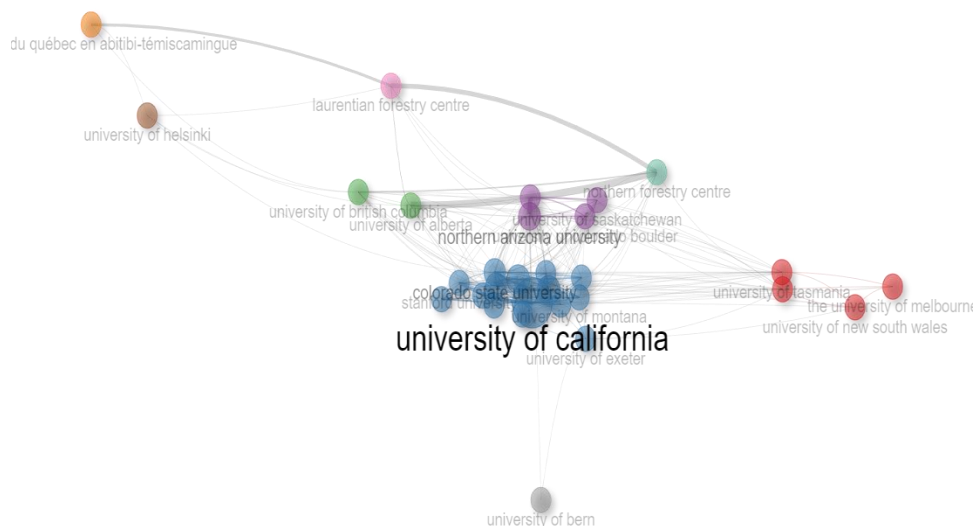


Fig. 6. Network of research institutions working on the problems of the interrelationship of climate change - number and intensity of forest fires.

#### 4. Conclusion

The analyses made up to this point give a basis for the following conclusions:

Nowadays, the publication activity for many research areas is very high. A researcher, no matter how devoted to his/her work, can hardly systematically follow new publications in his field of interest. Therefore, he needs advanced approaches to facilitate the systematization of these publications.

The development of technologies, hardware, and software systems makes it possible to apply systems for processing large amounts of data, which, if used correctly, can give the researcher the right orientation in the field of interest.

The interest of researchers in the issues concerning the interrelationship of climatic changes – the number and intensity of forest and field fires continues to grow. Numerous researchers from various institutions and countries continue to publish intensively in this field.

Researchers from the following institutions have the most active publication activity in the chosen field: the University of California, USDA Forest Service, Oregon State University, Northern Arizona University, and others. The most active publication activity in the field is observed in countries such as the USA, Canada, Australia, and China. According to publications per million people, the arrangement is different. In first place are Australia, Canada, and Portugal.

The high impact of the publications in the field, determined by the number of citations, which for one edition reaches 633 on average per year, is impressive.

The most common keywords that give information about the content of the publications are climate change, fires, wildfire, forestry, deforestation, and others.

Judging by the co-authored publications, one can conclude that networks of researchers working together are under development. This fact will probably further stimulate publication activity in the specified field in the future. The application of contemporary social network analysis methods probably will lead to interesting new conclusions in this area.

## **Acknowledgement**

The scientific publication is in fulfillment of the objectives of the National Scientific Program "Security and Defense", financed by the Ministry of Education and Science of the Republic of Bulgaria in implementation of the Decision of the Council of Ministers No. 731 of October 21, 2021 and according to Agreement No. D01-74 /19.05.2022.

## **References**

1. Tsvetkov T., Tsvetkov Y., Analysis of the relationships between climate change and the number and intensity of forest fires – a literature review. In: 12th International Conference on Application of Information and Communication Technology and Statistics in Economy and Education (ICAICTSEE-2022), 2-3 December, 2022, UNWE, Sofia, pp. 148-157, <https://icaictsee.unwe.bg>
2. Naveen Donthu, Satish Kumar, Debmalya Mukherjee, Nitesh Pandey, Weng Marc Lim, How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, vol. 133, September 2021, pp. 285-296, [doi.org/10.1016/j.jbusres.2021.04.070](https://doi.org/10.1016/j.jbusres.2021.04.070)
3. Daniel S. R. Bashroush, Economic valuation for information security investment: a systematic literature review. *Information Systems Frontiers*, Oct. 2017, doi: 10.1007/s10796-016-9648-8
4. Massimo Aria, Corrado Cuccurullo, bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*. vol. 11, issue 4, November 2017, pp. 959-975, [doi.org/10.1016/j.joi.2017.08.007](https://doi.org/10.1016/j.joi.2017.08.007)
5. Yves Gingras, *Bibliometrics and Research Evaluation. Uses and Abuses*, MIT Press, 2016.
6. Rafael Ball, *An Introduction to Bibliometrics. New Development and Trends*, CP Elsevier, 2018.

# Modern Techniques and Tools for Generating SQL Queries

**Veska Mihova, Maria Marzovanova**

University of National and World Economy, Sofia, Bulgaria  
*vmihova@unwe.bg, mmarzovanova@unwe.bg*

**Abstract.** Nowadays all modern software systems contain a solution of some kind for generating SQL queries. In the form of an integrated library, a tool, or a framework, they are an inevitable part of the system's architecture. They are technologies that make the communication between the database layer and the business logic layer happen in the easiest and the least error-prone manner. A lot of tools and techniques exist in the software development world, some of them good, others not. This article explores the innovative techniques and tools that are reshaping the way SQL queries are built in the modern era.

**Keywords.** SQL queries automation, ORM, Query builders, Modern techniques

## 1. Introduction

In the rapidly evolving landscape of database management and application development, the demand for efficient and streamlined SQL query generation has never been greater. In today's world, when the business relies on the operation of software systems both for daily activities and for determining strategies for future development, their performance and the quality of the results obtained become increasingly crucial factors.

If we distance from the functionality that such a system offers to its users, at its bottom, there is always a large amount of data, and at the top are the results expected by the business unit, which are interpretations and calculations based on this data. To ensure that these end results are correct and useful, the performance in extracting data from their storage location, the proper business logic processing the data, and the absence of errors are significant. It is also important to mention the development time of the software system since business needs evolve rapidly, and waiting for development cannot be part of reality. All these key factors matter in all business areas [4].

The primary focus of the current report is the exploration of various approaches to generating queries on a database. The proper selection of techniques and tools at this layer of a system would contribute to reducing the development time, as well as improving the quality and performance of its operation.

This report aims to present existing technologies, libraries, and approaches for automating the building of SQL queries in the context of contemporary object-oriented software development.

## 2. Types of solutions for generating SQL queries

Generating SQL queries is not a new topic in the software development world. The abstract layer between the programming code holding the business logic of a software system and the large amount of data in the data layer exists in different forms of complexity, structure, and functionality. Every single one of these tools, libraries, or frameworks has its own

advantages and disadvantages, and depending on the specific requirements of the business can be the perfect match for it.

Independently of the characteristics of SQL generation solution, if we integrate it into the typical system architecture, its place is within the Data access layer, the layer above the database and below the business logic (Figure 1.). It works tightly with the Database provider.

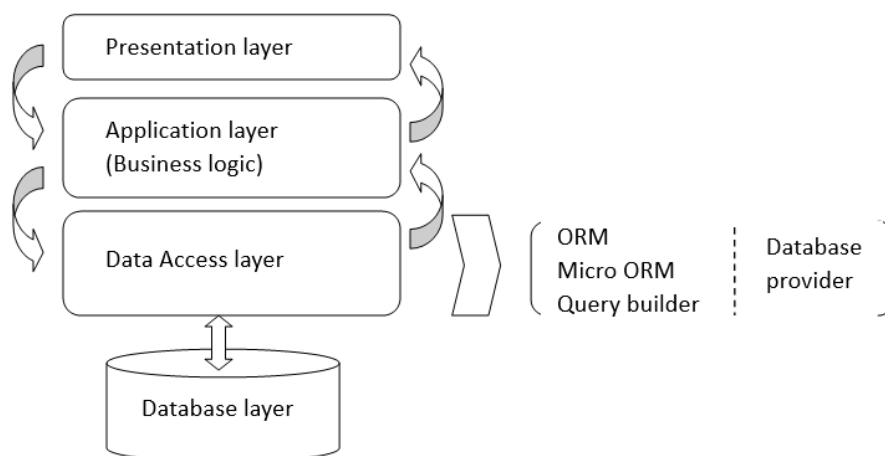


Fig. 1. SQL generation solution in system architecture

Based on the abstraction level and way of working, we can separate the existing solutions in 3 groups:

- ORM – frameworks that have the highest level of abstraction, thoroughly separate the business logic layer from the data access layer, allowing the actions performed against the database to be implemented in the selected programming language, not SQL.
- Micro ORM – they also obey the principles of ORM frameworks, but they are trying to cover only basic functionality, but are light and efficient.
- Query Builders – smaller tools or libraries, which in the common scenario provide an API for generating queries using methods or better, chained methods.

Each of the mentioned solutions has its advantages and disadvantages, and before making a choice, the needs of the developed system should be carefully analyzed. This analysis should consider the complexity of the queries that will be executed and the volume of data that will be processed. Another key factor is the developers themselves or, more specifically, the professional experience of the team with SQL.

### 3. ORM frameworks

Object-Relational Mapping (ORM) is a well-known technique in programming that provides an abstract layer over the database. This allows software developers to work with objects in the program code instead of using SQL. It is used to bridge object-oriented programming languages and relational databases, making the work of programmers easier and more optimized.

The role of the ORM layer is to automatically generate queries that correspond to basic data operations such as SELECT, INSERT, UPDATE, and DELETE. A relational database consists of tables with columns and rows, while object-oriented program code works with

objects defined by their properties and methods, and instances of these objects. To enable query generation, the ORM platform must be configured to be aware of the tables, their columns, and the relationships between them. This involves what is commonly referred to as mapping, or the correspondence between objects in the program code and tables in the database.

The main principles and ideas behind ORM can be defined as follows:

- Representation of the data structure from relational databases as structures of classes in the program code.
- Mechanism for establishing mappings between the classes in the program code and their properties with the tables in the database and their columns. In some platforms, this happens through configuration, while in others, it occurs automatically based on the names of the elements.
- Mechanism for the automatic management of relationships between objects, corresponding to the relationships between tables in the database.
- Mechanism for continuous monitoring of the correspondence between the structure of objects in the program code and the structure of the database.

When executing commands towards the database, the individual tools comprising the ORM platform take care of converting objects from classes into rows in tables and vice versa. Popular ORM solutions include Entity Framework (for .NET), Hibernate (for Java), SQL Alchemy (for Python), and Active Record (for Ruby). They, one might say, provide the highest level of abstraction and automation for working with data, completely isolating the developer from the SQL syntax of the executed queries.

As an example of such solutions, Entity Framework can be examined. It represents a full-fledged ORM tool embedded in the .NET Framework and .NET Core. To leverage EF in the developed system, it is necessary to define classes with the structure of the database tables and configure a context class, including corresponding sets for the tables. Applicable for all types of software development – web, mobile, desktop, etc. [5, 6, 8].

In the illustration in Figure 2Fig. 2., an example is given of a defined class User, which is equivalent to a table with the same name, Users, in the database. After creating the model and configuration in the business logic, objects of this class are used, and actions with them are transferred independently to the database without additional intervention [1].

Among the series of advantages of ORM that facilitate software development and interaction with relational databases, the following will be emphasized:

- Use of object-oriented code - facilitates and improves code readability.
- Correspondence between the structure of objects in the program and the structure of the database - facilitates data storage and retrieval.
- Absence of SQL code - allows a focus on business logic rather than database details.
- Included tools for migration and management of the database schema - facilitates maintenance and updating of the database structure.
- Reuse of program code - accelerates the development process and reduces code duplication.
- Built-in transaction management - facilitates data storage and updates within a single transaction.
- Abstraction and platform independence - facilitates portability between different relational databases.

- Testing - the possibility of using data simulations, which facilitates testing individual units of business logic using database queries.
- Security - can provide built-in protection against SQL injections and other security threats.

```
public class User
{
    public int Id { get; set; }
    public string FirstName { get; set; }
    public string LastName { get; set; }
    public string Email { get; set; }
    public string MobileNumber { get; set; }
}

public class ApplicationDbContext : DbContext
{
    public DbSet<User> Users { get; set; }

    protected override void OnConfiguring(DbContextOptionsBuilder
optionsBuilder)
    {
        optionsBuilder.UseSqlServer("ConnectionString");
    }
}
```

Fig. 2. Example of EF class and mapping configuration

Despite the numerous advantages, it is still important to consider the drawbacks of this approach. They are mainly in performance and cases with more specific needs that a universal platform usually does not strive to cover. As more important drawbacks, the following can be noted:

- Complexity - building correspondences between classes and tables and additional settings and configurations require additional code and time, which in some simpler software developments are seen as a burden.
- Performance - the abstraction inherent in this approach is undoubtedly convenient but at the same time leads to a loss of control over SQL code, which, in practice, is executed, and the possibility to optimize it when needed. If the generated queries are complex and inefficient, this leads to unnecessary load on the database and slower execution.
- Limited capabilities - encountering the limited capabilities of such a platform is inevitable when specific queries or formulas in SQL code need to be executed. In this category of complex scenarios, that are not covered, inheritance and polymorphism are also involved.

When choosing a solution for generating SQL queries, before selecting ORM as an approach, all these drawbacks should be considered and balanced with its advantages. In some cases, the choice between ORM and direct use of SQL may depend on the specific requirements of the project and the preferences of the team.

## 4. Micro ORM solutions

A micro ORM is an ORM framework but in its lightweight and minimalistic form. It's designed for simplicity and performance. Unlike full-fledged ORM frameworks, which often come with a wide range of features and abstractions, micro ORMs aim to provide a more focused and streamlined solution. As their name implies, they are smaller in size and number of features, and they have better performance, which is crucial when developing serverless solutions of any kind [7].

Micro ORM is a type of ORM platform, but it takes the form of a lighter and more minimalist layer compared to them. Essentially, these types of solutions are still abstractions over the database but are a bit closer to it. This allows them to address the needs for lightweights and higher performance in a software system. These tools facilitate interactions with databases by providing a balance between convenience and performance in handling data through objects in the program code.

Micro ORM is based on several key ideas and principles:

- Unnecessary abstractions and complex functionality are avoided, aiming to provide only essential functionality for mapping objects to the database.
- It relies on simplified configuration, aiming for easy integration with existing software systems.
- The focus is on performance, and therefore, developers are provided with access to the generated SQL code for control and optimization.
- By design, these tools are easy and intuitive to use. They are characterized by short and understandable code.
- Direct SQL Access: Micro ORMs may expose more direct access to SQL queries, allowing developers to write and execute raw SQL when needed. This can be beneficial in situations where fine-tuned control over the queries is required.

Most popular examples of micro ORMs include Dapper and NPoco, both for .NET.

Dapper is a popular micro ORM for .NET. It is known for its exceptional performance and simplicity. Dapper focuses on mapping database query results directly to object instances without introducing much abstraction.

This is a tool developed by people in a forum who created and expanded it according to their needs. Perhaps because of this, it is one of the most popular and widely used today. It uses ADO.NET to execute queries, providing more direct access to the database.

```
public class UserRepository
{
    private readonly IDbConnection _dbConnection;

    public UserRepository(IDbConnection dbConnection)
    {
        _dbConnection = dbConnection;
    }

    public User GetUserById(int userId)
    {
        return _dbConnection.QueryFirstOrDefault<User>("SELECT * FROM Users WHERE Id = @Id", new { Id = userId });
    }
}
```

Fig. 3. Example of Dapper usage

In the illustrated example in Figure 3, a row is extracted from the database table, which is automatically transformed into an object of type User. Although a manually written SQL query is used for execution, the business logic operates with objects [2].

NPoco is another lightweight micro ORM for .NET developed to be fast. It supports POCO (Plain Old CLR Object) mappings and is designed for flexibility and performance.

Examining the impact of the core ideas in Micro ORM solutions, we can easily describe their advantages:

- Simplicity and lightweight nature
- Performance
- Flexibility
- Easy integration

As a drawback, it should be noted that executing manually written queries, along with the freedom to use any specific functions and algorithms, control over performance, etc., inevitably leads to assumptions for errors. This is because the manually written SQL code does not compile and update automatically with changes in the database structure. The factor of human error will also have its impact.

While micro ORMs may not offer the extensive feature set of larger ORM frameworks, they can be valuable in situations where simplicity, performance, and direct control over database interactions are prioritized. Developers often choose micro ORMs for projects with specific requirements or when the overhead of a full ORM is deemed unnecessary.

## 5. Query builders

Query Builders are tools or libraries, which can be defined as an easy way to generate queries with methods called with program code. In the context of the architecture of a software system, they are used to achieve code readability, modularity, and easy maintenance.

In essence, Query builders provide an API that, by invoking various methods or a chain of methods, can generate queries ranging from the simplest to the most complex and intricate ones [3].

```
var books = db.Query("Books").Get();  
var books = db.Query("Books").WhereTrue("IsPublished").Get();  
var introToSql = db.Query("Books").Where("Id", 145).Where("Lang", "en").First();
```

Fig. 4. Method chain example

Query builders provide a programmatic and fluent interface for constructing SQL queries, offering a more intuitive and less error-prone approach compared to manually writing raw SQL statements. Here are the key aspects of how query builders achieve this:

- **Chaining Methods** - Query builders allow developers to chain methods together in a fluid manner. Each method typically corresponds to a specific SQL operation or clause, such as SELECT, FROM, WHERE, JOIN, etc. Chaining methods results in a readable and sequential representation of the SQL query, mirroring the natural flow of constructing a query.
- **Method Naming Conventions** - Query builder methods are often named in a way that reflects their SQL counterparts. This makes it easier for developers to understand the purpose of each method and its role in the query construction process. For example, a



method named `select` is likely used to specify the columns to be retrieved, while a method named `where` is used for adding conditions.

- **Fluent Syntax** - The fluent syntax of query builders enables developers to write code that closely resembles the structure of the desired SQL query. This results in more readable and maintainable code. Developers can build complex queries step by step, adding clauses and conditions in a natural and sequential order.

- **Type-Safe Operations** - Many query builders are designed to be type-safe, which means that the methods and parameters used in the construction of a query are validated by the compiler. This helps catch errors at compile-time rather than runtime. Type safety reduces the likelihood of mistakes, such as using incorrect column names or mismatching data types.

- **Abstraction of SQL Syntax** - Query builders abstract the underlying SQL syntax, providing a higher-level interface. This abstraction shields developers from the intricacies of SQL syntax, making the process more approachable, especially for those who may not be SQL experts. Developers can focus on expressing their intent in a programming language, leaving the query builder to translate that intent into valid SQL.

- **Dynamic Query Building** - Query builders often support dynamic query building, allowing developers to conditionally add or remove clauses based on runtime conditions. This flexibility is useful when dealing with variable search criteria or constructing queries with optional components.

- **Prevention of SQL Injection** - Since query builders use parameterized queries or prepared statements internally, they inherently protect against SQL injection attacks. Parameters are automatically sanitized, reducing the risk associated with user-input data.

In summary, query builders simplify the process of constructing SQL queries by providing a programmatic, readable, and type-safe interface. This approach enhances code quality, reduces errors, and makes it easier for developers to express their intent in a way that closely aligns with the structure of the desired SQL queries.

## 6. Conclusion

A modern software system cannot afford to be without a solution for generating SQL queries. Regardless of the chosen platform, library, or tool, isolating this layer leads to undeniable advantages for both the developers of the system and the software product itself. Each of them has its advantages and disadvantages, which can guide towards the right choice for the specific business, the specific team, and the needs related to performance, security, and complexity.

## Acknowledgement

The research activity of the authors is partially supported by National Science Fund of Bulgaria under the Grant № KP-06-H55/5 from 15.11.2021, entitled “Smart Risk Management for Business from Adverse Events and Natural Disasters”.

## References

1. Microsoft (2023). Getting Started with EF Core. *Cloud Skills Challenge*, 24.08.2023, <https://learn.microsoft.com/en-us/ef/core/get-started/overview/first-app?tabs=netcore-cli#create-the-model>
2. GitHub (2023, November). Dapper, DapperLib, <https://github.com/DapperLib/Dapper>

3. SqlKata on Github (2023, November). A fluent SQL query builder for C#, Build dynamic SQL queries, with confidence, <https://sqlkata.com/>
4. Lazarova V. (2019) Digitalization and Digital Transformation in Accounting. *Ikonomiceski i Sotsialni Alternativi*, issue 2, pp. 97-106, UNWE, Sofia, Bulgaria.
5. Milev P. (2021) Approach for Analysis and Comparison of Search Query Results in Web Publications. In: 11<sup>th</sup> *International Conference on Application of Information and Communication Technology and Statistics in Economy and Education (ICAICTSEE-2021)*, 25-26 November 2021, pp. 332-338, UNWE, Sofia, Bulgaria, ISSN 2367-7643.
6. Radoev M. (2021). Practical Aspects of Using HDFS Data in SQL Server 2019. In: 11<sup>th</sup> *International Conference on Application of Information and Communication Technology and Statistics in Economy and Education (ICAICTSEE-2021)*, 25-26 November 2021, pp. 363-367, UNWE, Sofia, Bulgaria, ISSN 2367-7643.
7. Stefanov G. (2021). Analysis of Serverless Cloud Data Warehouse Solutions. In: 11<sup>th</sup> *International Conference on Application of Information and Communication Technology and Statistics in Economy and Education (ICAICTSEE-2021)*, 25-26 November 2021, pp. 429-434, UNWE, Sofia, Bulgaria, ISSN 2367-7643.
8. Tsaneva, M. (2021). Applicability of ASP.Net Frameworks for Developing Web-based Business Information Systems. In: 11<sup>th</sup> *International Conference on Application of Information and Communication Technology and Statistics in Economy and Education (ICAICTSEE-2021)*, 25-26 November 2021, pp. 319-326, UNWE, Sofia, Bulgaria, ISSN 2367-7643.

# Using Code Generation in LLMs for Automatic Execution of Data Science Tasks

Venko Andonov

University of National and World Economy, Sofia, Bulgaria  
*vandonov@unwe.bg*

**Abstract.** Typically, most of the data science tasks start with data cleaning, preprocessing and feature engineering. The specifics of each one depends on the data itself as well as the final goal, and that is why it is not a straightforward process and requires human expertise and domain knowledge. The state of the art large language models (LLMs) have the capability to understand complex problems and generate programming code. In this paper, these capabilities are evaluated for proprietary and open source models and interpretation environments, for both public and private datasets in different domains. The results show that this is a feasible approach, but it requires the guidance of a human, especially for tasks that require specific knowledge about the data, its context and interpretation.

**Keywords:** Data science, LLM, Data processing

## 1. Introduction

Data science is a multifaceted discipline that encompasses a wide range of tasks, with data cleaning, preprocessing, and feature engineering forming the critical foundation of most projects. These initial steps are crucial for ensuring data quality and optimizing the performance of subsequent analytical and predictive models. Data cleaning involves identifying and rectifying errors, inconsistencies, and missing values in datasets [1]. Preprocessing encompasses various techniques such as normalization, scaling, and encoding to transform raw data into a format suitable for analysis. Feature engineering, a key component of the data preparation process, involves creating new variables or transforming existing ones to better represent the underlying patterns and relationships within the data.

The complexity and variability of these tasks are largely dependent on the nature of the data and the specific objectives of the analysis. Traditionally, these processes have heavily relied on human expertise and domain knowledge. Data scientists and analysts play a crucial role in interpreting the context of the data, identifying relevant patterns, and making informed decisions about appropriate data manipulations and transformations.

The automation of data science tasks has been a long-standing goal in the field, driven by the need to enhance efficiency, reduce manual effort, and minimize human error. However, the inherent complexity and diversity of real-world datasets have posed significant challenges to traditional automation approaches. Rule-based systems and conventional machine learning techniques often struggle to adapt to the nuanced requirements of different datasets and analytical goals.

The recent advancements in large language models (LLMs) have opened up new possibilities for automating various aspects of data science workflows. LLMs, such as GPT-4 and its counterparts, have demonstrated remarkable capabilities in understanding complex problems and generating high-quality programming code [2]. These models offer the potential

to revolutionize the way data scientists approach routine and repetitive tasks, potentially freeing up valuable time for more complex analytical work and strategic decision-making.

The primary objective of this paper is to evaluate the feasibility and effectiveness of using code generation capabilities of LLMs for the automatic execution of data science tasks. Specifically, this study aims to: assess and compare the capabilities of both proprietary and open-source LLMs in generating code for data cleaning, preprocessing, and feature engineering tasks; evaluate the performance of these models across a diverse range of datasets, including both public and private sources, spanning multiple domains; investigate the extent to which human guidance and intervention are necessary to ensure the accuracy, relevance, and contextual appropriateness of the generated code; identify and analyze the challenges, limitations, and potential pitfalls associated with using LLMs for automating data science tasks; explore the implications of this approach on the role of data scientists and the future of data science workflows.

## 2. Literature

The automation of data science tasks has been a long-standing goal in the field, leading to the development of various tools and frameworks designed to streamline the data preparation and analysis process. Traditional automation tools such as KNIME, Alteryx, and RapidMiner have been at the forefront of simplifying data preparation tasks. These platforms provide user-friendly interfaces that allow users to perform complex data manipulations without extensive programming knowledge. For example, KNIME offers a visual workflow designer that enables users to create data pipelines through drag-and-drop functionality; Alteryx specializes in data blending and advanced analytics, with a focus on geospatial analysis and predictive modeling; RapidMiner provides a comprehensive suite of data science tools, including automated feature engineering and model deployment capabilities [17]. More recent additions to this space include Trifacta and Dataiku, which have gained popularity for their collaborative features and integration with cloud platforms.

The powerful Large Language Models (LLMs) opened new avenues for automating various aspects of data science workflows, particularly in code generation. Models such as GPT-4, Claude, Gemini, and open-source alternatives like Llama, Mistral, and code-focused ones like StarCoder and CodeLlama have demonstrated impressive capabilities in understanding natural language and generating contextually relevant programming code [2]. These models are trained on large datasets that include multiple programming languages and problem domains, enabling them to address a wide range of coding tasks, including data manipulation, analysis, and visualization.

In the field of software engineering, LLMs have shown promise in various tasks: Code Documentation - Procko & Collins demonstrated the effectiveness of LLMs in automating code documentation, potentially saving developers significant time and improving code maintainability [10]; Bug Fixing - work by Sobania et al., and Fan et al., have explored the use of LLMs for automated bug detection and repair, showing promising results in identifying and correcting common coding errors [6] [18]; Code Completion - tools like GitHub Copilot, powered by OpenAI's GPT models, have gained widespread adoption for their ability to suggest code completions and generate entire functions based on natural language descriptions.

McInerney et al. explored the use of LLMs for generating data preprocessing scripts, finding that these models can significantly reduce the time required for initial data preparation [9]. Their work demonstrated that LLMs could generate code for tasks such as handling missing values, encoding categorical variables, and scaling numerical features. Xu et al. assessed the ability of LLMs to perform feature extraction and transformation. Their study

concluded that while LLMs can generate useful code snippets for feature engineering, the quality and accuracy of the output depend heavily on the clarity and detail of the input prompt [7].

Recent work by Jansen et al. has shown that LLMs can be used to automate various aspects of data analysis, including exploratory data analysis, statistical testing, and basic model building [8]. Their study demonstrated that LLMs could generate code to produce descriptive statistics, create visualizations, and perform simple hypothesis tests based on natural language descriptions of the data and analysis goals.

Despite the promising results, several studies have highlighted challenges in using LLMs for data science automation: Context Understanding - LLMs sometimes struggle to fully grasp the context of complex data science problems, leading to generated code that may not fully address the specific requirements of a task [3]; Data Privacy Concerns - the use of LLMs for data science tasks raises important questions about data privacy, especially when dealing with sensitive or proprietary datasets [15]; Reproducibility - issues with reproducibility in LLM-generated code, noting that different runs of the same model can produce varying results, which can be problematic in scientific and business contexts [13].

Recent developments in the field include domain-specific LLMs - there's a growing trend towards training LLMs on domain-specific datasets to improve their performance on specialized data science tasks; for example, Gu et al. introduced a model fine-tuned specifically for bioinformatics data analysis [4]; interactive code generation - new approaches are exploring more interactive ways of using LLMs for code generation, where the model engages in a dialogue with the user to refine and improve the generated code iteratively [11]; hybrid approaches - combining traditional automation tools with LLM-based code generation is an emerging area of research, as exemplified by the work of Yao et al., who proposed a framework that integrates AutoML techniques with LLM-generated feature engineering code [5].

### 3. Methodology

In this study, we leverage LangChain to develop a robust data processing pipeline. Our evaluation focuses on the capabilities of two state-of-the-art large language models (LLMs): the proprietary GPT-4-turbo API and the open-source Llama 2 70b model accessed via the Groq API. All code generation and model interactions are conducted remotely through these APIs, with local processing restricted to interpreting the generated code and executing the data processing tasks.

The selection of LLMs is based on performance, accessibility, and diversity. GPT-4-turbo, provided by OpenAI, is known for its advanced natural language understanding and superior code generation capabilities. On the open-source front, Llama 2 70b is selected for its competitive performance and transparency, providing a valuable comparison to proprietary models.

We utilize two datasets to test the models' versatility and effectiveness across different domains. The private dataset consists of financial data, encompassing various transactional and market variables.

The public dataset is sourced from the popular data science platform Kaggle – “House Prices - Advanced Regression Techniques” (<https://www.kaggle.com/competitions/house-prices-advanced-regression-techniques/data>).

The evaluation procedure begins with the definition of specific data science tasks, including data cleaning, preprocessing, and feature engineering. Detailed natural language prompts are crafted to describe these tasks, which are then inputted into the LLMs through LangChain.

Once the models generate the corresponding code snippets, these are executed on the respective datasets. The outputs are validated against expected results to assess accuracy. Performance metrics such as accuracy, computational efficiency, and the extent of human intervention required are recorded and analyzed.

For data cleaning and preprocessing, tasks include handling missing values, data normalization, and categorical encoding.

Feature engineering tasks involve creating new features based on domain knowledge and applying mathematical transformations to existing features. The relevance and quality of the generated features are assessed to determine their utility for predictive tasks.

Performance is measured using several metrics. Accuracy assesses how well the generated code performs the intended tasks. Efficiency is evaluated based on runtime and computational resource usage. Human effort is quantified by the time and effort required to correct or refine the generated code.

## 4. Results

Both GPT-4-turbo and Llama 2 70b are tasked with generating code to handle missing values in both datasets. The prompts specified various methods, such as mean imputation, median imputation, and mode imputation for numerical columns, and filling with the most frequent value for categorical columns.

An example prompt is: “Generate Python code to fill missing values in the 'LotFrontage' column of a pandas DataFrame using the mean value of the column.” (Fig. 1, Fig. 2).

```
import pandas as pd

# Sample DataFrame
data = {
    'LotFrontage': [80, 70, None, 60, None, 90]
}
df = pd.DataFrame(data)

# Fill missing values in 'LotFrontage' with the mean value of the column
mean_lot_frontage = df['LotFrontage'].mean()
df['LotFrontage'].fillna(mean_lot_frontage, inplace=True)

print(df)
```

Fig. 1. GPT-4's output of one of the queries

The code is then extracted from the response (it is typically surrounded by three backticks). We assume that our data is loaded in a dataframe with the name “df”, but in some of the responses this is not handled properly and needs further modification of the generated code.

The following are examples of relatively simple transformations of the data frame that have been requested:

Generate Python code to fill missing values in the 'Alley' column of a pandas DataFrame using the mode value of the column.

Data Normalization - Min-Max Scaling - Generate Python code to normalize the 'LotArea' column of a pandas DataFrame using Min-Max Scaling.

Categorical Encoding - One-Hot Encoding - Generate Python code to apply one-hot encoding to the 'Neighborhood' column of a pandas DataFrame.

```
Here is the Python code to fill missing values in the 'LotFrontage'
'''
import pandas as pd
import numpy as np

# assume 'df' is your pandas DataFrame

# fill missing values in 'LotFrontage' column with mean value
df['LotFrontage'].fillna(df['LotFrontage'].mean(), inplace=True)
'''

Here's how the code works:

1. `df['LotFrontage']` selects the 'LotFrontage' column of the Da
2. `.fillna()` is a pandas method that replaces missing values (N
3. `df['LotFrontage'].mean()` calculates the mean value of the 'L
4. `inplace=True` modifies the original DataFrame by filling the

After running this code, the missing values in the 'LotFrontage'

Note: If you want to fill missing values with a specific value in
`df['LotFrontage'].fillna(0, inplace=True)` would fill missing va
```

Fig. 2. Llama's output for the same query

Label Encoding - Generate Python code to apply label encoding to the 'MSZoning' column of a pandas DataFrame

Feature Creation - Creating a New Feature - Generate Python code to create a new feature 'TotalArea' by summing the '1stFlrSF', '2ndFlrSF', and 'TotalBsmtSF' columns of a pandas DataFrame.

Feature Transformation - Log Transformation - Generate Python code to apply a logarithmic transformation to the 'SalePrice' column of a pandas DataFrame.

The following are more complex examples and require deeper understanding of the underlying data and the final goal:

Domain-Specific Missing Value Imputation - Generate Python code to fill missing values in the 'GarageYrBlt' column by predicting the year based on the overall quality of the house and the neighborhood characteristics

Custom Outlier Detection and Treatment - Generate Python code to identify and treat outliers in the 'SalePrice' column based on a custom threshold that considers the median income of the neighborhood and the year the house was built.

Creating Interaction Terms Based on Domain Knowledge - Generate Python code to create interaction terms between 'OverallQual', 'GrLivArea', and 'Neighborhood' that capture the combined effect of these features on 'SalePrice', based on real estate market trends.

Temporal Feature Creation - Generate Python code to create a feature that measures the age of the house at the time of sale, and another feature that measures the average annual price appreciation rate based on historical sales data.

Additionally, similar queries have been submitted for the private financial dataset.

Overall, the results indicate that using LLMs for automating data science tasks is a feasible approach, particularly with the guidance of human experts to ensure accuracy and context-specific adjustments. It is still necessary to create quite specific prompts that require knowledge of the dataset and at what the end goal should be. These findings highlight the potential of LLMs to enhance productivity in data science workflows while acknowledging the need for ongoing human oversight.

Table 1. Experimental results for the queries, grouped by category, datasets and used LLM. The percentages show the success rate, where it is determined by successful execution of the Python script and correct (expected) output in the DataFrame.

Task Category	Dataset #1		Dataset #2	
	<i>GPT-4 Turbo</i>	<i>Llama 2 70b</i>	<i>GPT-4 Turbo</i>	<i>Llama 2 70b</i>
Handling missing values	5/5	5/5	5/5	5/5
Data normalization	4/4	3/4	5/5	3/5
Categorical encoding	2/2	2/2	2/2	2/2
Feature engineering	4/4	2/4	3/4	2/4
Feature transformation	4/4	3/4	3/3	1/3
Complex data preprocessing	2/3	1/3	3/3	2/3
Complex feature engineering with domain knowledge	3/4	2/4	4/4	2/4
Complex transformations	2/4	0/4	2/3	2/3

There is no significant difference between the used large language models; as expected the state-of-the-art GPT-4 Turbo has an advantage on the more complex queries. Further experiments with different models, different prompts, fine-tuned models and few-shot examples in the prompt are necessary to get a better understanding of the capabilities of this approach.

## 5. Conclusion

This study explored the potential of using large language models (LLMs) for automating various data science tasks, focusing on data cleaning, preprocessing, and feature engineering. By evaluating the performance of both proprietary (GPT-4-turbo) and open-source (Llama 2 70b) models across diverse datasets, we have demonstrated the feasibility and challenges of this approach.

Our results indicate that LLMs can successfully generate code for a wide range of data science tasks, from basic operations like handling missing values to more complex feature engineering requiring domain knowledge. The success rates varied across task categories and model types, with GPT-4-turbo generally outperforming Llama 2 70b, particularly on more complex queries.

The quality of generated code depends heavily on the specificity and clarity of the prompts, underscoring the continued importance of human expertise in framing data science problems. While LLMs show great promise in automating routine tasks, they are best viewed as tools to augment human expertise rather than replace it entirely. The role of data scientists may evolve to focus more on problem framing, interpretation, and strategic decision-making, with LLMs handling more of the routine coding tasks.

## References

1. Hadi, M. U., Qureshi, R., Shah, A., Irfan, M., Zafar, A., Shaikh, M. B., ... & Mirjalili, S. (2023). A survey on large language models: Applications, challenges, limitations, and practical usage. Authorea Preprints. <https://doi.org/10.2139/ssrn.4571236>



2. Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y. (2023). Sparks of artificial general intelligence: Early experiments with GPT-4. arXiv. <https://arxiv.org/abs/2303.12712>
3. Stringhi, E. (2023). Hallucinating (or poorly fed) LLMs? The problem of data accuracy. *i-lex*, 16(2), 54-63.
4. Gu, Y., Tinn, R., Cheng, H., Lucas, M., Usuyama, N., Liu, X., ... & Poon, H. (2021). Domain-specific language model pretraining for biomedical natural language processing. *ACM Transactions on Computing for Healthcare (HEALTH)*, 3(1), 1-23.
5. Yao, J., Xu, W., Lian, J., Wang, X., Yi, X., & Xie, X. (2023). Knowledge plugins: Enhancing large language models for domain-specific recommendations. arXiv preprint arXiv:2311.10779.
6. Fan, Z., Gao, X., Mirchev, M., Roychoudhury, A., & Tan, S. H. (2023, May). Automated repair of programs from large language models. In *2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE)* (pp. 1469-1481). IEEE.
7. Xu, F. F., Alon, U., Neubig, G., & Hellendoorn, V. J. (2022, June). A systematic evaluation of large language models of code. In *Proceedings of the 6th ACM SIGPLAN International Symposium on Machine Programming* (pp. 1-10).
8. Jansen, J. A., Manukyan, A., Al Khoury, N., & Akalin, A. (2023). Leveraging large language models for data analysis automation. *bioRxiv*. <https://doi.org/10.1101/2023.12>
9. McInerney, D. J., Young, G., van de Meent, J. W., & Wallace, B. C. (2023). Chill: Zero-shot custom interpretable feature extraction from clinical notes with large language models. arXiv preprint arXiv:2302.12343.
10. Procko, T., & Collins, S. (2023). Automatic code documentation with syntax trees and GPT: Alleviating software development's most redundant task. SSRN. <https://doi.org/10.2139/ssrn.4571367>
11. Ma, P., Ding, R., Wang, S., Han, S., & Zhang, D. (2023, December). InsightPilot: An LLM-empowered automated data exploration system. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: System Demonstrations* (pp. 346-352).
12. Liu, J., Wang, J., Huang, H., Zhang, R., Yang, M., & Zhao, T. (2023, October). Improving LLM-based health information extraction with in-context learning. In *China Health Information Processing Conference* (pp. 49-59). Springer Nature Singapore.
13. Kovacheva, M. (2023). Creation of an automated tool for evaluation and selection of NoSQL databases intended for financial services based on predefined criteria. *Innovations*, 11(1), 10-11.
14. Velkova, I. (2023). Unstructured social media data processing with artificial intelligence. *Industry 4.0*, 8(2), 65-67.
15. Zlateva, P., Steshina, L., Petukhov, I., & Velez, D. (2024). A conceptual framework for solving ethical issues in generative artificial intelligence. In *Electronics, Communications and Networks* (pp. 110-119). IOS Press.
16. Ni, A., Iyer, S., Radev, D., Stoyanov, V., Yih, W. T., Wang, S., & Lin, X. V. (2023, July). Lever: Learning to verify language-to-code generation with execution. In *International Conference on Machine Learning* (pp. 26106-26128). PMLR.
17. Milev, P. (2019). Conceptual approach for application of data science in web analytics. In *Proceedings of International Conference on Application of Information and Communication Technology and Statistics in Economy and Education (ICAICTSEE)* (pp. 223-228). ICAICTSEE.
18. Sobania, D., Briesch, M., Hanna, C., & Petke, J. (2023). An analysis of the automatic bug fixing performance of ChatGPT. In *2023 IEEE/ACM International Workshop on Automated Program Repair (APR)* (pp. 23-30). IEEE. <https://doi.org/10.1109/APR59189.2023.00012>

# RAG Strategies for a Hallucination-Free University Admission AI Assistant

Venko Andonov

University of National and World Economy, Sofia, Bulgaria  
*vandonov@unwe.bg*

**Abstract.** The development of a LLM-based university admission assistant requires the use of up-to-date and factually correct information. The state of the art proprietary and open source models are prone to hallucination (generating false information because of a lack of specific real information in their pretraining phase), which might cause confusion for the prospective students. A set of retrieval augmented generation (RAG) strategies on different models are evaluated, that ensure that the models output up-to-date and factually correct information and are also able to use personalized data to help the prospective students with the application process. We are using a specific case for UNWE, Sofia, which also adds an additional requirement for the support of Bulgarian language in the conversation.

**Keywords:** Retrieval augmented generation, University AI

## 1. Introduction

The advancement of large language models (LLMs) like GPT-4 Turbo and Llama 3 70b has significantly impacted various domains, including education. Universities worldwide are leveraging these models to develop AI-based assistants that streamline the admission process. These AI assistants aim to provide prospective students with accurate, personalized, and up-to-date information regarding admission requirements, application deadlines, program details, and other related queries.

However, a major challenge with LLMs is their propensity for hallucination, wherein the models generate false or misleading information due to gaps or biases in their pretraining data. Hallucinations can lead to confusion among prospective students, potentially impacting their decision-making process and overall experience. Therefore, ensuring the accuracy and reliability of the information provided by these AI assistants is crucial.

The primary objective of this research is to evaluate various Retrieval-Augmented Generation (RAG) strategies to minimize hallucinations in LLM-based university admission assistants. By integrating RAG strategies, the goal is to ensure that the AI models output factually correct and contextually relevant information. Additionally, the research focuses on the ability of these models to utilize personalized data to assist prospective students with their specific needs and queries during the application process.

This study centers on the University of National and World Economy (UNWE) in Sofia, Bulgaria. As a leading institution in the region, UNWE attracts a diverse pool of prospective students, necessitating an AI assistant that can handle a wide range of inquiries. The inclusion of UNWE as a case study adds an additional layer of complexity due to the requirement of supporting the Bulgarian language in the conversational interface.

This research evaluates the performance of both proprietary (GPT-4 Turbo) and open-source (Llama 3 70b) models when augmented with RAG strategies. The comparison aims to identify the most effective approach for delivering accurate and personalized information in

the context of university admissions. The study also examines the models' ability to handle Bulgarian language support, ensuring that the AI assistant can cater to local students efficiently.

## 2. Literature Review

The digitization of educational processes, particularly in career development and university admissions, has garnered significant attention in recent research. The integration of technological solutions to enhance these processes is crucial for improving student experiences and institutional efficiency. This literature review explores the latest advancements in digitizing career development and university admissions, emphasizing the role of AI and machine learning.

One of the main potential advantages of the technological solutions is in digitizing career development processes in higher education [1]. The study by Kirilov highlights the importance of effective career orientation and proposes a conceptual model to enhance student realization through digitization. By establishing a causal relationship between career development and education quality, the study underscores the necessity of integrating technological advancements to support student career guidance.

Kazakova extends this discussion to secondary education, examining the impact of digital learning methods as part of Bulgaria's Strategic Framework for Education, Training, and Learning. The research identifies key challenges and proposes solutions, demonstrating how digital technologies can significantly improve teaching methods and overall education quality [2].

Boyanov et al. focus on the security aspects of digital educational activities. Their study proposes a two-factor authentication approach using face recognition to prevent identity fraud during digital exams. This method ensures the authenticity of participants, thereby enhancing the integrity of university admission processes and other educational activities [3].

The application of AI and machine learning in university admissions has been explored in various contexts. Katti et al. investigate the use of Google Vertex AI for predicting university admissions. By leveraging machine learning techniques, their study aims to enhance the admission process by accurately forecasting student admissions based on relevant data. This approach can potentially streamline the admissions process, making it more efficient and data-driven [4].

Le Hoanh Su et al. introduce an AI chatbot designed to support university admissions and career guidance. The chatbot operates around the clock, processing student and parent requests using natural language processing and machine learning. This development represents a significant advancement in university admissions consulting, providing timely and personalized assistance to prospective students [5].

Similarly, Alothman et al. examine the acceleration of university admission systems through machine learning techniques. Their research focuses on analyzing historical data to optimize decision-making algorithms, aiming to enhance efficiency and reduce processing time. This study highlights the potential of machine learning to streamline administrative processes in higher education [6].

Several studies have focused specifically on the development and application of AI chatbots in university admissions. Nguyen et al. present a chatbot designed to support university admissions by answering common questions, providing information about admission requirements, and guiding applicants through the application process. This chatbot aims to reduce the workload on administrative staff and improve the applicant experience [9].

Prabha et al. introduce a student-focused chatbot powered by artificial intelligence. This chatbot assists prospective students during the admission process, addressing queries related

to eligibility criteria, deadlines, and required documents. The study emphasizes the role of AI in enhancing the accessibility and responsiveness of university admission services [10].

Sarker, Rahman, and Siam develop an Anglo-Bangla language-based AI chatbot for the Bangladeshi university admission system. This chatbot operates in both English and Bengali, providing personalized guidance to prospective students. The bilingual capability of the chatbot makes it particularly valuable in multilingual contexts, ensuring that language barriers do not hinder access to information [8].

Hamzah et al. explore the utilization of AI in developing a chatbot for new student admission support at Cokroaminoto University Palopo. The chatbot provides information about programs, admission requirements, and deadlines, assisting applicants throughout the admission process. This study highlights the broader applicability of AI chatbots in various educational contexts [11].

Lira et al. take a different approach by exploring the use of artificial intelligence to assess personal qualities in college admissions. Their study analyzes essays, recommendation letters, and other qualitative data to provide a holistic view of applicants. This research demonstrates the potential of AI to move beyond quantitative metrics and incorporate qualitative assessments in the admissions process, offering a more comprehensive evaluation of prospective students [7].

### 3. Methodology

To ensure the accuracy and relevance of the information provided by the university admission assistant, data was collected from multiple authoritative sources:

a) **Primary Source:** The official website of the University of National and World Economy (UNWE), Sofia, provided comprehensive and up-to-date information on admission requirements, deadlines, program descriptions, faculty information, and frequently asked questions.

b) **Supplementary Sources:** Additional data was gathered through direct communication with the admissions office and other relevant university departments.

c) **User-Specific Data:** Customized information was compiled based on individual queries from prospective students, encompassing previous interactions, preferred academic programs, academic backgrounds, and specific personal preferences.

To maintain data integrity and ensure compatibility with the AI models, the following steps were undertaken:

a) **Data Cleaning:** The collected data underwent meticulous review and correction for errors or inconsistencies. This process involved standardizing formats, rectifying discrepancies, and cross-verifying information against official sources to ensure accuracy.

b) **Data Structuring:** The cleaned data was organized into structured formats suitable for efficient retrieval and integration with the Large Language Models (LLMs).

Two state-of-the-art language models were selected for this study:

a) **GPT-4 Turbo:** This proprietary model developed by OpenAI was chosen for its advanced language processing capabilities. Retrieval mechanisms were integrated into GPT-4 Turbo to enhance response accuracy by grounding outputs in real-time, relevant data.

b) **Llama 3 70B:** This open-source model, known for its flexibility and customization potential, was tailored to handle admission-related queries and integrated with retrieval-augmented techniques to leverage external data sources.

A comprehensive evaluation of various RAG strategies was conducted, focusing on their methodologies and theoretical underpinnings. The strategies assessed include:

a) Traditional Search-Based Retrieval: This method uses specific keywords from user queries to search a database or document set. The retrieved information is then used to inform the model's responses, ensuring outputs are based on actual, up-to-date information rather than relying solely on pre-trained knowledge.

b) Context-Aware Retrieval: This advanced method enhances traditional search by considering query context. It employs semantic search techniques to understand the broader meaning and intent behind user queries, retrieving the most relevant information even when exact keywords are absent. This approach leverages natural language processing to improve the relevance and accuracy of retrieved data.

c) Knowledge-Augmented Techniques: These techniques integrate structured knowledge bases, such as knowledge graphs or databases, with the LLMs. The AI assistant uses these structured sources to validate and supplement its responses, ensuring outputs are both accurate and contextually appropriate. This strategy is particularly effective for handling complex queries requiring specific and detailed information.

d) Hybrid Methods: This approach combines traditional search-based retrieval, context-aware retrieval, and knowledge-augmented techniques to maximize the accuracy and relevance of the provided information. The hybrid method leverages the strengths of each approach, ensuring the AI assistant can handle a wide range of queries with high accuracy.

For contextual information, the ChromaDB vector store database is used to store the chunked and preprocessed information that has been gathered. Also, standard SQL connections are used to retrieve data about the admission process and the prospective students.

## 4. Results

The integration of Retrieval-Augmented Generation (RAG) strategies with GPT-4 Turbo significantly enhanced the accuracy and relevance of the model's responses. The data-driven approach ensured that the information provided was grounded in real-time, authoritative sources. Similar improvements were observed with Llama 3 70b, though the gains were slightly less pronounced compared to GPT-4 Turbo.

The hallucination rate, defined as the frequency of incorrect or misleading information, decreased significantly with the introduction of RAG.

The implementation of RAG strategies greatly enhanced GPT-4 Turbo's ability to handle queries in Bulgarian.

GPT-4 Turbo outperformed Llama 3 70b in terms of accuracy, relevance, and reduced hallucination rates. This suggests that proprietary models may have an edge in certain applications due to more extensive training data and optimization.

Both models showed significant improvements in handling Bulgarian queries post-RAG, though GPT-4 Turbo maintained a slight advantage in overall performance metrics.

Как мога да кандидатствам за Маркетинг – дистанционна форма в УНСС?

За да кандидатствате за програмата „Маркетинг“ – дистанционна форма в Университета за национално и световно стопанство (УНСС), можете да следвате следните стъпки:

Проучване на изискванията за кандидатстване:

Посетете уебсайта на УНСС и разгледайте секцията за кандидатстване. Тук ще намерите информация за необходимите документи, сроковете за кандидатстване и другите изисквания. Подготовка на необходимите документи:

Обикновено се изискват следните документи:

Диплома за завършено средно или висше образование.

Копие от личната карта.

Мотивационно писмо.

CV (автобиография).

Други документи, които университетът може да изисква (например, сертификати за владение на чужди езици и т.н.).

Регистрация и подаване на документи:

Регистрирайте се в системата за онлайн кандидатстване на УНСС. Това може да се направи на уебсайта на университета.

Fig. 1. Direct response with hallucinations - it includes false statements that might cause problems for the prospective students (in Bulgarian, the question is: “How can I apply for Marketing is distance form at UNWE?”).

Как мога да кандидатствам за Маркетинг – дистанционна форма в УНСС?

За да кандидатствате за специалността "Маркетинг" в дистанционна форма на обучение в УНСС, трябва да следвате стандартния процес за прием. Според "Правилника за прием ОКС бакалавър 2023-2024", стъпките за кандидатстване включват:

Регистрация в онлайн системата: Всички кандидати трябва да се регистрират в електронната система за кандидатстване на УНСС. Подаване на документи: Необходимите документи включват диплома за завършено средно образование и други специфични документи, които могат да се изискват за конкретната програма. Изпити или приемни изпити: Възможно е да се изисква полагане на приемни изпити или предоставяне на резултати от национални изпити.

Такси за кандидатстване: Заплащането на такса за участие в кандидатстудентската кампания.

За конкретна информация относно сроковете и изискванията за специалността "Маркетинг" в дистанционна форма, е препоръчително да се консултирате с официалния сайт на УНСС или да се свържете с приемната комисия на университета.

Fig. 2. Response after applying a RAG strategy - it return correct, factual information.

Каква е таксата за кандидатстване?

Съжалявам, не успях да намеря точна информация за таксата за кандидатстване за учебната година 2023–2024 в предоставения документ. За да получите най-актуалната информация, препоръчвам да се свържете с приемната комисия на УНСС или да посетите официалния сайт на университета. Там обикновено се публикуват всички подробности за таксите и сроковете за кандидатстване.

Каква е таксата за кандидатстване?

Таксата за кандидатстване в УНСС обикновено е в диапазона между 30 и 50 лева за едно направление. За да получите точната сума, препоръчвам да се свържете директно с отдела за прием на студенти на УНСС или да се консултирате с актуалната информация, предоставена в официалните документи на университета, които сте получили или които са достъпни на уебсайта на УНСС.

Fig. 3. Comparison of responses for information that is not available at the given moment, about the application fee - applying the proposed strategy (above) returns an inconclusive response that doesn't contain false information and tells the prospective students to check and ask for more information. The direct response (below) contains false information that might mislead them.

The implementation of RAG strategies has proven to be highly effective in enhancing the performance of both proprietary and open-source models. By grounding responses in real-time, authoritative data, these strategies significantly reduce hallucinations, improve accuracy, and increase user satisfaction. The results highlight the potential of RAG to transform AI-assisted university admission processes, providing reliable and personalized support to prospective students.

## 5. Conclusion

This study has demonstrated the significant potential of Retrieval-Augmented Generation (RAG) strategies in developing a reliable and accurate AI-based university admission assistant. By integrating RAG techniques with state-of-the-art language models like GPT-4 Turbo and Llama 3 70B, we have successfully mitigated the problem of hallucinations, ensuring that prospective students receive factual and up-to-date information.

The implementation of RAG strategies resulted in a marked decrease in hallucination rates and a substantial improvement in the accuracy and relevance of responses. This was particularly evident in the case of GPT-4 Turbo, which showed superior performance in handling queries, especially in the Bulgarian language context. The ability to provide accurate information in multiple languages is crucial for institutions like UNWE.

The success of this RAG-enhanced AI assistant in the context of UNWE's admission process points to broader applications in higher education. Such systems can significantly streamline administrative processes, reduce workload on staff, and provide round-the-clock support to prospective students, enhancing their application experience.

## References

1. Kirilov, R. (2020). Technological Opportunities for the Digitization of the Career Development Processes. *Economic Alternatives*, 1, 184-195.
2. Kazakova, M. (2023). Development of Secondary Education through Digital Learning Methods. *Ikonomiceski i Sotsialni Alternativi*, (2), 130-140.
3. Boyanov, L., Kyoreva, K., & Andonov, V. (2018). Improving Identification in Educational Activities Through Two Factor Authentication. In *Proceedings of International Conference on Application of Information and Communication Technology and Statistics in Economy and Education (ICAICTSEE)* (pp. 105-111). International Conference on Application of Information and Communication Technology and Statistics in Economy and Education (ICAICTSEE).
4. Katti, J., Agarwal, J., Bharata, S., Shinde, S., Mane, S., & Biradar, V. (2022, March). University admission prediction using Google Vertex AI. In *2022 First International Conference on Artificial Intelligence Trends and Pattern Recognition (ICAITPR)* (pp. 1-5). IEEE.
5. Le Hoanh Su, T. D. H., Thi-Yen-Linh, T., Thi-Duyen-Ngoc, N., Bao-Tuyen, L., & Ha-Phuong-Truc, N. (2020). Development of an AI Chatbot to support admissions and career guidance for universities. *International Journal of Emerging Multidisciplinary Research*, 4(2), 11-17.
6. Alothman, B., Alazmi, H., Ali, M. B., Alqallaf, N., & Khan, M. (2022, July). Accelerating university admission system using machine learning techniques. In *2022 Thirteenth International Conference on Ubiquitous and Future Networks (ICUFN)* (pp. 439-443). IEEE.
7. Lira, B., Gardner, M., Quirk, A., Stone, C., Rao, A., Ungar, L., ... & Duckworth, A. L. (2023). Using artificial intelligence to assess personal qualities in college admissions. *Science Advances*, 9(41), eadg9405.
8. Sarker, K. C., Rahman, M. M., & Siam, A. (2023, June). Anglo-Bangla language-based AI chatbot for Bangladeshi university admission system. In *2023 International Conference on Communications, Computing and Artificial Intelligence (CCCAI)* (pp. 42-46). IEEE.
9. Nguyen, M. T., Tran-Tien, M., Viet, A. P., Vu, H. T., & Nguyen, V. H. (2021, November). Building a chatbot for supporting the admission of universities. In *2021 13th International Conference on Knowledge and Systems Engineering (KSE)* (pp. 1-6). IEEE.
10. Prabha, M., Saraswathi, P., Karuppasamy, M., JansiRani, M., Dharshana, V., & RS, G. K. (2023, November). Student Chabot for University Admission Using Artificial Intelligence. In *2023 3rd International Conference on Advancement in Electronics & Communication Engineering (AECE)* (pp. 512-515). IEEE.
11. Hamzah, M. A., Siaulhak, S., Iriansa, I., Jumardi, A., & Aman, A. (2023). Utilization of Artificial Intelligence in Chatbot Development for New Student Admission Support at Cokroaminoto University Palopo. *Brilliance: Research of Artificial Intelligence*, 3(2), 434-440.



# Features and Challenges of Natural Language Processing in the Modern Digital World

**Yavor Tabov**

University of National and World Economy, Sofia, Bulgaria

*jtabov@unwe.bg*

**Abstract.** Natural Language Processing (NLP) is a rapidly evolving field focused on enabling computers to understand, interpret, and generate human language. This paper provides an overview of NLP, detailing some of its core features, such as machine translation, sentiment analysis, and speech recognition. It also explores significant challenges faced by NLP, including technical and linguistic difficulties. Finally, conclusions from the study are presented, summarizing the features and challenges of NLP.

**Keywords:** Natural Language Processing, Natural Language Understanding, Natural Language Generation, Sentiment Analysis, Machine Translation, Question Answering, Text Classification, Speech Recognition

## 1. Overview of natural language processing

The modern era of technology has a tremendous impact on society, revolutionizing how we live, work, and communicate. In the context of natural language processing (NLP), technology has transformed industries and daily chores by enabling machines to understand and interact with human language [6].

NLP can be applied in various domains such as machine translation, email spam detection, information extraction, summarization, and question answering [8]. NLP focuses on enabling computers to understand human language. It is closely related to literature and linguistics, often intersecting with language theory, psychology, and cognitive science. As a major branch of artificial intelligence and computational linguistics, NLP includes various subfields like machine translation (converting one natural language to another), speech recognition (identifying human voice signals), and text processing [14].

The general methodology of NLP is presented in Fig. 1.

The general methodology of NLP includes the following five phases:

- **Lexical analysis.** This phase involves identifying and analyzing the structure of words. The process breaks down a given sentence into paragraphs, sentences, and then individual words;
- **Syntactic analysis.** During this phase, a parsing tree is constructed to establish the relationships between words and generate the tree structure;
- **Semantic analysis.** This phase focuses on understanding the correct context of the words used;
- **Discourse integration.** This phase examines how sentences connect with one another, ensuring that the context of previous statements is appropriately linked to the current ones;
- **Pragmatic analysis.** This phase involves interpreting language aspects that require real-world knowledge[14].

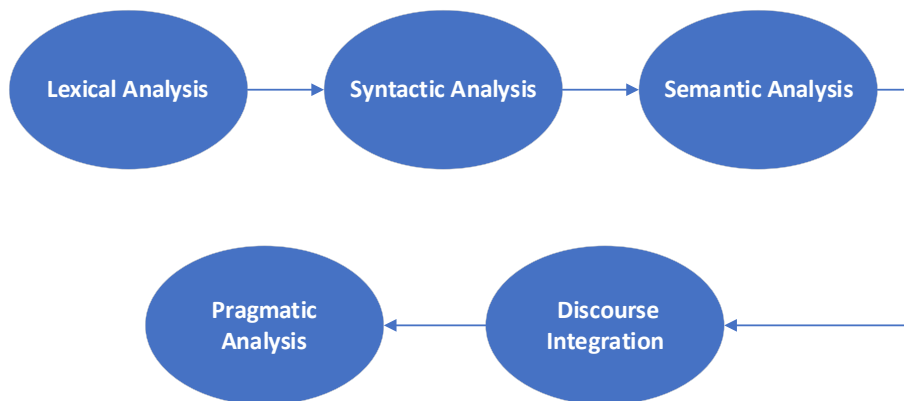


Fig. 1 General methodology of NLP

NLP involves processing natural language input from humans and generating meaningful responses from machines, allowing for natural language interactions with computers. NLP has two primary components: Natural Language Understanding (NLU) and Natural Language Generation (NLG) [3]. NLU allows machines to comprehend and analyze natural language by extracting concepts, entities, emotions, and keywords.

It is commonly used in customer support to interpret and address issues reported by users, whether spoken or written. On the other hand, NLG involves creating coherent phrases, sentences, and paragraphs from an internal representation [8]. Technologies such as voice assistants like Siri, speech-to-text features in chat apps, machine translation, and text generation methods all rely on NLP [3].

## 2. Features of natural language processing

NLP has significantly impacted various fields by enabling machines to process and use human language in meaningful ways. This capability has transformed areas such as customer service, where automated systems can now handle inquiries and support, and content creation, where machines can assist in generating text and media. The literature in the field of NLP explores the following features:

- **Text classification.** Text classification involves assigning categories to text documents based on specific rules that identify each class. While handwritten rules can be effective, they are labor-intensive to create and maintain, requiring significant manpower. Text classification is a particularly challenging area in machine learning due to the complexity of understanding natural language. A crucial step in the classification process is text representation, which is essential for accurately interpreting and categorizing text. As text classification continues to evolve, efficient methods for representation and classification are critical to managing the complexities of diverse textual data [4].

Classification is a fundamental problem in NLP that involves categorizing text into predefined categories. Its importance has surged with the expansion of social networks, blogs, forums, and online academic resources, making it a critical component of information retrieval systems and search engines. Additionally, text classification is valuable for applications such as email and SMS spam filtering. Key techniques in text classification include feature extraction, data reduction, deep learning model selection, and model evaluation. These techniques enable classification at various levels of granularity-document, paragraph,

sentence, and clause-demonstrating the versatility and significance of text classification in managing and interpreting diverse textual data [17];

- **Sentiment analysis.** Sentiment analysis is a crucial technique in NLP that focuses on identifying and extracting subjective information from text. It involves categorizing sentiments expressed in text as positive, negative, or neutral, and deciphering underlying opinions, attitudes, emotions, and feelings. The increasing volume of user-generated content on social media, blogs, reviews, and discussion forums has significantly boosted the relevance of sentiment analysis. Businesses leverage this technique to gain insights into customer feedback, monitor brand reputation, and inform strategic decisions. By analyzing various sources of textual data, sentiment analysis helps organizations understand public sentiment and respond effectively [13];

- **Machine translation.** Machine Translation (MT) is a process where computer software automatically translates text from one natural language to another. It is distinct from computer-aided translation and interactive translation methods. MT plays a crucial role in facilitating cross-lingual communication by automating the translation process, making it an essential tool for breaking language barriers [10].

Machine translation is a key benchmark for evaluating language understanding in computational systems. It involves two main components: language analysis and language generation. In a typical machine translation system, texts are translated from a source language into one or more target languages. The process includes analyzing the structure of the original text, breaking it down into translatable units, and then generating a semantically equivalent sequence in the target language. Machine translation is capable of processing large volumes of text quickly, with the primary objective of ensuring that the meaning of the original text is accurately conveyed in the translated output [1];

- **Speech Recognition and Speech Synthesis.** Speech recognition is the process of converting spoken language into text through computer programs or algorithms [5]. It plays a vital role in signal processing and is commonly used in applications like voice-activated assistants, such as Siri and Google. These systems rely on speech recognition to translate spoken queries into text, which is then used to perform actions or searches. Modern speech recognition techniques are adept at managing variations in accents and distinguishing similar-sounding words based on context, enhancing their accuracy and functionality.

Text-to-Speech (TTS) and Speech-to-Text (STT) are complementary but distinct technologies. TTS, or speech synthesis, enables machines to "speak" written text, either by concatenating recorded speech segments or generating entirely synthetic speech through vocal tract models. While TTS focuses on producing audible speech from text, STT is concerned with converting spoken language into text. Both technologies play crucial roles in enhancing human-computer interaction by bridging the gap between text and speech. [11];

- **Conversational agents and chatbots.** Conversational Agents (CA), also known as Virtual Assistants or Chatbots, are sophisticated software programs designed to interact with users using natural language. They are increasingly employed across various domains, including tourism and healthcare, to provide continuous support and perform tasks. The development of CA requires expertise in multiple areas, such as software engineering, machine learning, deep learning, and natural language processing. These systems integrate various speech and language technologies, including Automatic Speech Recognition (ASR), Natural Language Understanding, Dialogue Management (DM), Natural Language Generation, and Text-to-Speech Synthesis (TTS). As a result, CA represent a complex convergence of technologies aimed at facilitating natural and effective human-computer interactions [12].

Chatbots are advanced computer programs designed to facilitate interactive communication between humans and virtual assistants. Their popularity has surged in recent years, driven by significant advancements in artificial intelligence, machine learning, neural networks, and natural language processing. These improvements enable chatbots to effectively engage with users through dynamic and interactive queries, making them a valuable tool for various applications [6];

- **Question answering.** Question answering systems (QAS) have garnered significant attention in the artificial intelligence community due to advancements in learning technologies [7]. These systems, designed to automatically answer questions posed in natural language, are utilized across various domains including general knowledge, customer service, technical support, and healthcare. Modern QAS often leverage machine learning models like GPT-3, which have demonstrated impressive capabilities in generating accurate and high-quality answers. A QAS typically comprises two key components: question processing, which involves understanding the user's query through tasks like named entity recognition and parsing, and answer generation, which involves producing responses through information retrieval, text summarization, and natural language generation. The growing popularity of QAS in chatbots and virtual assistants underscores their importance and has driven ongoing research and development to enhance their performance and effectiveness [2];

- **Text summarization.** Summarization is the process of condensing a text into a shorter version while preserving essential information and meaning. This task involves either extracting key elements from the original text or generating new, coherent content that reflects the source's core information [18].

Text summarization is crucial for managing the increasing volume of data, as manually summarizing large documents is both labor-intensive and time-consuming. Text documents can be summarized using either extractive methods, which select key phrases from the original text, or abstractive methods, which generate new sentences to convey the main ideas. Summarizing text effectively supports various natural language processing tasks, such as question answering and text classification, and enhances data retrieval by reducing information access times. Text summarization systems improve efficiency by processing large amounts of text and provide more consistent results than manual summarization [15].

The reviewed features are shown in Fig. 2.

### **3. Challenges of natural language processing**

The primary goals of NLP are to interpret, analyze, and manipulate natural language data using various algorithms, tools, and methods. However, achieving these objectives is challenging due to the complexity and variability of natural language data. As a result, no single approach can address all NLP challenges, prompting extensive research and development of diverse tools and methods. This has garnered significant attention from researchers, driving advancements in NLP and related fields [8]. In today's digital age, the overwhelming abundance of information, particularly multilingual content on the Internet, necessitates the use of automatic text summarization to manage data efficiently. Despite the rapid growth of data, businesses only analyze a small fraction of it, leading to significant economic costs. As data creation continues to soar, advanced data processing tools become increasingly vital for handling and making sense of this vast amount of information [18]. With the world increasingly online, making data accessible to everyone poses significant challenges, particularly due to language barriers. The diverse languages, each with unique sentence structures and grammar, complicate this task. Machine Translation, using statistical engines like Google Translate, helps bridge this gap by translating phrases between languages.

However, the main challenge lies not in the direct translation of words but in preserving the meaning of sentences, along with correct grammar and tenses. Effective machine translation is crucial for ensuring meaningful and accurate communication across different languages [8]. According to researchers, machine translation has a major challenge towards accurate translation into natural language [14]. Automated text summarization faces numerous challenges, including defining what constitutes a good summary, whether it should be extractive, abstractive, general, or query-driven. Implementing human summarization practices is difficult, requiring significant resources in terms of tools and data. Ensuring informativeness and coherence in machine-generated summaries, selecting appropriate tools, and dealing with a lack of annotated data are major hurdles. Evaluation methods, both intrinsic and extrinsic, are subjective and complex, making it hard to automate the evaluation process fully. Despite over fifty years of research and numerous improvements, text summarization performance remains average. Integrating automated summarization with other systems could enhance its intelligence and effectiveness [18].

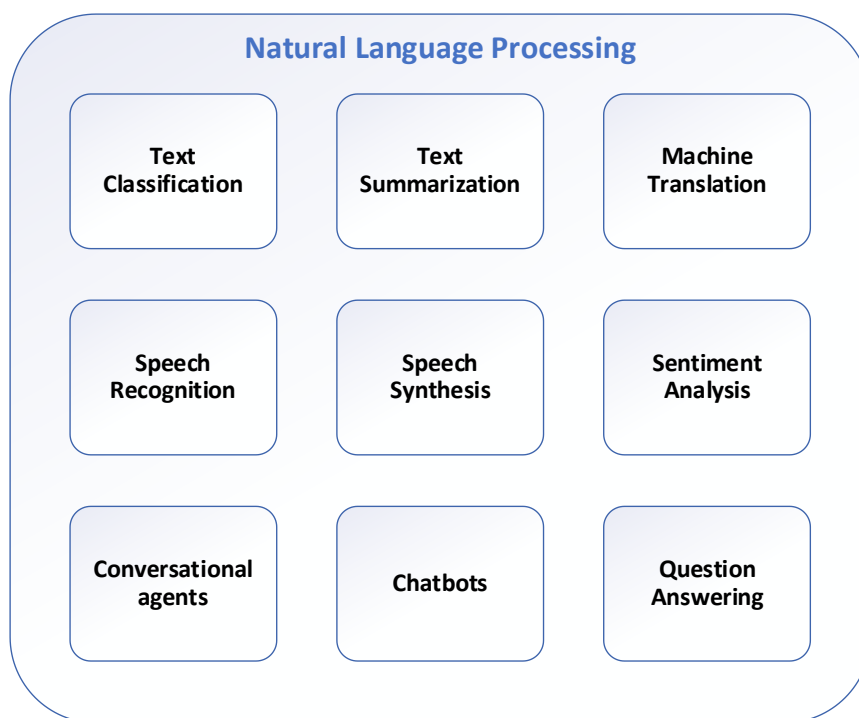


Fig. 2 Features of NLP

NLP is confronted with multiple challenges due to the inherent complexity and variability of human language. Contextual words and phrases, synonyms, homonyms, and varying levels of complexity in language make it difficult for algorithms to accurately interpret and process text. Sarcasm, irony, and ambiguous sentences further complicate the task. Additionally, informal phrases, idioms, and culture-specific expressions pose significant hurdles for creating universally applicable models. Differences in language use across various domains and geographic areas add another layer of difficulty. Despite advances in autocorrect

and grammar correction, predicting the writer's intent remains challenging. While NLP models have made significant progress for widely used languages, there is still a pressing need to develop more inclusive and adaptable models that cater to diverse linguistic and cultural contexts [8].

Natural Language Understanding in NLP faces significant challenges due to its requirement for multiple roles, including Natural Language Generation. Current NLU algorithms struggle to fully interpret natural language because human language often deviates from strict rules. Humans can understand incomplete or grammatically incorrect sentences through context, a capability that is difficult for machines to replicate. Training state-of-the-art deep learning models is computationally intensive and time-consuming, often taking up to a week on modern GPUs and being highly dependent on initialization and hyper-parameters.

Moreover, replicating previous results and tuning new models remains complex due to the need for re-implementing NLP components. Conventional algorithms often lack the precision required for unique contexts, and understanding the output of these algorithms is as important as their accuracy. Clustering methods can also fail to accurately define each class's meaning, sometimes necessitating manual recoding of data, which increases costs. Selecting an appropriate model for a specific study is another time-consuming and effort-intensive process. Insights from neuroscience and cognitive science, which study how the human brain processes language, could be valuable in developing more effective NLP models [1].

Challenges in NLP can be broadly divided into technical and linguistic aspects. Technically, NLP depends on statistical and learning capabilities, with corpus data being crucial for machine learning. However, constructing high-quality corpora is labor-intensive and time-consuming, limiting the availability of essential data and impeding progress in language acquisition. Linguistically, NLP models often lack the shared historical, cultural, and common sense knowledge that humans use in everyday conversations. This deficiency makes it difficult for machines to understand and process texts and conversations effectively, as they fail to grasp key contextual factors. Addressing these challenges is essential for advancing the capabilities of NLP systems. From a linguistic perspective, machines encounter several significant challenges in understanding natural language. These challenges include the diversity of languages, which complicates the development of a universal processing approach; ambiguity, where words or sentences can have multiple meanings; robustness, requiring consistent performance across different accents and conversation styles; dependency on practical knowledge, which affects language comprehension; and contextual understanding, where meaning depends on the surrounding environment. These linguistic challenges are often more critical than technical ones, highlighting the need for deep analysis and optimization of language processing methods. Addressing these issues, particularly through the lenses of phonetics, semantics, and pragmatics, is crucial for improving machine performance in natural language understanding [3].

NLP merges linguistics with Artificial Intelligence and is used in technologies like voice assistants and translation services. The field has advanced significantly over the past decade, largely due to the impact of machine learning and deep learning, which have improved both existing solutions and enabled new breakthroughs [9].

Business Process Management (BPM) aims to streamline organizational processes to ensure effective delivery of products and services. With the maturity of NLP and the availability of public frameworks, NLP has the potential to enhance BPM practices significantly. NLP researchers face several challenges when applying their methods to BPM. Key issues include enhancing the performance of semantic and pragmatic analyzers, adapting NLP tools to specific organizational domains, defining new tasks such as detecting process exclusivity and concurrency, incorporating world knowledge to fill in gaps left by common

sense, and improving information extraction from event logs. Addressing these challenges is crucial for effectively integrating NLP into BPM and improving process analysis and optimization [16].

#### 4. Conclusion

In conclusion from the presented research of the features and challenges of natural language processing, we can summarize the following notes:

- The goal of NLP is to enable computers to understand human language, intersecting with literature, linguistics, language theory, psychology, and cognitive science;
- General methodology of NLP includes five phases: lexical analysis, syntactic analysis, semantic analysis, discourse integration and pragmatic analysis;
- The features of NLP include text classification, sentiment analysis, machine translation, speech recognition, speech synthesis, conversational agents, chatbots, question answering and text summarization;
- Despite significant advancements and extensive research in NLP, challenges in interpreting, analyzing, and summarizing natural language data, managing data overload, and ensuring accurate machine translation across diverse languages remain substantial, necessitating the continuous development of advanced tools and methods;
- NLP relies heavily on statistical and machine learning capabilities, which require high-quality corpus data;
- NLP models lack the shared historical, cultural, and common sense knowledge humans use in conversations. These results in difficulties with language diversity, ambiguity, robustness across accents, dependency on practical knowledge, and contextual understanding, making it challenging for machines to process texts and conversations effectively.

#### References

1. Abro, A., Talpur, M., Jumani, A., Natural Language Processing Challenges and Issues: A Literature Review, Gazi University Journal of Science, volume 36, issue 4, 2023, ISSN: 1303-9709
2. Beta, T., Question-And-Answer System Using Natural Language Processing, 2023, <https://www.researchgate.net/>
3. Chen, D., Challenges of Natural Language Processing from a Linguistic Perspective, International Journal of Education and Humanities, Vol. 13, No. 2, 2024, 2770-6702
4. Dogra, V., Verma, S, Kavita, Chatterjee, P., Shafi, J., Choi, J., Ijaz, M., A Complete Process of Text Classification System Using State-of-the-Art NLP Models, Computational Intelligence and Neuroscience, Volume 2022, ISSN:1687-5273
5. Gaikwad, S, Gawali, B., Yannawar, P., A Review on Speech Recognition Technique. International Journal of Computer Applications, Volume 10, No. 3, November 2010, ISSN 0975-8887.
6. Gupta, A., Hathwar, D., Vijayakumar, A., Introduction to AI Chatbots, International Journal of Engineering Research & Technology (IJERT), Vol. 9 Issue 07, July -2020, ISSN: 2278-0181
7. Ishwari, K. S. D., Aneez, A. K. R. R., Sudheesan, S., Karunaratne, H. J. D. A., Nugaliyadde, A., Mallawarrachchi, Y., Advances in natural language question answering: A review, 2019, arXiv preprint arXiv:1904.05276.
8. Khurana, D., Koli, A., Khatter, K., Singh, S., Natural language processing: state of the art, current trends and challenges, Multimedia tools and applications, volume 82, 2022, ISSN: 1573-7721
9. Krasadakis, P., Sakkopoulos, E., Verykios, V., A Survey on Challenges and Advances in Natural Language Processing with a Focus on Legal Informatics and Low-Resource Languages, Electronics, Volume 13, Issue 3, 2024, ISSN: 2079-9292
10. Okpor, M., Machine Translation Approaches: Issues and Challenges. International Journal of Computer Science Issues, Vol. 11, Issue 5, No 2, September 2014, ISSN 1694-0784.

11. Olorunfemi, A., The depth of natural language processing on speech recognition synthesis model, 2024, <https://www.researchgate.net/>
12. Ouaddi, C., Benaddi, L., Jakimi, A., Architecture, Tools, and DSLs for Developing Conversational Agents: An Overview, *Procedia Computer Science* 231, 2024 , ISSN: 1877-0509
13. Prasad Gunasekaran, K., Exploring Sentiment Analysis Techniques in Natural Language Processing: A Comprehensive Review, 2023, arXiv e-prints, arXiv-2305.
14. Sharma, P., Nagashree, N., Survey on Natural Language Processing and its Applications, *INTERNATIONAL JOURNAL OF COMPUTATIONAL LEARNING AND INTELLIGENCE*, Vol. 1, Issue. 2, October 2022
15. Srivastava, A., Chauhan, K., Daharwal, H., Mukati, N., Kavimandan, P., Text Summarizer Using NLP (Natural Language Processing). *IRE Journals*, Volume 6, Issue 1, ISSN: 2456-8880
16. Van der Aa, H., Carmona, J., Leopold, H., Mendling, J., Padro, L, Challenges and opportunities of applying natural language processing in business process management, In *COLING 2018: The 27th International Conference on Computational Linguistics: Proc. of the onference: August 20-26, 2018 Santa Fe, New Mexico, USA*, Association for Computational Linguistics, ISBN 978-1-948087-50-6
17. Wan, Z., Text Classification: A Perspective of Deep Learning Methods, 2023, arXiv preprint arXiv:2309.13761.
18. Yadav, D., Desai, J., Yadav, A., Automatic Text Summarization Methods: A Comprehensive Review, 2022, arXiv: 2204.01849



## The Cash of The Big Bulgarian Companies: What Do Their Cash Flow Statements Show and What Do Not?

**Fanya Filipova, Teodora Zapryanova, Atanas Atanasov, Rumyna Marinova**

University of Economics - Varna, Varna, Bulgaria  
*fanya\_filipova@ue-varna.bg, teodorazap@ue-varna.bg*  
*atanasov\_at@ue-varna.bg, r\_marinova@ue-varna.bg*

**Abstract.** The subject of the study is the information about cash and cash flows of ten largest Bulgarian companies, which in recent years are almost constantly present in the annual ranking of the international insurance company Coface "CEE Top 500 Ranking". The researched period is between 2019 - 2022. The aim is to present and comment on several of the most important financial indicators, also used by banks creditors, calculated mainly from statement of cash flow (SCF) and their relationship with other statements of the annual financial report (with the profits/losses of these companies; with their revenues; with their indebtedness; with the net change in cash and cash equivalents, etc.). Based on this, the information is defined that the SCF directly does NOT provide, or is not visible in the SCF prepared by the direct method. Conclusions are formulated related to some identified gaps and weaknesses in the SCFs of the 10 Bulgarian companies, including the disclosure of the information (minimum disclosure requirements) related to them, according to the requirements of IAS 7. Through the research, an attempt is made to define the usefulness of the information from the SCF, but also to emphasize the need to check its quality, through cross-checking (binding) with other data in financial statements and through proposals for changes in IAS 7.

**Keywords:** Cash Flow Statement; direct and indirect method; IAS 7; SCF analysis; free cash flow; cash and cash equivalents.

### 1. Introduction

The statement of cash flows (CFS) as a mandatory component of the public annual financial statements (AFS) of enterprises appeared much later (in the 1980s) compared to the balance sheet and the income statement. To this day, it seems that he remains in the shadow of these two leading statements, despite the fact that he provides information about one of the most important driving forces in business – the CASH. Confirmation of a certain neglect of this report in Bulgaria is the fact that according to Art. 29, para. 3, 4 and 5 of the Accounting Law, the financial statement of certain sole traders, micro and small enterprises, may not contain a Cash Flow Statement. On the other hand, the listed categories of enterprises are the most numerous in our country, and at the same time they often use bank lending to finance their activities. And because of this, they should voluntarily prepare the CFS (because they are not obliged by law), if, for example, a creditor (the bank) demands it. In addition, according to the rules of the Bulgarian NAS 7 Statement of cash flows, the indirect method is not allowed in preparing of CFS, which would further complicate some domestic companies if they are owned by a foreign parent company, which according to its accounting policy, in its consolidated financial statements uses the indirect method. Through the indirect method, the transformation of accrued profit/loss (the accrual principle) into cash and cash equivalents (the cash principle) is very clearly presented. The exclusion of the indirect method as a possible

approach in our National Accounting Standards greatly reduces the possibility for interested users (including banks as creditors) to identify possible disproportions, deviations or violations in the reporting and presentation of the financial result (through the accrual principle), and its relation to actual cash flows (inflows and outflows).

The above raises the following questions:

✓ What is and how significant is the informational utility of SCF? Is there one at all in the current format of this report, regulated in IAS 7, and if "YES", who uses this information?

✓ What motivates the choice of a method (direct or indirect) when preparing the SCF by the enterprises, in cases where they apply the International Standards for Financial Statements/Reporting (IAS/IFRS) as an accounting base?

✓ What are the leading financial indicators related to the money and cash flows of enterprises, and to what extent does the current SCF format facilitate or complicate their calculation?

✓ Which financial indicators, based mainly on information from the SCF, show that there is a solid, robust relationship between an enterprise's ability to generate sufficient cash from its business and the amount of its bank financing? Or is it difficult to derive such dependencies?

The actuality of the problem related to money, cash flows of enterprises and information about them, on the one hand, and their relationship with bank lending, on the other, are of particular relevance for Bulgaria, as well as for most countries in Eastern Europe, where traditionally banking credit is the main source of business financing. According to Eurostat data for 2022 the value of the short-term and long-term loans represent 25,1 % of all liabilities of non-financial corporations in Bulgaria. (Eurostat, 2023) According to data for 2022, housing loans and loans to non-financial enterprises make up the majority of loans in every country in the European Union. The distribution of these loans to non-financial enterprises is as follows: in Bulgaria there are 57.33%, in the Czech Republic: 36.7%, Slovakia: 32.49%, Romania: 50.72%, Poland: 36.1%, Hungary: 54.88%. In addition, over the past 7-8 years, the IAS Board has made some changes to IAS 7 to improve the usefulness of the information in this statement to users, and discussions about other upcoming changes are ongoing. This undoubtedly provides an opportunity for research and discussion on the subject.

The purpose of this study is to make an attempt to define the usefulness of the information from the CFS, as well as the need to check its quality, by cross-checking (and binding) with the other statements of the annual report.

To achieve the above purpose, the following tasks are defined:

1). To present several most important indicators (also used by banks), calculated mainly from data from the annual CFS of a sample of 10 largest Bulgarian companies for the period 2019 - 2022. To briefly clarify what information they provide and to track how they change (mostly compared to the profits/losses of these companies, their revenues, with the net change in cash and cash equivalents) over the 4-year period under study.

2). To comment on what information (and a basis for calculating certain indicators) the Statement of cash flows does NOT directly provide (and data from the other statements should also be used – balance sheet / Statement of financial position and Statement of profit and loss and the other comprehensive income), or given data cannot be directly obtained from the SCF if it is prepared using the direct method.

The research seeks to show and justify what information users see in the companies' SCF, and also what they do not see directly, or it is possible that the information presented in the SCF is unclear, confusing, insufficient. In this way, an attempt is made to define the usefulness of the information from the SKF, but an emphasis is also placed on the need to

check its quality, through cross-sections (binding) with other statements and through proposals for changes in IAS 7. All this, however, in many greater detail, through a richer range of research approaches and based on a larger sample of enterprises, is the subject of ongoing research. The size of the sample (only 10 largest Bulgarian enterprises), as well as the limitations of the applied research approaches (content analysis of the information related to the Financial Statements and quantitative analysis of a selected group of financial indicators, based mainly on data from the Financial Statements), also define the current limitations of the study, which is the initial phase of a larger study that the authors are engaged in.

The main research statements (hypotheses) to which the study tries to give an initial, still incomplete answer are the following:

H1: For the largest Bulgarian companies, which are predominantly non-public companies, the direct method of preparing the SCF is dominant because of the greater ease of its application, and not because of the information convenience of banks as creditors.

H2: The direct method of preparation of the SCF does not help to establish manipulations (or assumptions about such) when applying the accrual principle, characteristic of items in the Statement of Financial Position and the Statement of Comprehensive Income.

H3: The calculation of financial indicators on the basis of the SCF (and in combination with data from the other elements of the financial statement) partially helps to identify gaps and manipulations in the accounting information based on the accrual principle.

H4: The direct method of preparing the SCF is more closely related to bank lending on the basis of assets.

## **2. Conceptual basis of the study**

### **a. Purpose and usefulness of information from the Statement of Cash Flows**

The phrase "Cash is king!" is commonly known. But, based on the principle of double-entry bookkeeping, accounting has always given priority to the periodic summarization and presentation of accounting information, to those data related to assets, liabilities and capital (of which, of course, money and financial results are part of). Therefore, the balance sheet (and then the income statement) historically appears first, before the cash flow statement. However, experts point out that "in 1863, the Northern Central Railroad published a statement of its financial transactions, including information on its annual cash inflows and outflows." (Beresford, D. R., 1988).

Based on the understanding that working capital (as the difference between current assets and current liabilities) approximates the short-term potential cash supply as a cash flow representation, in 1902 the United States Steel Corporation prepared a statement of changes in so-called "funds" during the year, as in this case, funds meant the difference between current assets and current liabilities. This statement (of funds resulting from changes in working capital) became very popular in advanced capitalist economies after 1920. In 1971, the American Accounting Standards Board (APB), the predecessor of today's Financial Accounting Standards Board (FASB), published opinion No. 19 (Opinion No. 19), which requires the statement of funds to be included in the annual financial report and to be audited by the independent auditors. This report is called the Statement of Changes in Financial Position. In the 1980s, only 10% of Fortune magazine's 500 companies had a cash and cash flow focus in the fund report. The remaining 90% report information on changes in working capital. In 1985, already 70% of them had a reporting focus on cash.

Just then, the American Financial Accounting Standards Board (FASB) published Statement of Financial Accounting Concepts No. 5, in which the Board sets out the understanding that a complete set of financial statements should also include a Statement of Cash Flows. In late 1987, the FASB published Statement No. 95, which replaces APB Opinion

No. 19. And in it, the FASB regulates the preparation of the CFS without engaging in the discussion of what the term "funds" means and what they include (cash or working capital or both). Thus, this statement now replaces the statement of changes in funds (statement of changes in financial position). (Alves et al, 2015) In 1992, the IAS Board also created a standard on the preparation of the Statement of Cash Flows, and its predecessor was the standard published in 1977 on the Statement of Changes in Financial Position. Consequently, for nearly four decades businesses (and especially those that prepare full financial statements) have been creating, presenting and disclosing information related to their cash and cash flows. It would be hard to find arguments to dismiss the usefulness of the third-party information that this reference has. Through it (depending on the form of preparation) it is possible to see the movements, directions of incoming and outgoing cash flows; the net cash change in each of the three activities – operational, investment and financial; the total cash change for the reporting period; the transformation of the accrued operating profit into money, etc. And on this basis, to calculate a number of financial indicators, tying the cash and its change during the year, with other important financial values such as income, expenses, assets, total debt, including short term and long term.

Based on this, external users can make judgments, forecasts and assumptions about management's ability to manage the company's cash resources; to make correct investment decisions; to use sources adequate in size and terms for financing the activity, as well as the possibility to distribute dividends. Information about the size and movement of the company's cash and cash equivalents should be of essential importance for creditors (banks, etc.). How and to what extent they effectively use the data from the Cash Flow Statement of their borrowers is a question that, according to the authors of the present study, has not been sufficiently well studied. Especially when it comes to countries like Bulgaria and those from Eastern Europe (members of the EU), where bank lending is the dominant source of business financing. For the needs of the management itself, weekly, monthly and longer-term cash flow forecasts are an important tool for planning, forecasting and managing the liquidity and solvency of the enterprise within reasonable limits, as well as for important investment decisions.

**b. Legal framework – IAS 7 Statement of cash flows (and IAS 7 of the same name); American ASC 230 “Statement of Cash Flows”. Problem areas.**

This research does not aim to examine in detail the regulatory framework related to the Statement of Cash Flows. Rather, the aim is to highlight some of the more debatable and problematic points related to the preparation of this report, namely:

- Which method to apply for the preparation of the CFS: the direct method, which covers and classifies only cash transactions in the form of cash inflows and outflows, or to continue using the indirect method, by which the accrual principle is tied to the cash flow principle and the accrued operating profit is almost transformed into cash profit;
- Why do businesses in countries with developed capital markets (and here in Bulgaria – mostly banks) prefer the indirect method for preparing the SCF? Why do the American Financial Accounting Standards Board (FASB) and the IASB give priority (mandatory application) to the direct method?
- Whether to clarify and detail the requirements related to the way cash flows are classified and resp. to remove some currently existing business choices, e.g. related to the classification of dividends and interest paid; whether to introduce more specific criteria for classifying cash flows, incl. and according to the characteristics of different industries?
- Should the SCF for banking institutions remain in the same form and if "no", what should the change be?

- Should information regarding the type and effect of non-monetary transactions as a source of financing for the enterprise be provided within the framework of the SCF itself, or in the disclosure part related to it, and what exactly should this information include?
- Should requirements be introduced for comparison, cross-checking of the information from the SCF with the other features, items in the financial statement to which it is related, as well as a requirement for reporting by the management in case of significant fluctuations, or even illogical connections and proportions?
- Can the usefulness of the SCF be improved through additional disclosure? (under the SCF itself, or in a separate note (explanatory appendix) to the financial statement?)

The authors' opinion on the above discussion questions related to the Statement of Cash Flows is a subject of ongoing research.

### **c. Changes to IAS 7 Statement of Cash Flows made and discussed by the IASB Board**

Unchanged for almost 25 years (1992-2016), in 2016 the IAS Board made a change to IAS 7 related to the Disclosure Initiative, namely: enterprises should disclose changes in their obligations arising in their financial activities. Some minor changes were made in IAS 7 and through IFRS 10, IFRS 11, IFRS 16, IFRS 17. Discussions in the IAS Board during the period 2017-2019 regarding the classification of cash flows related to interest and dividends did not lead to a change in IAS 7 and the existing alternatives remained - opportunities for choice and judgment on the part of enterprises when classifying them in SCF. Discussions continued (and continue) and about: whether a change is needed when defining the purpose of preparing the SCF (e.g. due to the increasingly shared need of users of financial information to receive information (by way of disclosure) about non-monetary financing, as well as regarding other, non-monetary transactions). In 2023, the IAS Board published changes to IAS 7 and IFRS 7, regarding the disclosure of quantitative and qualitative information about financial commitments related to suppliers (where payment to suppliers is tied to factoring contracts, for example). They are in force from 01.01.2024, and their earlier application is allowed. At the present time (early 2024), the following important issues related to the SCF remain under discussion by the IAS Board:

- Is there a need to change in defining the purpose of the Statement of Cash Flows?
- Is there a need to eliminate the alternative (indirect method) method for preparing the SCF?
- Is there a need for more detailed and strict regulation of criteria for classifying cash flows in the three activities: operational, investment and financial?
- Is there a need to change the structure of the SCF for banks and non-bank financial institutions?
- Is there a need for more extensive disclosure, including about non-monetary deals and transactions, and how and where, in which part of the annual financial statement should this be done?

The authors seek answers to the above discussion questions in their ongoing research.

## **3. Research methodology**

The research methodology includes the use of different research tools in order to confirm or reject the formulated hypotheses. The methods used include: analysis of the content of the annual financial statements of the 10 Bulgarian companies included in Cofface Top 500, as well as calculation and analysis of selected cash flow indicators, based on information from

the financial statements of the companies. CEE Top 500 Ranking is a joint project of Coface offices in Central Europe. This ranking covers the largest companies in the region - based on sales revenue (turnover). The study includes the following countries: Bulgaria, Croatia, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Serbia, Slovakia and Slovenia. The CEE Top 500 identifies the largest 10 companies in each of the specified countries (with a turnover  $\geq 300$  million euros). Financial service providers such as banks, insurance companies, leasing companies and brokers are excluded from the ranking. In addition to revenue, the CEE Top 500 survey also includes other key corporate indicators, e.g. the net profit, the number of employed persons and the corresponding changes compared to the previous year. (COFACE, 2022) The sample covers a period of four years 2019 – 2022 and the main information source was the annual financial statements of the enterprises prepared on an individual basis. The variety of indicators for analysis based on information from the Statement of Cash Flows (and other related financial information from the financial statement) is large. In the present study, the authors are limited to examining the size and dynamics of 10 indicators for the 10 most studied Bulgarian enterprises. The authors consider that these are the most used indicators for analyzing the liquidity, creditworthiness and solvency of companies - credit seekers, by banks - creditors.

These indicators include:

- ✓ Current Cash Flow (CCF) = Net operating profit adjusted by non-cash income and expenses; increased by the financial income, reduced by the financial expenses, as well as by the corporate tax paid;
- ✓ Current cash flow (CCF) as % of sales;
- ✓ Current creditworthiness;
- ✓ Net current creditworthiness;
- ✓ Current creditworthiness / long-term debt;
- ✓ Free cash flow;
- ✓ Current creditworthiness based on free cash flow;
- ✓ Operating cash flow;
- ✓ Current creditworthiness based on operating cash flow = Current cash flow before interest expense and taxes/Interest expense;
- ✓ Interest coverage ratio.

#### **4. Data and results**

The structure and format of the presented information regarding the company's cash resources and their change over time is based on the cash flow report prepared by them for the relevant reporting period. Since the companies in the sample apply IAS/IFRS as a basis for preparing their financial statements, we believe that an important indicator for the presentation of information in the Statement of Cash Flows is the chosen method for preparing the statement - direct or indirect. Table 1 presents the results in this regard:

Of all the companies, only one chose to prepare its statement of cash flows using the indirect method. This enterprise is also the only public enterprise registered on the Bulgarian Stock Exchange, which supports our thesis that the presence of a stock market listing can be a driving motive for using the indirect method, along with the ownership structure - concentrated or dispersed. We believe that the last point regarding the ownership structure should be the subject of a separate study, since in Bulgarian practice there are many cases of public companies that, however, do not use the indirect method for preparing the SCF. In this regard, the presence of dispersed ownership in combination with the listing on a stock exchange may

prove to be of key importance in the choice of a method for the preparation of the SCF by the enterprises.

Table 1. Method for preparing the Statement of Cash Flows applied by the surveyed enterprises

No	Enterprise	Base for preparation	Method used
1	Bulgarian Energy Holding AD	IAS/IFRS	direct
2	Aurubis Bulgaria AD	IAS/IFRS	direct
3	National Electric Company EAD	IAS/IFRS	direct
4	Eurohold Bulgaria AD	IAS/IFRS	indirect
5	Lukoil Bulgaria EOOD	IAS/IFRS	direct
6	Astra Bioplant Ltd	IAS/IFRS	direct
7	Kozloduy NPP EAD	IAS/IFRS	direct
8	Advance Properties Ltd	IAS/IFRS	direct
9	Bulgargaz EAD	IAS/IFRS	direct
10	Saksa Ltd	IAS/IFRS	direct

\*The order of presentation is the order of Coface CEE Top 500 Ranking

It is important to note that this enterprise during the period under study reports systemic losses, which for this period increase almost 10 times, and this is combined with a negative cash flow from operational activity, which, however, is 25 times less than the reported losses. At the same time, the results confirm the formulated hypothesis that the direct method of preparing the Cash Flow Statement is dominant for the studied sample, which may be due to the greater ease of its application. Despite the fact that seven of these companies use bank lending, they apply the direct method for preparing the SCF, by which the banks as creditors could not directly calculate the indicators important to them, which are indicated in Appendix 1. This logically reinforces the understanding expressed by hypothesis 1: the direct method of drawing up the SCF is dominant because of the greater ease of its application, and not because of the information convenience of banks as creditors. Only two of the investigated companies (Aurubis Bulgaria AD and Lukoil Bulgaria) have foreign mothers and do not have bank loans, but they also use the direct method for preparing the SCF. This fact also raises the question of whether the origin of ownership - local or foreign - affects the choice of method for preparing the Statement of Cash Flows.

In connection with the initial testing of Hypothesis 2 regarding the direct method of preparing the SCF and establishing manipulations (or assumptions of such) when applying the accrual principle (in the Statement of Financial Position and in the Statement of Comprehensive Income), the authors are directed to the links and the interdependencies between the following several financial values of the investigated companies: net amount of the annual financial result (profit/loss) - net change in cash - sales revenue - cash availability at the end of the reporting period. Table 2 presents the relevant data.

Table 2. Net amount of the annual financial result, net change in cash, sales revenue, cash availability at the end of the reporting period of the studied enterprises

Enterprise	2019				2020				2021				2022			
	Net profit/loss	Net change in cash	Sales revenue	Cash at 31.12.	Net profit/loss	Net change in cash	Sales revenue	Cash at 31.12.	Net profit/loss	Net change in cash	Sales revenue	Cash at 31.12.	Net profit/loss	Net change in cash	Sales revenue	Cash at 31.12.
Enterprise 1	70551	-840	183681	223	41376	-152	53724	71	309099	10254	327032	10325	264249	-7999	312208	2326
Enterprise 2	324892	156415	1314050	395583	275810	90165	1257968	485106	890304	429159	2706252	914193	429463	-174838	6038599	738743
Enterprise 3	14165	-2467	1902432	3506	39460	15641	1681747	18300	244717	-7501	3199419	11692	90028	6301	7280815	17900
Enterprise 4	265785	26322	4719734	39450	264780	4864	5226288	44314	292077	-17424	6508513	26890	499257	-24346	7853054	2544
Enterprise 5	34429	-20	1404739	80	39657	171932	698762	171857	63905	-172809	2135224	211	-100431	207635	4927370	207395
Enterprise 6	584255	266028	355607	413835	3007	-241009	169909	173465	320828	558437	763804	730011	582061	1564602	1022022	2289617
Enterprise 7	-14631	-146	7515	138	-16754	136	4580	267	-116388	-159	1923	115	-102147	2131	3684	2238
Enterprise 8	68228	-1404	3311529	4227	9675	-1035	2168434	3267	75352	740	3217820	3888	149215	92509	7374093	96316
Enterprise 9	-1260	-5023	2781206	28032	46314	70507	2518934	98455	650096	347549	4075756	445633	1083256	-354167	5232919	91890
Enterprise 10	22827	1835	1783925	31706	16302	-10153	1407717	21553	18719	17096	2011721	38649	47969	-18526	2660830	20123

\* The order of the enterprises does not match the order in Table 1

An initial analysis of the above data (which limits the present study) indicates the following:

- The dynamics in the change in the net profit/loss of the companies during the research period (2019-2022) often does not have a direct reflection and dependence with the change in the net cash flows and cash availability at the end of the respective year;
- One of the enterprises – enterprise 7, for the entire investigated period reports a loss (increasing in size), but parallel to this, its cash availability also increases at the end of the respective year, although there is a decrease in the value of its revenue from sales;
- Another enterprise – enterprise 5, reports a significant loss (BGN 100,431 thousand) in 2022, against the background of reported net profits in the previous three years in the range between BGN 34 and 64 thousand. At the same time, its cash in the end of 2022 in the amount of BGN 207395 thousand are the largest for the entire period under study, which may be related to its drastically increased sales revenue in the same year 2022, (almost more than twice, compared to the previous year 2021), but it is not clear from SCF what the correlation is between the significant reported loss for 2022, the significantly increased revenues and the greatly increased cash availability at the end of the same year;
- At enterprise 4, the net financial result at the end of the researched period - 2022 has grown almost twice compared to the beginning of the researched period - 2019, and at the same time, its income has also grown almost twice, and its cash at the end of 2022 compared to 2019 has decreased almost 15 times.

Similar observations can be made about the remaining companies in the sample, which at this stage of ongoing research on the subject by the authors, leads to the conclusion that the in-depth establishment of correlations between such important financial values as net profit/loss, net change of cash, sales revenue, cash on hand at the end of the reporting period, can reveal interesting cash management approaches of the companies under review, as well as whether there are unusual relationships between accrued results and revenue, and their relationship to cash.

Regarding the initial testing of Hypothesis 3 in relation to leading financial indicators primarily used by banks as creditors, calculated on the basis of data from the Cash Flow Statement (and in combination with data from the other statements of the annual report) and establishing gaps and manipulations in accounting information based on the accrual basis, the data are presented in Appendix 1 for the surveyed companies.

Enterprise 1 has achieved improvements in certain aspects of liquidity and cash flow, such as a reduction in CCF and an increase in the interest coverage ratio. However, challenges remain, particularly with respect to the decline in current liquidity and free cash flow in 2022.



Further investigation into the reasons behind these trends and potential corrective actions is needed.

In the case of Enterprise 3, the main conclusions from the indicated data show that the company has experienced fluctuations in terms of various indicators of liquidity and cash flows. There have been improvements in some areas, such as CCF and the ratio of current liquidity to long-term debt. However, negative operating cash flow in 2022 and fluctuations in free cash flow require closer examination to identify factors contributing to these trends. The overall analysis of the results for Enterprise 2 shows an improvement in the company's cash flow management and liquidity over the past years. The interest coverage ratio is high and indicates the company's ability to cover its interest by a large margin. Current creditworthiness remained almost at zero during the period, which can be interpreted as a low dependence on current financing obligations. However, it is necessary to carefully consider the factors that led to changes in free cash flow and operating cash flow in 2022.

Due to publication volume limitations, the present study does not present the individual analysis for the remaining 7 companies in the sample. In Appendix 1, their financial indicators are systematized, calculated primarily on the basis of the Cash Flow Statement.

With regard to the initial testing of hypothesis 4 regarding the relationship between the direct method of preparing the SCF and bank lending based on assets, the results of the conducted research provide grounds for confirming the formulated hypothesis of a close relationship. The data show that 3 out of 10 enterprises (30%) have not received bank loans. These companies should be excluded from the analysis, as they do not meet one of the main requirements – to have received bank loans. After reducing the sample, its scope includes seven enterprises that received bank loans during the period under review. Of them, 3 enterprises, i.e. 42.86% (3 out of 7) have secured their bank loans with a mortgage on real estate and/or a special lien on non-current assets in combination with a lien on securities and/or receivables. Two of the enterprises (28.57 %) disclosed as collateral only a pledge of shares and/or shares. One enterprise (14.29 %) indicated that the loan it received was secured by a state financial guarantee given the specifics of its ownership structure, and one enterprise (14.29 %) did not specify information regarding the collateral for the loans received. All enterprises that secured the received loans with a mortgage and/or a pledge of non-current assets used the direct method for compiling the Statement of cash flows during the three investigated periods. This finding, although based on a limited sample, confirms the formulated hypothesis that the direct method is more closely related to asset-based bank lending.

An important question that this study raises and analyzes at an initial stage is: *What CANNOT BE SEEN in the Cash Flow Statement of the companies under study?* As an example, the accounting of non-monetary transactions can be indicated: in four of the analyzed enterprises, during the period, such transactions were observed, which were mainly of an investment and financial nature, in which no cash or cash equivalents were used and which were not reflected in the Cash Flow Statement? It is noteworthy that in one of the investigated enterprises, these deals just for 2022 are 27! The specified transactions, in which no cash was actually used, include a variety of operations such as:

- ✓ Agreement for the loan of financial instruments with related and unrelated persons;
- ✓ Debt Substitution Agreement
- ✓ Transfer/Assuming of Obligation by Deed of Assignment
- ✓ Agreement to offset receivables or liabilities with related and unrelated counterparties
- ✓ Contract for assignment of receivables

- ✓ Agreement for the purchase and sale of financial instruments (repurchased own bonds) as the obligation with the transferor is settled against the transfer of shares of a subsidiary
- ✓ Tripartite agreement for set-off of a receivable from a transferee with a loan obligation with a related party
- ✓ Concluded cash pool contracts with an ultimate parent company, by virtue of which the company can absorb credit on its own account in the cash pool in the amount of several hundred million euros
- ✓ Offsetting loans between related parties within an economic group

All these transactions are not directly reflected in the SCF, since they do not lead to the occurrence of an incoming or outgoing cash flow, or if such occurs, it partially represents the changes in the property and financial situation, and at the same time, the monetary equivalent of these non-monetary transactions is essential. The stated findings raise the question of the need for more extensive disclosure related to the company's cash flows, including regarding non-monetary deals and transactions carried out during the period. We believe that this is necessary in order to reveal the full volume of the activity carried out and the impact of this type of transaction on the financial indicators of the enterprise, as well as in order to show the "full" picture of the cash flows of the enterprise.

## 5. Conclusion

The usefulness of the information in the public Cash Flow Statements of enterprises is indisputable. Through it, the movement of cash (incoming and outgoing cash flows) is visible, through which new aspects of the management of these essential company assets are presented. Such information cannot be obtained from the statement of financial position or the statement of comprehensive income, which are based on the accrual basis. At the same time, however, the information in the statement of cash flows is closely related to other financial values from the remaining elements of the annual report, and atypical or significant illogical dependencies between them should be a reason for a deeper analysis.

At an early stage of research, the present study confirms the main research claims, namely:

- 1). For the largest Bulgarian companies included in the sample, which are predominantly non-public companies as well, the direct method of preparing the SCF is dominant because of the greater ease of its application, and not because of the information convenience of creditor banks.
- 2). The direct method of preparing the statement of cash flows does not help to establish manipulations (or assumptions about such) when applying the accrual principle, which is characteristic of the items in the statement of financial position and the statement of comprehensive income.
- 3). The calculation of financial indicators based on information from the Statement of cash flows (and in combination with data from the other elements of the financial statement) partially helps to identify gaps and manipulations in the accounting information based on the accrual principle.
- 4). The direct method of preparing the SCF is more closely related to bank lending on the basis of assets.

The object of further research is to establish a precise quantitative correlation between the 4 financial values of the enterprises from the sample formed above (Table 2), as well as their relationship with such a leading indicator for banks (in the credit analysis of their customers), such as EBIDTA, and also the relationship between the four financial dimensions,

EBITDA and the indicators most often used by banks, related to liquidity and solvency of customers, indicated in Appendix 1.

## References

1. Alves, T. M. et al. (2015). The Role of Cash Flow Statement in The Anticipation of Bankruptcy: The Case of Banco Privado Português, *Revista Universo Contábil*, ISSN 1809-3337, FURB, Blumenau, v. 11, n. 3, p. 190-210, jul./sept., 2015
2. Beresford, D. R. (1988). The 'Balancing Act' in Setting Accounting Standards, *Accounting Horizons*, March 1988, pp. 1–7.
3. COFACE CEE TOP 500 COMPANY RANKINGS, Available from: COFACE CEE TOP 500 - 2022 EDITION / Coface CEE Top 500 / News and Publications - Coface, Last seen: 15.02.2024.
4. Eurostat (2023). Non-financial corporations - statistics on financial assets and liabilities, Available from: Non-financial corporations - statistics on financial assets and liabilities - Statistics Explained (europa.eu), last seen: 05.02.2024.
5. IAS 7 Statement of Cash Flows, Effective from: 01.01.2024. Available from: IFRS - IAS 7 Statement of Cash Flows Last seen: 12.02.2024.
6. National Accounting Standard 7 Statement of Cash Flows, *State Gazette* 86/2007.
7. Statista (2024). Composition of loans to households and non-financial corporations in the European Union in 2022, by country, Available from: <https://www.statista.com/statistics/1301197/composition-of-loans-in-the-eu-by-country/>, Last seen 10.02.2024.
8. Annual Individual Financial Statement of Advance Properties Ltd for 2019, 2020, 2021, 2022.
9. Annual Individual Financial Statement of Astra Bioplant Ltd for 2019, 2020, 2021, 2022.
10. Annual Individual Financial Statement of Kozloduy NPP EAD for 2019, 2020, 2021, 2022.
11. Annual Individual Financial Statement of Aurubis Bulgaria AD for 2019, 2020, 2021, 2022.
12. Annual Individual Financial Statement of Bulgargaz AD for 2019, 2020, 2021, 2022.
13. Annual Individual Financial Statement of Bulgarian Energy Holding EAD for 2019, 2020, 2021, 2022.
14. Annual Individual Financial Statement of Eurohold Bulgaria AD for 2019, 2020, 2021, 2022.
15. Annual Individual Financial Statement of Lukoil Bulgaria Ltd for 2019, 2020, 2021, 2022.
16. Annual Individual Financial Statement of National Electric Company for 2019, 2020, 2021, 2022.
17. Annual Individual Financial Statement of Saksa Ltd for 2019, 2020, 2021, 2022.

## Appendix 1

### Specific financial indicators, based on Cash flow statement

Indicators	Enterprise 1					Enterprise 2					Enterprise 3					Enterprise 4				
	2019	2020	2021	2022	2019	2020	2021	2022	2019	2020	2021	2022	2019	2020	2021	2022	2019	2020	2021	2022
CCF	70551	41376	309099	264249	747946	679662	1092776	954256	14071	41667	242683	86538	337409	321820	383818	601658				
CCF as % of sales	38.41%	77.02%	94.32%	84.69%	56.50%	53.59%	40.28%	15.78%	0.79%	2.48%	7.54%	1.19%	7.12%	6.08%	5.85%	7.69%				
Current Creditworthiness	1.68	2.17	1.80	0.10	0.09	0.05	0.02	0.00	0.00	6.40	2.27	0.44	2.51	1.09	1.57	1.62	0.85			
Net Current Creditworthiness	1.68	1.80	0.06	0.08	-0.48	-0.70	-0.84	-0.77	6.17	1.83	-0.08	0.99	0.95	1.36	1.50	0.82				
Free Cash Flow (FCF)	1.22	1.33	0.08	0.07	0.05	0.02	0.00	0.00	4.69	1.75	0.28	0.49	0.24	0.23	0.18	0.10				
Current creditworthiness based on FCF	-	61015.00	335312.00	163891.00	-	506662.00	1595715.00	722046.00	-	-2656.00	105384.00	10469.00	-	118415.00	168454.00	398651.00				
Opening Cash Flow (OCF)	0.00	44353.00	242678.00	167037.00	0	608059.00	1205344.00	831600.00	0.00	14727.00	115767.00	-31063.00	0.00	193818.00	261267.00	477831.00				
Current creditworthiness based on OCF	5.43	6.76	48.75	51.52	371.32	448.53	527.52	135.27	3.46	8.53	35.61	8.18	75.52	87.16	85.07	96.29				
Interest coverage ratio																				

\* The order of the enterprises does not match the order in table 1 in the paper above.

## Challenges for Auditors in Relation to Disclosure of Sustainability Information in Financial Statements

**Iavor Bachev**

University of National and World Economy, Sofia, Bulgaria

*ybashev@unwe.bg*

**Abstract.** In recent years, there have been global efforts to improve the role of sustainable development as a measure for controlling global warming. In June 2020, Regulation (EU) 2020/852 was published, which sets out six environmental objectives for businesses: climate change mitigation; the process of stabilizing the average global temperature below 2°C by 2050 and making efforts to limit it to 1.5°; adaptation to climate change; sustainable use and protection of water and marine resources; transition to a circular economy; pollution prevention and control; and biodiversity protection and restoration. The International Sustainability Standards Board (ISSB) was founded in November 2021, and its primary function is to produce sustainability disclosure standards. Two years later, the board agreed and adopted the first worldwide standards for sustainability disclosure, IFRS S1 and IFRS S2. On 5.01.2023, Directive (EU) 2022/2464 on corporate sustainability reporting entered into force, which allows users of information to identify environmental risks and opportunities that are material to enterprises. The paper seeks to identify areas of concern regarding the regulation and auditing of sustainability information. The following tasks are set: outline the requirements of Directive (EU) 2022/2464 and the changes it involves in the accounting directive; define the specifics of the standardization in the Republic of Bulgaria, as well as future auditor requirements.

**Keywords:** Sustainability, Sustainability standards, Auditor, Materiality

### 1. Introduction

Directive (EU) 2022/2464 went into effect in 2023, and it outlines rules for disclosing information about sustainability issues. The fundamental reason for the shift is the inadequate comparability and trustworthiness of the sustainability information that has been released thus far, as well as the asymmetry between stakeholders' information demands and the information presented [1].

Directive (EU) 2022/2464 [2] makes modifications to Directive 2013/34/EU relating to sustainability reporting. A new clause called "Sustainability reporting" is added, which governs the duty to publish information about the impact of the undertaking on sustainability issues [§1 Art. 19a]. This requirement applies to large, small, and medium-sized businesses that work in the public good. To effect change, they should also give consumers with knowledge that will allow them to understand how sustainability issues affect the development, performance, and health of the firm. This information will be included in a distinct section of the management report, in accordance with sustainability reporting standards. The activity report must be created in an electronic format for reporting.

The information that the companies will disclose relates to: a description of the business model and strategy of the undertaking and the extent to which they can minimise the risks related to sustainability issues and are consistent with the interests of stakeholders; the adopted

objectives related to sustainability issues, tied to specific deadlines and the progress achieved; competences and skills of the management staff; its role in relation to sustainability issues and incentive schemes; a description of the organizations's sustainability policies; a description of actual or potential adverse impacts related to the entity's operations and its value chain and the measures taken to mitigate those impacts; a description of the main risks and how they are managed by the entity [§2 Art.19a]; The aim is to provide consumers with information that will help them understand the group's impact on sustainability issues and how sustainability issues affect the group's health [§1 Art. 29a]. The information they are required to disclose is similar to that disclosed by entities in their individual reporting.

Separate sections entitled 'Sustainability Reporting Standards' and 'Sustainability Reporting Standards for Small and Medium-sized Enterprises' set out a reporting framework aimed at ensuring the provision of comprehensible, meaningful, verifiable and comparable information. Depending on the subject matter of the specific reporting standard, it will specify the information that the entity should disclose on: environmental, social and human rights factors, and governance factors. Small and medium-sized enterprises that are listed on the stockexchange may report according to separate standards, which will require less information to disclose.

Commission Delegated Regulation (EU) 2023/2772 of 31 July 2023 adopted European sustainability reporting standards. EFRAG is currently developing two draft standards that apply to listed SMEs and non-listed micro enterprises - the Exposure Draft ESRS for listed SMEs (ESRS LSME ED) and the Exposure Draft for the voluntary reporting standard for non-listed SMEs (VSME ED).

The financial statements of public-interest entities shall be audited by independent auditors who shall express an auditor's opinion as to whether the accounts have been prepared in accordance with the applicable financial reporting framework. In order to increase the confidence of users of information, Directive (EU) 2022/2464 introduces the requirement for mandatory assurance of sustainability reporting. The auditor or audit firm shall present its views in an assurance report on sustainability reporting. That report shall be prepared in accordance with the requirements of assurance standards adopted by the Commission.

Directive (EU) 2022/2464 allows an auditor or an audit firm, other than those carrying out the statutory audit of the financial statements, to express an opinion. The status of an "independent audit service provider" is introduced [According to Article 3 "Amendments to Directive 2006/43/EC", "independent audit service provider" means a conformity assessment body accredited in accordance with Regulation (EC) No 765/2008 of the European Parliament and of the Council for the specific conformity assessment activity referred to in point (aa) of the second subparagraph of Article 34(1) of Directive 2013/34/EU], which allows audits to be carried out by persons who have not acquired the qualification of a chartered accountant. In order to ensure the quality of the audit, these providers must meet certain conditions. Expertise in sustainability reporting and assurance on sustainability reporting, maintain this qualification, be independent, objective and ethical. To be an "audit service provider" you need to have a developed system of procedures, a quality assurance system and sufficient resources. Audit service providers that are accredited or in the process of accreditation as of 1 January 2024 are not subject to the requirements for training and assurance examinations. They will acquire the necessary knowledge of sustainability reporting and assurance through the requirement of continuing education. The rest take an examination of professional competence, guaranteeing the necessary level of theoretical knowledge and the ability to apply this knowledge in practice.

## 2. Transposition and challenges

According to Article 5 of Directive (EU) 2022/2464, Member States must transpose the changes in their legislation by July 6, 2024. In the Republic of Bulgaria, the issues related to the audit and the reporting framework are regulated by the Law on the Independent Financial Audit and the Accountancy Act. Following the established traditions of similar processes in the past, in the authors opinion it will be much more effective if this is done by changing the existing laws, and not by signing new ones.

The Accountancy Act should define the enterprises that will prepare a sustainability report as a separate section of the activity report. The law must define the content of the information to be disclosed by the enterprises/groups in the individual or consolidated sustainability report.

The Law on the Independent Financial Audit should regulate the obligations, rights and responsibilities of auditors in relation to the expression of sustainability assurance. The auditor's opinion should be formed on the basis of audit procedures allowing to conclude that there is compliance of the sustainability report with the principles, rules and requirements of the applicable sustainability reporting framework. Who will express such an opinion must take exams corresponding to the four areas defined in the Directive - the reporting framework, standards, verification and analysis procedures [Art.3 Directive (EU) 2022/2464]. It is mandatory for auditors to have theoretical knowledge of the applicable reporting framework. More precisely for:

- 1). Sustainability disclosure requirements (EU Regulatory Framework), including:
  - 1.1). Corporate sustainability reporting directive (nature and environmental objectives; concept of double materiality; differences between the non-financial reporting directive and the sustainability reporting directive; reporting requirements of the directive (individual and collective); reporting of non-EU entities);
  - 1.2). European standards for sustainability reporting (standards with common disclosure rules and requirements (ESRS1, ESRS2); standards for reporting environmental issues (ESRS E); standards for reporting social issues (ESRS S); standards for reporting on corporate governance issues (ESRS G); standards for public small and medium-sized enterprises (ESRS LSME); sectoral sustainability reporting standards; content and structure of sustainability reports) [4];
  - 1.3). Specifics of the requirements of European standards (ESRS E1 Climate change; ESRS E2 pollution; ESRS E3 Water and Marine Resources; ESRS E4 Biodiversity and Eco Systems; ESRS E5 Resource Use and Circular Economy; Social standards (ESRS S1, ESRS S2, ESRS S3, ESRS S4);
  - 1.4). Taxonomy Regulation (Taxonomy Regulation (EU) 2020/852 [5]; Delegated Regulation (EU) 2021/2178 for financial and non-financial corporations (July 2021); Delegated Regulation (EU) 1214/2022 (nuclear and gas) in force from 1 January 2023).
- 2). IFRS sustainability disclosures, including:
  - 2.1). Historical review (role and objectives of the International Sustainability Accounting Standards Board (ISSB); functions of the Climate Disclosure Standards Board (CDSB), the Value Reporting Foundation (VRF) and the Climate-Related Financial Disclosure Task Force (TCFD); standards of the Sustainability Accounting Standards Board (SASB);
  - 2.2). Overall characteristics of the Sustainability Accounting Standards Board (ISSB) (overall requirements; qualitative aspects of the material);
  - 2.3). Sustainability Disclosure Standards (IFRS S1 and IFRS S2 - Objectives, Purposes, and Requirements);

- 2.4). Issues with Sustainability Reporting (comparability of sustainability information; application of the notion of materiality; abilities and skills of accountants and auditors);
- 3). Fulfillment of assurance commitments (relevant standards and ethics; planning, execution, and completion of sustainability assurance engagements; and reporting on sustainability assurance engagements).

Another issue confronting auditors is the use of the idea of materiality and what information should be provided. The information is critical if its exclusion is expected to influence consumer decisions about purchasing, selling equity instruments, giving loans, or exercising their right to vote [1]. The first step in defining the necessary information to be given in the sustainability statement is to identify the key concerns surrounding the impacts, risks, and opportunities related to environmental, social, and governance elements. The risk here is that material hazards and opportunities are not acknowledged, or that only insignificant ones are discovered and reported.

### 3. Conclusion

The implementation of Directive (EU) 2022/2464 [2] presents a variety of issues. One of them is about defining auditors' roles, rights, and responsibilities in terms of expressing sustainability assurance. It is critical to determine who will convey this security, as well as their competencies and capabilities. Another significant challenge is determining the materiality of an enterprise's consequences, risks, and opportunities. An erroneous assessment approach reduces the value of disclosures made available to stakeholders.

### References

1. Bacheva Sn., Pozharevska R. (2023). Disclosure of Sustainability Information in Financial Statements-Regulations and Problems, Scientific Papers, Vol.5, pp.11-33) Bacheva Sn., R, Pozharevska (2023) Opovestiavane na informasia za ystoichivostta vuv finansovite otcheti-problemi I regulasii, nauchni trudove, tom 5).
2. Directive (EU) 2022/2464 of the European Parliament and of the Council of 14 December 2022 amending Regulation (EU) No 537/2014, Directive 2004/109/EC, Directs 2006/43/EC and Directive 2013/34/EU, as regards corporate sustainability reporting, (Corporate Sustainability Reporting Directive (CSRD)), (Direktiva (ES) 2022/2464 na Evropeyskia parlament i na Saveta ot 14 dekemvri 2022 za amendenie Reglament (ES) No 537/2014, Direktiva 2004/109/EO, Direktiva 2006/43/EO i Direktiva 2013/34/ES, po otnoshenie na otchi-taneto na predpriyatiyata vav vrazka s ustoychivostta, (Direktiva za otchitane na korpo-rativna ustoychivost), <https://eur-lex.europa.eu/legal-content/BG/TXT/HTML/?uri=CELEX:32022L2464>
3. Regulation (EU) 2020/852 of the European Parliament and of the Council of 18 June 2020 [3] establishing a framework to facilitate sustainable investment, and amending Regulation (EU) 2019/2088 (Text with EEA relevance), Official Journal of the European Union, L 198/13, (Reglament (ES) 2020/852 na Evropeyskia parlament i na Saveta Reglament (ES) 2020/852 na Evropeyskia parlament i na Saveta ot 18 yuni 2020 g. za sazdavane na ramka za ulesnyavane na ustoychivite investitsii i za izmenenie na Reglament (ES) 2019/2088 (tekst ot znachenie za EIP), Ofitsialen vestnik na Evropeyskia sayuz, L 198/13), <https://eur-lex.europa.eu/legal-content/bg/txt/?uri=celex%3a32020r0852> (accessed 15.08.2023)
4. ESRS1 General principles (2023), [https://www.efrag.org/Assets/Download?assetUrl=%2Fsites%2Fwebpublishing%2FSiteAssets%2FEDESRS\\_1.pdf](https://www.efrag.org/Assets/Download?assetUrl=%2Fsites%2Fwebpublishing%2FSiteAssets%2FEDESRS_1.pdf)
5. European Commission. (2023). EU Taxonomy for Sustainable Activities – Regulation (EU) 2020/852, [https://ec.europa.eu/info/business-economy-euro/banking-and-finance/sustainable-finance/eu-taxonomy-sustainable-activities\\_en](https://ec.europa.eu/info/business-economy-euro/banking-and-finance/sustainable-finance/eu-taxonomy-sustainable-activities_en)



# Analysis on the Applications of Retrieval Augmented Generation (RAG) Architecture

**Bozhidar Bahov**

University of National and World Economy, Sofia, Bulgaria

*bozhidar.bahov@unwe.bg*

**Abstract.** This paper explores the applications and implications of Retrieval-Augmented Generation (RAG) architectures in various business domains, highlighting their ability to overcome the limitations of traditional Large Language Models (LLMs). Unlike fine-tuning, which demands significant computational resources and struggles with domain-specific nuances, RAG efficiently integrates external knowledge to deliver accurate and contextually relevant responses. The paper discusses the core components of RAG systems, including vector databases and retrieval models, and showcases their practical and business.

**Keywords:** Retrieval-Augmented Generation, Large Language Models, Vector Databases, Information Retrieval, Natural Language Processing

## 1. Introduction

Large Language Models (LLMs) have revolutionized natural language processing by demonstrating remarkable capabilities across various tasks, from language generation to comprehension. However, they are not without significant limitations, such as difficulties in updating domain-specific knowledge, high computational demands for fine-tuning, and challenges in maintaining transparency and accuracy. These constraints have spurred the exploration of alternative methods to enhance the utility of LLMs, particularly in knowledge-intensive applications. One promising approach that has emerged is Retrieval-Augmented Generation (RAG), which integrates information retrieval mechanisms with generative models. Unlike traditional fine-tuning, which often requires substantial computational resources and may still fail to adequately cover all domain-specific nuances, RAG offers a more flexible and efficient solution. By retrieving relevant external knowledge and incorporating it dynamically into the generation process, RAG provides accurate, up-to-date, and contextually enriched responses. This paradigm shift not only addresses some of the inherent limitations of LLMs but also opens new avenues for more robust and scalable AI systems.

## 2. Limitations of Large Language Models

Despite the impressive capabilities across various Natural Language Processing (NLP) tasks, LLMs have several notable limitations.

### **Risk of hallucinations**

Unless fine-tuned explicitly for this, most LLMs risk producing factually incorrect output, a phenomenon known as "hallucinations." This happens because LLMs generate responses based on patterns learned from training data without a deep understanding of factual

accuracy. The tendency to hallucinate can result in the dissemination of misinformation, mainly when users depend on these models for accurate information.

#### **Contextual limitations**

It is difficult to adapt LLMs to specific contexts, resulting in a lack of niche knowledge or specialized vocabulary required for some industries or business domains. This is mainly because most LLMs are trained on publicly available data which does not have enough of that specialized knowledge.

#### **Auditability**

Another significant issue is the difficulty in tracing back and verifying the sources of the model outputs. This lack of transparency complicates the assessment of the model's reliability and the trustworthiness of the generated content and raises ethical concerns.

#### **Knowledge cutoff**

Last but not least, LLMs are limited by the static nature of their training data, which usually has a cutoff date. This means that they do not have information about new or recent developments, making their responses potentially misleading for users who are not aware of these imitations.

### **3. Addressing the limitations of LLMs**

One way to address the above limitations is through fine-tuning, which is the process of further training a pre-trained LLM on specific data or tasks to tailor its parameters to reflect the specificities of the target domain. This allows it to adapt to specific knowledge, vocabulary, or tasks, thereby enhancing its performance on specialized applications.

However, even fine-tuned models still have a number of limitations:

- **Substantial hardware and computational requirements:** The process of fine-tuning requires significant computational resources, including high-performance GPUs, which can be costly and inaccessible to many organizations. It also typically involves high levels of time and energy consumption.
- **Expertise Intensive:** Fine-tuning requires considerable expertise in machine learning and domain-specific knowledge as well as time to prepare the training data and conduct the training process.
- **Difficult to Keep Up-to-Date:** As new data and information become available, continuously updating the fine-tuned models requires frequent retraining to maintain relevance and update the knowledge cutoff to a more recent date.
- **Limited Transparency:** Fine-tuned models still lack transparency in their decision-making processes, making it difficult to trace how specific outputs are generated.
- **Persistent Hallucinations:** Despite targeted training, fine-tuned models may still produce hallucinations.

Thus, while fine-tuning enhances the applicability of LLMs to specific domains, it does not entirely eliminate their limitations and introduces new challenges, particularly regarding resource requirements and model maintenance.

### **4. Introduction to Retrieval-Augmented Generation**

The term "Retrieval-Augmented Generation" (RAG) was first introduced by Lewis et al. in their paper "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks" (2020). RAG represents a significant advancement in the field, providing a hybrid approach

that combines the strengths of retrieval-based and generation-based models. The retriever component, identifies relevant documents from a dense vector index, while the generator component (the LLM) conditions on these documents to generate outputs. This approach addresses the limitations of pre-trained language models in knowledge-intensive tasks like dynamic knowledge access, provenance, and efficient updates.

Li et al. (2022) outlined the general framework of retrieval-augmented text generation, emphasizing three key components:

- **Retrieval Sources:** Data can be retrieved from training corpora, external datasets, or large-scale unsupervised corpora.
- **Retrieval Metrics:** These include sparse-vector retrieval, dense-vector retrieval, and task-specific retrieval methods.
- **Generation Models:** Integration methods for retrieved information encompass data augmentation, attention mechanisms, and skeleton extraction.

### **Vector Databases**

Vector databases, also referred to as vector search engines or indexing systems, are specialized storage solutions designed to efficiently manage and query high-dimensional vector data. These systems capture the semantic relationships between data points through compact numerical vectors. Unlike traditional databases, which rely on SQL and are geared towards discrete data types, vector databases excel at handling continuous, high-dimensional data. This makes them particularly well-suited for machine learning applications, where tasks such as information retrieval, recommendation systems, and natural language understanding require quick identification of semantically similar items.

In the context of RAG architectures, vector databases play a crucial role. They enable the rapid retrieval of relevant document embeddings by comparing them against user query embeddings. This allows the system to efficiently identify and fetch the most pertinent information, which can then be integrated into the input for generative models. The flexibility of vector databases, along with their ability to handle evolving embeddings as models are updated or fine-tuned, adds to their utility. They support a range of indexing and search algorithms, including approximate nearest neighbor (ANN) methods, which provide a good balance between accuracy and speed. This combination of scalability, efficiency, and adaptability makes vector databases an essential component in RAG systems, allowing them to leverage extensive external knowledge sources while ensuring fast and relevant responses.

## **5. RAG Architecture**

The diagram below showcases a typical RAG application architecture, designed to combine information retrieval with generative language models for enhanced performance in knowledge-intensive tasks. The architecture consists of several components that interact in both user-activated and automated processes.

### **Workflow Overview**

- 1). A user submits a query via the UI.
- 2). The query is embedded using the Embedding Model.
- 3). The Retrieval Model searches the Vector DB for relevant document embeddings.
- 4). If similar data exists, the LLM Cache may serve the request directly, or new data is retrieved and combined with the query to form the LLM Prompt.
- 5). The LLM generates a response, which is sent back to the UI for the user to view.

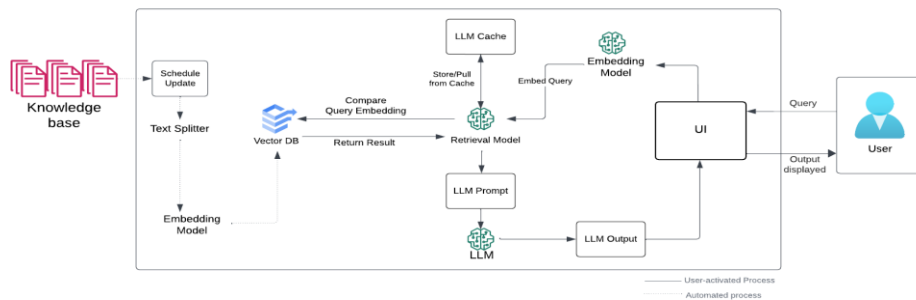


Fig. 1. An example RAG architecture of a conversational knowledge system

In more details, the elements of the architecture are as follows:

- **Knowledge Base:** could be any existing source of data, from databases to API connections.
- **Text Splitter:** divides large documents into smaller, manageable chunks. This step is crucial for creating embeddings for each chunk and ensuring relevant sections can be retrieved.
- **User Interface (UI):** serves as the entry point for users to submit queries. It also displays the output generated by the system.
- **Embedding Model:** converts textual data into vector representations, known as embeddings, which capture semantic information. The embedding model is applied to both user queries and document chunks enabling the Retrieval model to compare them.
- **Vector Database:** stores embeddings of document chunks. It enables fast retrieval of relevant documents by comparing the embeddings using similarity metrics.
- **LLM Cache:** stores previously retrieved and processed information. It aims to reduce latency by reusing relevant information from past interactions.
- **Retrieval Model:** compares the query embedding against stored document embeddings in the Vector DB.
- **Large Language Model LLM:** generates the final output based on the prompt it receives.
- **LLM Prompt:** formulates a comprehensive prompt for the language model by integrating the user's query with the retrieved information.

## 6. RAG Applications

The Retrieval-Augmented Generation architecture has emerged as a preferred solution across various business domains due to its numerous advantages. It significantly improves the accuracy and relevance of AI-generated content by accessing a broader range of information sources, both internal and external. This results in more tailored and informative responses, enhancing user satisfaction. Additionally, RAG is cost-effective, as it requires less computational power and data compared to traditional fine-tuning or large-scale training of language models. Its ability to handle large datasets and intricate queries also provides enhanced scalability, overcoming the limitations of conventional LLMs constrained by their context window size.

RAG-based conversational systems are also much better at providing information on the sources of the data used for the LLM's output due to the ability to directly point to the specific part of the Vector DB that the retriever model fetched. This level of transparency remains challenging for other approaches and LLM-based systems.

For example, in healthcare, RAG enhances the analysis of Electronic Health Records and clinical trial data, enabling more accurate and comprehensive search outcomes. It facilitates interactive data conversations, allowing non-technical users to interact with databases through natural language queries, simplifying complex data interactions. In customer support, particularly in IT and manufacturing sectors, RAG-powered chatbots provide precise and contextually relevant responses to customer inquiries, improving service quality and efficiency. In education, RAG streamlines tasks such as grading coursework or creating summarized study materials, while in finance and legal industries, it assists in drafting contracts and condensing regulatory documents, aiding data-driven decision-making processes.

However, RAG is not without its challenges. Despite its strengths, it can still produce hallucinations if the retrieval database lacks comprehensive information. As the size of the database grows, managing scalability and ensuring efficient data retrieval becomes increasingly complex. Furthermore, biases present in the retrieval database can influence the generated responses, highlighting the need for mechanisms to identify and mitigate these biases. Nonetheless, RAG represents a promising advancement in AI, offering significant benefits and a wide range of applications, albeit with certain limitations that need to be addressed.

## 7. Conclusion

In summary, while Large Language Models have significantly advanced the field of NLP, they possess inherent limitations that hinder their effectiveness in specific, knowledge-intensive tasks. Retrieval-Augmented Generation (RAG) has emerged as a compelling alternative, leveraging external knowledge bases to enhance the accuracy and relevance of generated content. This paper has explored various aspects of RAG, from its architectural components to practical applications, highlighting its advantages over traditional fine-tuning approaches.

Looking ahead, the evolution of RAG is expected to continue, with innovations such as Agentic RAG offering more sophisticated capabilities. This advanced architecture introduces an evolved, agent-based framework, capable of handling complex, multi-document question-answering tasks. The future of RAG research lies in further refining these frameworks, enhancing the autonomy and adaptability of agents, and developing systems capable of proactive, anticipatory actions. Such advancements will likely expand the scope of RAG applications, making it an indispensable tool in fields requiring nuanced information processing and decision-making. As RAG systems continue to evolve, they promise to set new standards in the integration of retrieval and generation technologies, driving the next generation of intelligent systems.

## References

1. Ji, Ziwei, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. "Survey of hallucination in natural language generation." *ACM Computing Surveys* 55, no. 12 (2023): 1-38.
2. Kucharavy, Andrei. "Fundamental Limitations of Generative LLMs." In *Large Language Models in Cybersecurity: Threats, Exposure and Mitigation*, pp. 55-64. Cham: Springer Nature Switzerland, 2024.

3. Hadi, Muhammad Usman, Rizwan Qureshi, Abbas Shah, Muhammad Irfan, Anas Zafar, Muhammad Bilal Shaikh, Naveed Akhtar, Jia Wu, and Seyedali Mirjalili. "A survey on large language models: Applications, challenges, limitations, and practical usage." *Authorea Preprints* (2023).
4. Petroni, Fabio, Aleksandra Piktus, Angela Fan, Patrick Lewis, Majid Yazdani, Nicola De Cao, James Thorne et al. "KILT: a benchmark for knowledge-intensive language tasks." *arXiv preprint arXiv:2009.02252* (2020).
5. Lewis, Patrick, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler et al. "Retrieval-augmented generation for knowledge-intensive nlp tasks." *Advances in Neural Information Processing Systems* 33 (2020): 9459-9474.
6. Li, Huayang, Yixuan Su, Deng Cai, Yan Wang, and Lema Liu. "A survey on retrieval-augmented text generation." *arXiv preprint arXiv:2202.01110* (2022).
7. Pan, James Jie, Jianguo Wang, and Guoliang Li. "Survey of vector database management systems." *The VLDB Journal* (2024): 1-25.
8. Zhu, Yinghao, Changyu Ren, Shiyun Xie, Shukai Liu, Hangyuan Ji, Zixiang Wang, Tao Sun et al. "REALM: RAG-Driven Enhancement of Multimodal Electronic Health Records Analysis via Large Language Models." *arXiv preprint arXiv:2402.07016* (2024).
9. Jauhiainen, Jussi S., and Agustín Garagorry Guerra. "Evaluating Students' Open-ended Written Responses with LLMs: Using the RAG Framework for GPT-3.5, GPT-4, Claude-3, and Mistral-Large." *arXiv preprint arXiv:2405.05444* (2024).
10. Wiratunga, Nirmalie, Ramitha Abeyratne, Lasal Jayawardena, Kyle Martin, Stewart Massie, Ikechukwu Nkisi-Orji, Ruwan Weerasinghe, Anne Liret, and Bruno Fleisch. "CBR-RAG: case-based reasoning for retrieval augmented generation in LLMs for legal question answering." In *International Conference on Case-Based Reasoning*, pp. 445-460. Cham: Springer Nature Switzerland, 2024.

# Unlocking and Exploring the Power of Unstructured Big Data Using Pentaho Business Analytics

**Mariana Kovacheva**

University of National and World Economy, Sofia, Bulgaria  
*mkovacheva@unwe.bg*

**Abstract.** This paper presents Pentaho Business Analytics capabilities for unstructured data and visualizations. It aims to provide what types of data integration are possible, the connectivity with Big Data platforms, as well as transformation and cleansing data, before using it for visualization, text analytics and ad-hoc analysis. Working with unstructured data is extremely difficult and for the user to be able to make conclusions after, and because of that when working with this type of data, there should be processes strictly followed.

**Keywords:** Big Data, Pentaho Business Analytics, unstructured data, visualizations

## 1. Introduction

In today's digital and data-driven world, companies are seeking ways to harness the vast amounts of data – semi-structured and unstructured that is generated from different sources. Semi-structured and unstructured data may include emails, documents, images, social media posts, feeds, etc. which if analysed in a correct way hold an immense potential. One of the biggest challenges in terms of integration, transformation and analysis is that semi-structured and unstructured data are very complex and have a lot of variations.

This paper probes into the capabilities of Pentaho Business Analytics for unlocking and exploring the power of unstructured Big Data. As Pentaho is one of the leading platforms in the data analysis, offers tools for integrating different types of unstructured data, connecting with different Big Data platforms and performing data transformation. These are very important steps in order to prepare the data for analytics, visualizations and for the users to be able to derive the needed insights in an effective way.

## 2. Pentaho Business Analytics functionalities

Pentaho Business Analytics is a comprehensive data integration and business analytics platform. Some of the key features are the following (Fig.1).

### 2.1. Data Integration

Pentaho Business Analytics supports ETL processes (Extract, Transform and Load). The tool helps for the data to be extracted from different sources, transforms it as per business requirements and then loads it to where it will be stored. [1]

Also, Pentaho offers integration with many data sources – relational databases as MySQL, PostgreSQL, non-relational databases as MongoDB, Cassandra, files as Excel and CSV and Big Data platforms as Hadoop. [2]

Based on the good ETL processes and connectivity, Pentaho Business Analytics also offers data blending. This means that the tool can combine data from multiple sources into a single cohesive view for analysis.

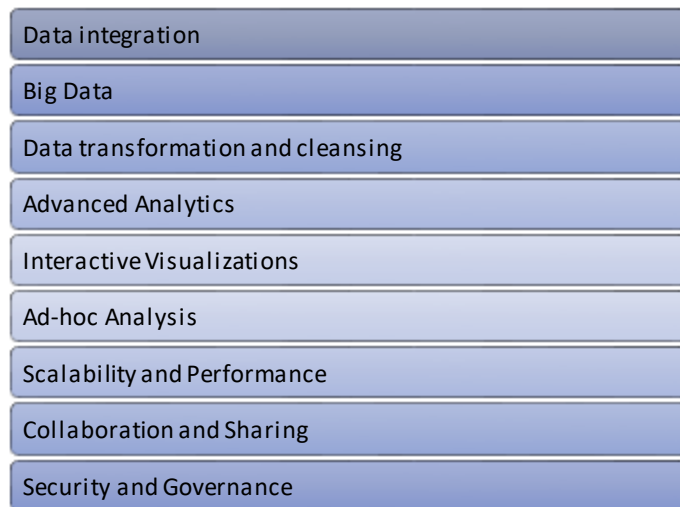


Fig. 1. Pentaho Business Analytics functionalities

## 2.2. Big Data

Pentaho Business Analytics has integration with Hadoop and its components like HDFS, MapReduce, Hive, HBase. The tool uses Apache Spark for faster, in-memory data processing and advanced analytics. Also supports real-time processing through Apache Kafka and Spark Streaming.

## 2.3. Data transformation and cleansing

Pentaho has functions for cleansing, deduplicating, validating and enriching data. Also, Pentaho Business Analytics has transformation capabilities which include filtering, aggregating, joining, splitting and calculating data.

## 2.4. Advanced Analytics

Pentaho offers different types of analytics: predictive, text and statistical analytics:

- Predictive analytics – integrations with Machine Learning (ML) libraries can be made for developing predictive models.
- Text analytics – Pentaho has capabilities for analysing unstructured data as text for performing sentiment analysis and extracting key information for the business.
- Statistical analytics – advanced statistical functions.

## 2.5. Interactive Visualizations

Pentaho Business Analytics offers dashboards, customizable reports and different types of analysis. [1] [5]

- Dashboards for interactive, real-time visualizations for key metrics and trends.
- Customizable reports – tools for creating detailed and visually appealing reports made to business needs.

## 2.6. Ad-hoc Analysis

Pentaho Business Analytics makes it easier for the users to create their own analysis with the user-friendly drag and drop interface without having any advanced technical skills. Also, the tools give opportunity for the users to create their own reports and dashboards, and to create OLAP to explore the data from the different perspectives.



### **2.7. Scalability and Performance**

Pentaho Business Analytics is designed to handle increasing volumes of data while maintaining and using the data. The tool is using in-memory computing for faster processing of data and better real-time analytics. Also offers load balancing which means that distributes the tasks that are being processed to optimize the performance of the systems that are using it. [5]

### **2.8. Collaboration and Sharing**

Pentaho has features, which make team collaboration on different projects easier. Report distribution and interactivity is supported by Pentaho Business Analytics which makes the tool more appealing to the business.

## **3. Pentaho Business Analytics, Big Data and Hadoop**

The integration of Pentaho Business Analytics with Big Data and Hadoop creates a powerful synergy, enabling organizations to manage and analyze large volumes of data efficiently and effectively. Here's how these three elements connect:

### **3.1. Pentaho Business Analytics**

Pentaho Business Analytics is a comprehensive platform that provides tools for data integration, advanced analytics, and interactive visualization. It is designed to cater to the diverse needs of data-driven organizations, offering functionalities such as ETL (Extract, Transform, Load), data blending, and real-time processing. Pentaho's intuitive interface and self-service capabilities empower users to perform ad-hoc analysis and create detailed reports and dashboards without extensive technical knowledge. [1] [6]

### **3.2. Big Data**

Big Data refers to extremely large datasets that traditional data processing systems cannot handle efficiently. These datasets are characterized by their volume, velocity, and variety. The sheer size and complexity of Big Data require specialized tools and platforms to store, process, and analyze it. With the bigger amounts of data generated every minute by every sphere, every business needs to discover the proper way of storing data depending on its type. [3]

### **3.3. Hadoop**

Pentaho's ETL tools seamlessly integrates with Hadoop, enabling the extraction, transformation, and loading of large datasets stored in HDFS while supporting connectivity with various Big Data sources. This allows organizations to blend Hadoop data with other sources like relational databases and cloud services. Leveraging Hadoop's distributed computing capabilities, Pentaho efficiently processes and analyses large-scale data, and integration with Apache Spark enhances performance through in-memory computing for faster processing and real-time analytics. Pentaho's advanced analytics tools, including predictive analytics and text mining, can be applied to Hadoop data for deeper insights, with integration to machine learning libraries such as Spark MLlib for building and deploying predictive models.

Visualization tools in Pentaho enable the creation of interactive dashboards and reports from Hadoop data, providing real-time insights and facilitating data-driven decision-making, while geospatial analysis capabilities help visualize geographical data trends. Additionally, Pentaho supports real-time data ingestion and processing via frameworks like Apache Kafka and Spark Streaming, essential for applications requiring immediate responses, such as fraud detection and real-time marketing [3] [4] [6].

## 4. Pentaho Business Analytics – advantages and disadvantages for Big Data

In this digital era there are many challenges for the companies when we are talking about Big Data and instruments that could be used to get insights from the data provided. This point provides some advantages and disadvantages of using Pentaho Business Analytics for Big Data, providing view of its capabilities and limitations.

### 4.1. Advantages



Fig. 2. Advantages of Pentaho Business Analytics for Big Data

- Comprehensive Data Integration

Pentaho Business Analytics stands out for its comprehensive data integration capabilities, which are particularly advantageous for Big Data environments. The platform's robust ETL tools facilitate the seamless extraction, transformation, and loading of large datasets from a diverse array of sources, including relational and NoSQL databases, cloud services, and Hadoop. This connectivity ensures that organizations can unify their data for comprehensive analysis [1] [2] [8].

- Advanced Analytics

Pentaho's advanced analytics features, such as predictive analytics and text analytics, empower users to develop sophisticated models and extract meaningful insights from unstructured data.

- Scalability and Performance

The scalability and performance of Pentaho Business Analytics are also significant advantages. By leveraging in-memory computing through Apache Spark, the platform ensures faster data processing and real-time analytics. Its ability to utilize Hadoop's distributed computing capabilities allows it to handle large-scale data processing tasks efficiently.

- Interactive Visualizations

Pentaho offers powerful interactive visualization tools, enabling users to create custom dashboards and detailed reports that provide real-time insights. The platform's geospatial analysis capabilities are particularly useful for visualizing and analysing geographical data trends [1].

- User-Friendly Interface

Pentaho's user-friendly interface and self-service BI capabilities enable non-technical users to perform ad-hoc analysis and create visualizations independently. The intuitive drag-and-drop functionality reduces dependency on IT and empowers business users to make data-driven decisions [6] [7].

- Collaboration and Sharing

Additionally, Pentaho fosters collaboration and sharing through shared workspaces, automated report distribution, and the creation of interactive reports that can be easily explored and customized by stakeholders [7].

## 4.2. Disadvantages



Fig. 3. Disadvantages of Pentaho Business Analytics for Big Data

### - Complexity in Setup and Configuration

Despite its many advantages, Pentaho Business Analytics also presents some challenges, particularly in the context of Big Data. The initial setup and configuration of Pentaho can be complex and time-consuming, requiring specialized knowledge and expertise. Integrating the platform with various Big Data sources and ensuring smooth data flow can also pose significant challenges, demanding substantial effort and technical proficiency [7].

### - Performance

Performance considerations are another area of concern. Large-scale data processing and advanced analytics tasks can be resource-intensive, necessitating significant computational and storage resources. While Pentaho is designed to scale, its performance can be constrained by the underlying infrastructure and the complexity of data workflows. Organizations must ensure they have the necessary resources to support Pentaho's operations effectively.

### - Learning

There is also a learning curve associated with Pentaho, especially for non-technical users. Despite its user-friendly interface, fully leveraging the platform's capabilities may require extensive training and support. Advanced features, particularly in analytics and integration, necessitate a deep understanding of the platform and related technologies.

### - Cost

The costs associated with Pentaho can be significant. Licensing, maintenance, and support expenses, coupled with the need for robust infrastructure, can be substantial, especially for small to medium-sized enterprises.

### - Limited Native Features

Finally, while Pentaho integrates with various machine learning libraries, it may not offer as many native advanced analytics features as some specialized platforms. Real-time processing capabilities, although present, might not be as advanced or efficient as those offered by dedicated real-time analytics platforms. Organizations must weigh these limitations against the platform's benefits to determine if Pentaho Business Analytics is the right fit for their Big Data needs.

## 5. Conclusion

Pentaho Business Analytics emerges as a robust platform for unlocking and exploring the potential of unstructured Big Data. Its comprehensive suite of tools facilitates seamless data integration, advanced analytics, and interactive visualizations, making it an invaluable asset for organizations navigating the complexities of Big Data. The platform's strong connectivity with Hadoop and other Big Data sources ensures efficient processing and analysis of large-scale datasets, while its user-friendly interface and self-service capabilities empower users to derive meaningful insights with ease. Despite some challenges in setup, performance,

and cost, Pentaho's strengths in data integration, scalability, and advanced analytics make it a compelling choice for organizations aiming to leverage unstructured data for strategic decision-making. By adhering to structured processes and utilizing Pentaho's capabilities, businesses can transform vast amounts of complex data into actionable intelligence, driving innovation and competitive advantage.

## References

1. 'Pentaho<sup>TM</sup> Business Analytics End-to-End Data Integration and Analytics at Enterprise Scale'. Hitachi Vantara. [Online]. Available: <https://www.hitachivantara.com/en-us/pdf/datasheet/lumada-dataops-pentaho-business-analytics-datasheet.pdf>
2. Data Integration: Ingest, Blend, Orchestrate, and Transform Data'. [Online]. Available: <https://pentaho.com/products/pentaho-data-integration/>
3. Get Started with Pentaho Data Integration and Analytics. Pentaho, Big Data, and Hadoop'. [Online]. Available: <https://docs.hitachivantara.com/r/en-us/pentaho-data-integration-and-analytics/9.5.x/mk-95pdia000/pentaho-big-data-and-hadoop>
4. What is Apache Hadoop?' [Online]. Available: <https://cloud.google.com/learn/what-is-hadoop>
5. Tarnaveanu D., 'Pentaho Business Analytics: a Business Intelligence Open Source Alternative', Database Systems Journal, vol. III, no. 3/2012, pp. 23–34.
6. Singh N. K., 'Pentaho and Big Data', Code360. [Online]. Available: <https://www.naukri.com/code360/library/pentaho-and-big-data>
7. 'Cloudera Distribution for Hadoop vs Pentaho Business Analytics comparison', PeerSpot. [Online]. Available: [https://www.peerspot.com/products/comparisons/cloudera-distribution-for-hadoop\\_vs\\_pentaho-business-analytics](https://www.peerspot.com/products/comparisons/cloudera-distribution-for-hadoop_vs_pentaho-business-analytics)
8. 'Modernizing Applications and Data for Hybrid Clouds'. Hitachi, Mar. 2023. [Online]. Available: <https://pentaho.com/wp-content/uploads/2024/05/451-Research-Modernizing-Applications-and-Data-for-Hybrid-Clouds.pdf>

# Theoretical and Practical Financial Analysis of a Real Estate Intermediary Company and the Role of Software

**Zlatan Enchev**

University of National and World Economy, Sofia, Bulgaria  
*zenchev@gmail.com*

**Abstract.** The current research examines the nature of financial analysis of real estate companies, including real estate agencies, facility companies, and other similar businesses. A special emphasis is placed on the DuPont system and the possibilities of building analysis systems based on this model, in compliance with the legislation in the field of accounting in Bulgaria. The proposed analysis systems include an example of an approval conducted concerning one of the largest real estate agencies in the country – Address Real Estate Ltd. In addition, we pose the question of the contribution of the software in the implementation of the financial analysis. The use of software greatly assists financial analysts, i.e. its use saves time and effort in analysing relevant information. The methodology developed provides valuable information both to specialists dealing with real estates as well as to teachers and students engaged in the problems of financial analysis of real estate companies. The work presented offers practical tools and approaches that can be integrated in real business conditions to improve the financial management and analysis of companies in the real estate sector.

**Keywords:** Financial Analysis of Real Estate Companies, Application of the DuPont System, Software,

## 1. Introduction

We briefly indicate in the material presented by us the essence of the financial analysis of real estate intermediary companies, as well as the sources of information for implementing this analysis. We have also developed systems for financial analysis of real estate companies engaged in intermediary services, and these are systems consistent with the well-known DuPont analysis system and the legislation in Bulgaria regulating accounting. To put it more precisely, these are systems in which capital ratios are present, i.e. certain capitals are present in the numerators and denominators of these ratios. These ratios concern revenue, costs, profits, equity and total capital. Costs and revenue are respectively associated with a decrease and an increase in equity, and profit, as the part of income exceeding the corresponding expense, also represents an increase in equity. It is even an element of equity capital and, accordingly, one of the elements forming this capital. The proposed systems have a pyramidal structure, and the values of some ratios represent the product of the values of other ratios. This shows the relationship between ratios measuring certain aspects of the business activity of real estate companies providing intermediary services in this field.

## 2. Theoretical aspect of the financial analysis

The financial analysis of real estate companies is an analysis of their monetary relationships, i.e. their finances. This is regarded as a principle that applies to all companies, regardless of the type of economic activity they carry out. There is no legal definition of

financial analysis. In the past there was a definition for financial and accounting analysis. One of the repealed national accounting standards, namely National Accounting Standard 13 “Indicators for financial and accounting analysis of the enterprise” stated that financial analysis is “is the selection, comparison and evaluation of accounting and other information in the management process in order to examine significant relationships and trends, uncharacteristic fluctuations and processes, to determine the directions of development, to measure the influence of the factors that determine them, and to adopt optimal management decisions”. This legal definition was also present in the previously existing National Accounting Standard 29 “Indicators of financial and accounting analysis of the enterprise”. Some indicators specified in these standards had to be indicated in the appendix, which is a part of the annual financial report of companies in the real sector.

Real estate companies have monetary relations with: owners, staff, customers, suppliers, commercial banks, bondholders, the state and municipalities, etc.

The owners (they are individuals and/or legal entities) of an agency own the equity capital available to it, which can be characterised as money that they have invested in the agency and which it has used (part of its assets represent ways of use of this money).

The companies have monetary relations with staff in relation to the payment of remuneration and with relevant organisations regarding social, health and supplementary pension insurance of staff.

There are also monetary relations of the companies with their suppliers of electricity and heat energy, water, communication services, and with the customers in relation to the sales of the services.

The companies receive loans from commercial banks and then there is a monetary relationship between them and the banks. The situation is similar if the agencies receive money from bond sales, i.e. agencies have a monetary relationship with bondholders. This possibility is relevant for agencies that have adopted such a legal form as a joint-stock company and have issued bonds (joint-stock companies can issue bonds).

Real estate companies have monetary relations with the state in relation to the taxes they pay into the state budget in accordance with the Law on Corporate Income Taxation and relations with the municipalities concerning the local taxes and fees they pay into the municipal budgets in accordance with the Law on Local Taxes and Fees. All this can be summarised as follows (Fig.1).

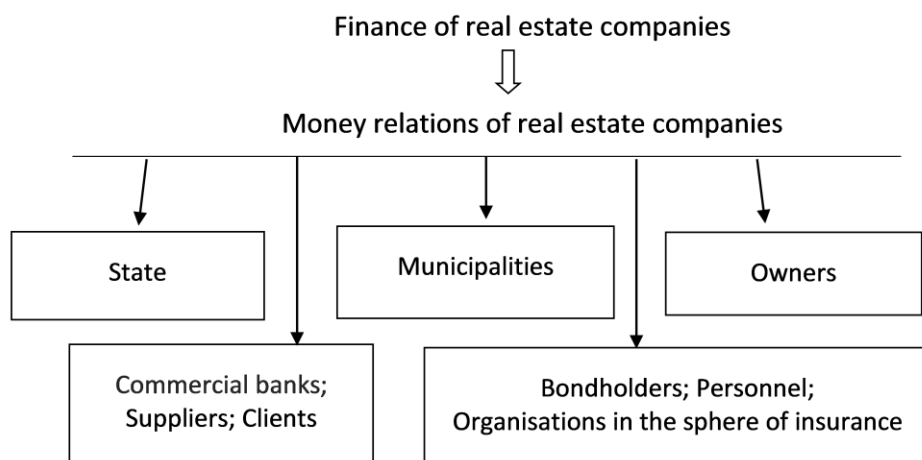
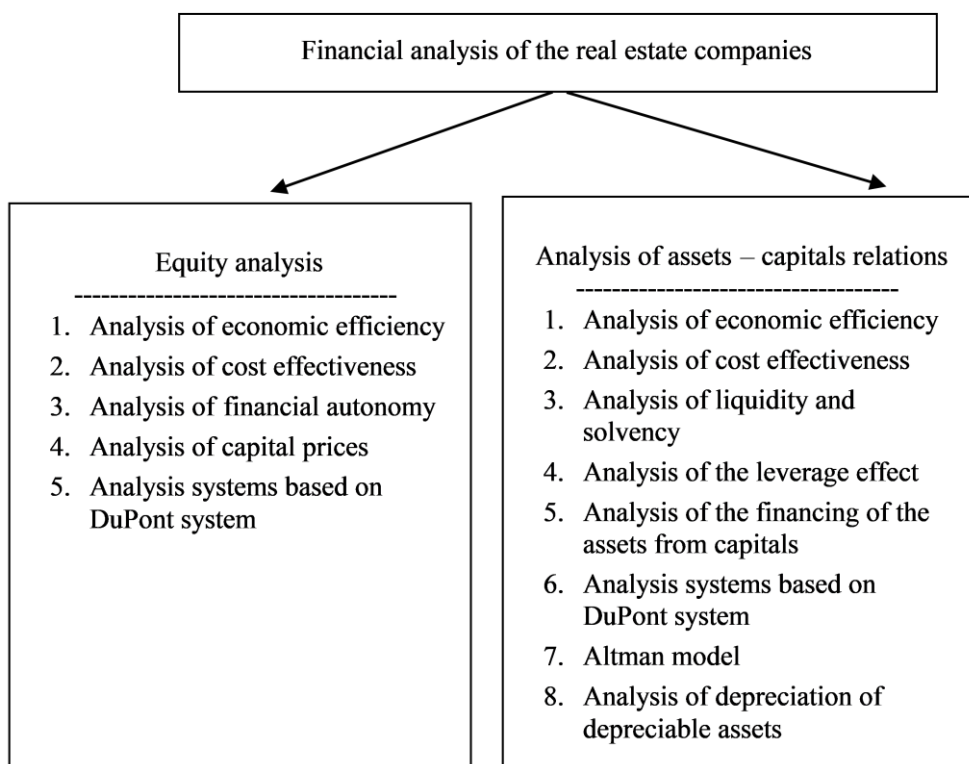


Fig.1. Finances of real estate companies

The financial analysis of real estate companies, which we will examine, covers some aspects of their activity – economic efficiency, profitability, cost effectiveness, etc. In addition, we shall look at questions related to the financial sustainability of these companies. We will try to find the relations between individual indicators of financial analysis within certain models that are based on the DuPont model. All this can be summarised in two groups – equity analysis and analysis of assets – capitals relations. This can be illustrated in the following diagram „Financial analysis of the real estate companies“ (Фиг. 2.).



Фиг. 2. Components of financial analysis of real estate companies – a dissertation study

We wish to emphasise that the present study does not cover all the issues regarding the financial analysis of real estate companies. These issues, in our opinion, can be the subject of some scientific works.

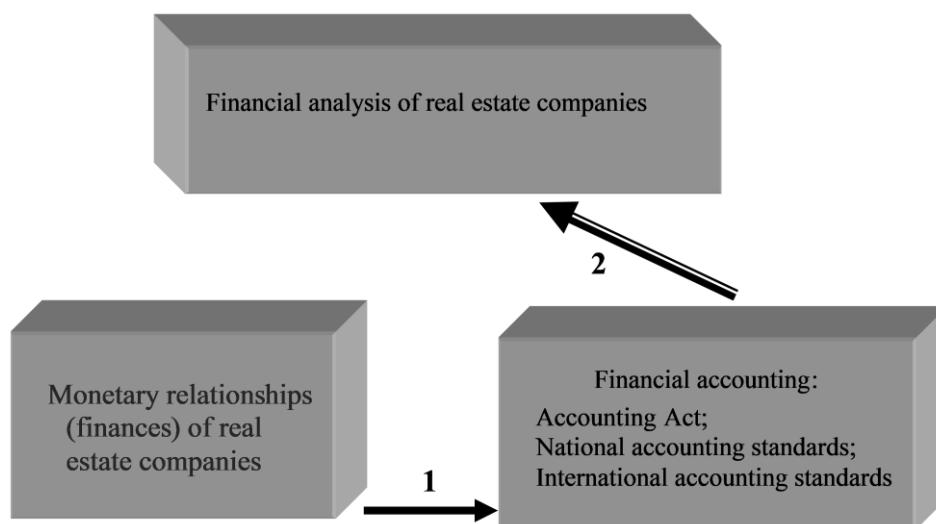
### 3. Sources of financial analysis

The monetary relationships of real estate companies are administered by their accounting departments, which in their work comply with the regulated accounting knowledge, i.e. financial accounting, which is one of the divisions of accounting science. The knowledge that we refer to is contained in the Accounting Act and also in the National Accounting Standards (NAS) and possibly in the International Accounting Standards (IAS). It is possible that some of the real estate companies apply these standards. These standards include: International Accounting Standards (IAS), International Financial Reporting

Standards (IFRS) and interpretations for their application (SIC – IFRIC interpretations). International Accounting Standards also include subsequent amendments and additions to these standards and interpretations for their application, future standards and interpretations for their application, adopted by the European Commission pursuant to Regulation (EC) No. 1606/2002 of the European Parliament and of the Council of 19 July 2002 on the application of international accounting standards.

After the changes in 1989, two accounting laws were in force in our country – one from 15 April 1991 to the end of 2001, and another – from the beginning of 2002 to the end of 2015. From the beginning of 2016, the Accounting Act entered into force, which is the current act. With regard to the National Accounting Standards, a number of changes were made.

It can be seen that there is a relationship between real estate company finance, financial accounting and financial analysis of real estate companies. We could illustrate this in the following way (Фиг.3):



Фиг. 3. Relationship between financial accounting, finance of real estate companies and financial analysis: Arrow 1 – indicates the monetary relationships of the real estate company in its accounting department based on financial accounting; Arrow 2 - Financial accounting is a source of information for carrying out a financial analysis of the real estate company.

According to the Accounting Act, companies, including real estate companies, prepare both an annual financial report as of December 31 for the reporting period (the calendar year from January 1 to December 31) and interim financial reports for periods shorter than the reporting period when required by law or by decision of the head of the company. It is possible for a real estate company that is the parent company to also prepare a consolidated financial report, but this does not happen often.

The components of the annual financial report are the balance sheet, the income and expenses statement, the statement of equity, the statement of monetary flows and an appendix. In the research, we use information from the annual financial reports of certain real estate companies. They are published in the commercial register, and access to the database constituting this register is freely available.



Financial and accounting documentation is a source of information for financial analysis, but more specifically such a source are the financial reports and accounts. Financial reports are information sources for carrying out financial analysis of companies. They include the information about the finances of the companies from the financial accounts they use (the information in the accounts can be considered as an intermediate product, and that in the reports as the final product of financial accounting). Through these accounts, information about the finances of the companies is also reflected, which, however, is summarised in the financial reports. Internal experts, i.e. employees of the company, can work with the information from the accounts, or it is possible for outside people to access this information in certain cases. The company has its own chart of accounts, i.e. a set of accounts with which it works (it contains sections, and the sections include groups of accounts). It is approved by the head of the company. The individual chart of accounts can be aligned to a certain extent with the National Chart of Accounts and Sample National Chart of Accounts that existed in the past (since the beginning of 2005, there is no National Chart of Accounts). We will indicate here the probable sections of an individual chart of accounts of a facility company:

- Section 1 Capital accounts,
- Section 2 Fixed asset accounts,
- Section 3 Inventory accounts,
- Section 4 Accounts for estimates,
- Section 5 Accounts for funds,
- Section 6 Expenses accounts,
- Section 7 Revenue accounts.

Accounts are known to be synthetic and analytical. While the synthetic account contains a summarised information about the reported subject, the analytical accounts that are related to this synthetic account, provide more detailed information about the subject. With the help of the analytical accounts, one can carry out more detailed analyses for the needs of management. It is possible for the analytical accounts, which are related to some synthetic accounts containing information of great importance, to have costing articles, which are the second phase of the analytical reporting (the first phase represents the analytical accounts). It is possible to organise analytical reporting to the synthetic account “Costs for main activities” by using relevant analytical accounts and the respective costing articles. Here is an example of what the costing articles to the analytical accounts of a facility company might contain:

- 1). Costs for materials for the activity,
- 2). Costs for external services,
- 3). Fuel and energy costs,
- 4). Depreciation costs,
- 5). Salary expenses,
- 6). Insurance costs,
- 7). Other costs

The facility company accounts for administration costs through a synthetic account “Administrative costs”. This account can also induce the keeping of analytical accounts, and they on their part can have costing articles. Costing articles can also be a source for financial analysis.

What has been said about the sources of information for the implementation of financial accounts can be summarised in the following diagram (Fig.4.):

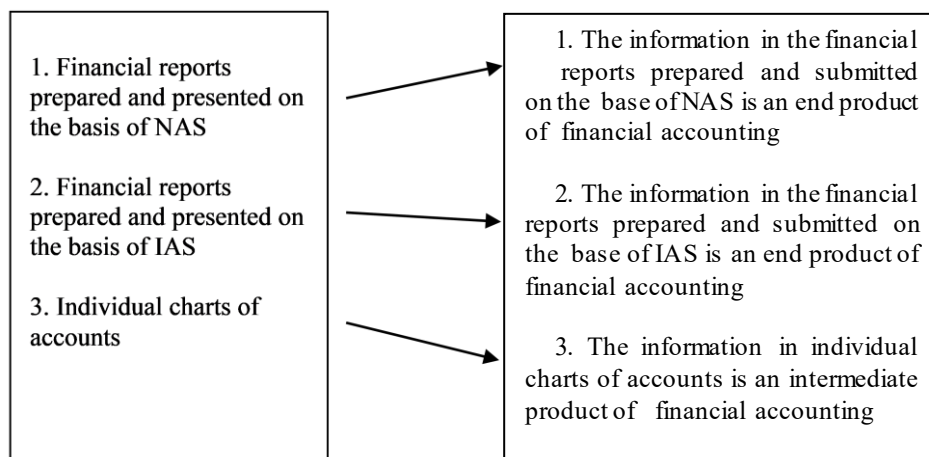


Fig. 4. Practical sources of information for financial analysis of real estate companies

Knowledge of the sources of information for financial analyses of real estate companies and solid knowledge of the implementation of these analyses are prerequisites for the implementation of qualitative analyses.

#### **4. Practical aspect of the financial analysis – analysis systems concerning the capitals, based on the DuPont system, as per the example of Address Real Estate Ltd**

##### **Origins of the DuPont System**

The DuPont system was developed in the 1920s by financial analysts at the DuPont Corporation, one of the leading chemical companies in the United States. The method was created to improve financial management and evaluate the effectiveness of various business divisions. It was named after the company and has since been widely used in various industries to analyse financial performance.

##### **Essence of the DuPont System**

The DuPont System is a method of breaking down return on equity (ROE) into three main components: net margin, total asset turnover, and financial leverage. This breakdown allows analysts to identify the main factors that influence a company's financial performance.

##### **Major Components of the DuPont System**

**Net Profit Margin:** This indicator measures a company's profitability by calculating the percentage of revenue that remains as net profit after all expenses.

$\text{Net Margin} = \text{Net Profit} / \text{Revenue}$

$\text{Net Margin} = \text{Revenue} / \text{Net Profit}$

**Total Asset Turnover:** This ratio shows how efficiently the company uses its assets to generate revenue.

$\text{Total Asset Turnover} = \text{Revenue} / \text{Total Assets}$

$\text{Total Asset Turnover} = \text{Total Assets} / \text{Revenue}$

Financial Leverage (Equity Multiplier): This ratio measures the degree of use of borrowed funds to finance the company's assets.

$$\text{Financial Leverage} = \text{Total Assets} / \text{Equity}$$

$$\text{Financial Leverage} = \text{Equity} / \text{Total Assets}$$

### **DuPont System Formula**

Return on equity (ROE) is calculated as the product of the three components:

$$\text{ROE} = \text{Net Margin} \times \text{Total Asset Turnover} \times \text{Financial Leverage}$$

### **Types of DuPont System**

Classic DuPont System: It is based on the three main components – net margin, asset turnover and financial leverage.

Advanced DuPont System: Includes additional indicators such as operating margin, current asset utilisation efficiency, and other factors that can be broken down for more detailed analysis.

### **Application of the DuPont System**

The DuPont System is applicable in a variety of industries, including real estate companies. Here are some key aspects of its application:

Estimating Profitability: Through a net margin analysis, companies can evaluate the profitability of their operations.

Asset Efficiency: Asset turnover shows how well a company is using its assets to generate revenue.

Financial Structure: Financial leverage helps understand how much the company relies on debt to finance its assets.

Identifying Problems: Breaking down of ROE allows for the identification of specific problems and areas for improvement.

Management Decisions: Helps managers make informed decisions to optimise financial strategy and improve performance.

The DuPont System is a powerful financial analysis tool that provides detailed insight into the factors affecting return on equity. It is widely applicable and can help real estate companies improve their financial management and make better strategic decisions.

We have developed two analysis systems aligned with the DuPont system. The “B” equity analysis system includes seven ratios and can be used by external analysts as it contains public data visible in the company's annual tax returns. We present this system together with the values of the ratios for 2022 (when calculating the values, we consider information contained in the annual accounting balance as of 31.12.2022 and the income and expenses statement for 2022) of Address Real Estate Ltd. Here is what the calculations show:

$$\text{Company profit} / \text{Equity} = 2\,211\,000 / 2\,373\,500 = 0.9 \quad (1)$$

$$\text{Company profit} / \text{Company revenue} = 2\,211\,000 / 8\,930\,000 = 0.2 \quad (2)$$

$$\text{Company revenue} / \text{Equity} = 8\,930\,000 / 2\,373\,500 = 3.8 \quad (3)$$

$$\text{Company profit} / \text{Company costs} = 2\,211\,000 / 6\,459\,000 = 0.3 \quad (4)$$

$$\text{Company costs} / \text{Company revenue} = 6\,459\,000 / 8\,930\,000 = 0.7 \quad (5)$$

$$\text{Total capital} / \text{Equity} = 3\,710\,000 / 2\,373\,500 = 1.6 \quad (6)$$

$$\text{Company revenue} / \text{Total capital} = 8\,930\,000 / 3\,710\,000 = 2.4 \quad (7)$$

Remark: We work with the 2022 average balance sheet values of equity and total capital. Costs include operating costs and financial costs, i.e. we work with the costs of ordinary activity. Revenue includes income from ordinary activities.

The system with the ratios and their values according to data for 2022 is shown in Fig.5.

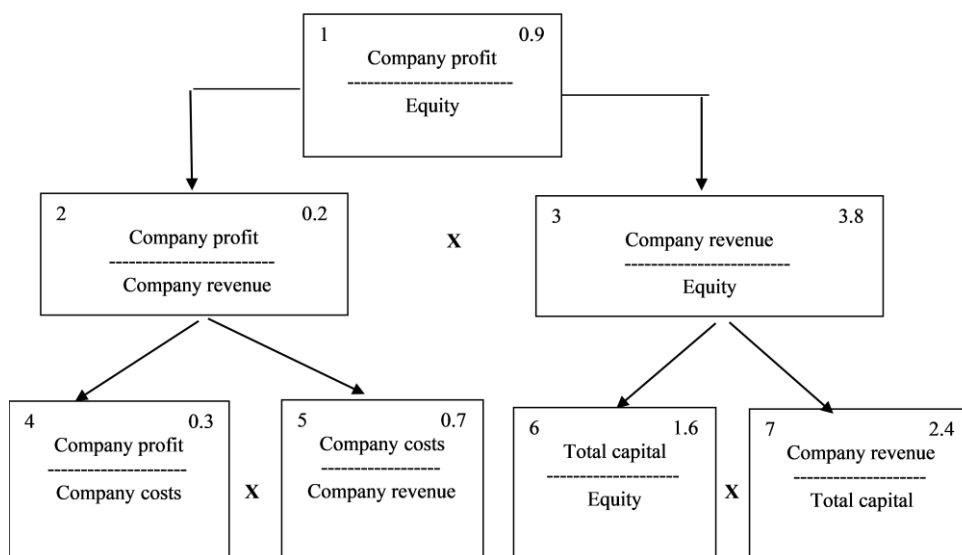


Fig. 5. Analysis system “B” including revenue and costs of the real estate company, the company’s profit, equity and total capital.

Remark: In the top left corner in the rectangles, the number of the corresponding ratio is indicated, and in the top right, its value for the corresponding year.

In Table 1 we present the values of the ratios for the years from 2018 to 2022 inclusive.

We specify that in some cases, if we multiply the values of two ratios, we get the value of the third ratio, which slightly diverges from the indicated value. This is so because the specified values were obtained by working with data from the balance sheets and income and expenses statements, and accordingly the obtained values were rounded to the first decimal place.

In the last three years of the period (2020-2022 inclusive) there is an increase- (0.6;0.8;0.9) in the value of the ratio with number 1 (Company Profit/Equity), which is a profitability ratio, and that value has the highest peak in 2022. Different ways to increase this value can be considered. And these ways concern the increase in the values of the ratios with numbers 2 (Company Profit/Company Revenue) and 3 (Company Revenue/Equity), which are ratios for profitability and economic efficiency, respectively. Expanding the network of agency offices could be considered, as well as attracting quality personnel, more advertising of the agency, reducing the prices of services, which would lead to greater turnover and will accordingly result in a greater profit, improving the quality of services using better software and looking for opportunities to reduce operating costs.

The agency can also consider options related to renting properties that are rented out, and then becoming the landlord itself, as well as buying properties that are offered for sale so that once they are purchased, they can be renovated and be sold or leased accordingly. In

addition, representatives of the agency should participate in the various forums dedicated to real estate, where the agency can promote its activity.

Table 1. The values of the ratios for the years from 2018 to 2022 inclusive

No	Ratio	2018	2019	2020	2021	2022	2018	2019	2020	2021	2022	Indicator for
	SYSTEM "B"											
1	Company profit/Equity	0,5	0,4	0,6	0,8	0,9		profitability				
2	Company profit/Company revenue	0,1	0,1	0,1	0,2	0,2		profitability				
3	Company revenue/Equity	4	5	3,9	3,9	3,8		economic efficiency				
4	Company profit/Company costs	0,2	0,1	0,2	0,3	0,3		profitability				
5	Company costs/Company revenue	0,8	0,9	0,8	0,8	0,7		profitability				
6	Total capital/Equity	2	2,9	3,1	2,1	1,6		financial autonomy				
7	Company revenue/Total capital	4	1,7	1,3	1,9	2,4		economic efficiency				
	SYSTEM "C"											
1	Operating activity profit/Operating activity costs	0,2	0,1	0,2	0,3	0,4		profitability				
2	Operating activity profit/Operating activity revenue	0,5	0,1	0,2	0,2	0,3		profitability				
3	Operating activity revenue/Operating activity costs	1,2	1,1	1,2	1,3	1,4		economic efficiency				
4	Operating activity profit/Total capital	0,3	0,3	0,2	0,4	0,7		profitability				
5	Total capital/Operating activity revenue	0,5	0,6	0,8	0,5	0,4		economic efficiency				

*Remark:* Here, 1 is a function of 3 and 4, and 2 is a function of 4 and 5. Based on the data for the five-year period from 2018 to 2022 inclusive, it is possible to track certain trends and formulate recommendations to the management of Address Real Estate Ltd.

What has been written about the first three ratios is also valid for ratios with numbers 4, 5 and 7, which are also ratios of profitability and economic efficiency. The ratio with number 6 shows how much total capital is included in one BGN of equity capital. In fact, this is a ratio for financial autonomy – for 2022, its value is 1.6, i.e. one BGN of equity capital is included in BGN 1.60 of total capital, which means that BGN 0.60 of borrowed capital is included in this total capital. The agency should strive to increase the values of profitability ratios with numbers 1, 2 and 4, and also to increase the values of economic efficiency with numbers 3 and 5. On the other hand, it should strive to reduce the value of the ratio with number 5, which is

a ratio for economic efficiency, since the economic effect is in the denominator, and the cost of obtaining it is in the numerator. The reduction of the value of this ratio means the investment of less costs by the agency in accomplishing one BGN economic effect.

In the presented system of financial analysis of real estate companies from the position of capitals, we see the dependence of some values of the ratios on others (Fig. 6.).

The value of the ratio with number 1 depends on the values of ratios numbered 2 and 3. The value of the ratio with number 2 depends on the values of ratios numbered 4 and 5. And the value of the ratio with number 3 depends on the values of ratios numbered 6 and 7 (Fig. 6.).

We also present another analysis system based on the equity (the “C” system) that includes five ratios – it can also be used by outside analysts. We approve the system in relation to Address Real Estate Ltd using information from the company’s annual accounting balances at the end of each of the years for the period 2018-2022 and the income and expenses statements for these years of this company. In the top left, in the rectangles, the number of the corresponding ratio is indicated, and in the top right, its value for the corresponding year.

Here is what the calculations show (we work with the average book value for 2022 of the total capital):

$$\text{Operating activity profit/Operating activity costs} = \quad (1)$$

$$= (8\,929\,000 - 6\,423\,000) / 6\,423\,000 = 2\,506\,000 / 6\,423\,000 = 0.4$$

$$\text{Operating activity profit/Operating activity revenue} = \quad (2)$$

$$= 2\,506\,000 / 8\,929\,000 = 0.3$$

$$\text{Operating activity revenue/Operating activity costs} = \quad (3)$$

$$= 8\,929\,000 / 6\,423\,000 = 1.4$$

$$\text{Operating activity profit/Total capital} = \quad (4)$$

$$2\,506\,000 / 3\,710\,000 = 0.7$$

$$\text{Total capital/Operating activity revenue} = \quad (5)$$

$$3\,710\,000 / 8\,929\,000 = 0.4$$

$$\text{Remark: Operating activity profit} = \text{Operating activity revenue} - \text{Operating activity costs}$$

Based on the data for the five-year period from 2018 to 2022 in Table 1, inclusive, certain trends can be traced and recommendations could be made to the management of Address Real Estate Ltd.

It is possible for analysts working in a particular real estate company to develop analysis models similar to the models proposed here that will enrich the financial analysis of that company.

In the table we present the results of approving the two analysis systems in relation to Address Real Estate Ltd for the period from 2018 to 2022 inclusive. The values of the ratios and the way of their calculation were presented, taking into account data for 2022. We also present in the table calculations of the ratio values included in the two systems, and these values were obtained when working with respective data for the previous years – from 2018 until 2021 inclusive. They are calculated by considering the data entered in the annual balance sheets at the end of each of the years in the specified period and the income and expenses statement for each of these years. We also present the analysis systems with the data included. Now we will analyse the values obtained for the period of consideration.

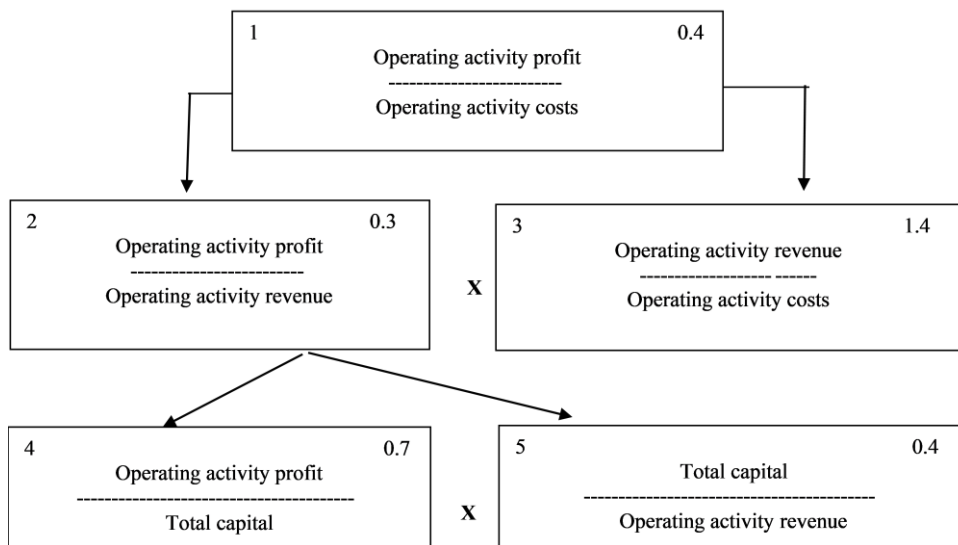


Fig. 6. Analysis system "C" includes operating activity revenue, operating activity costs, operating activity profit of real estate companies and total capital

Three of the ratios represent profitability ratios – ratios numbered 1, 2 4, and the remaining two ratios are the economic efficiency ratios.

The agency should strive for the increase of the values of the ratios, with the exception of the ratio with number 5. In relation to this ratio, we should seek to decrease its value, which would mean with less resource (in this case, a smaller average availability of equity within a given calendar year) to be engaged in achieving a slight economic effect, i.e. revenue from operational activity. The increase in the values of the other ratios would mean the attaining of greater effect when investing one BGN of costs or one BGN of equity, and in the case of a ratio with number 2, the increase in its value means that one BGN of operating activity revenue would lead to a greater profit from operating activity.

The value of ratio number 1 is the highest for 2022 when compared to the values of this ratio for the previous four years. It depends on the values of the ratios with numbers 3 and 4. The value of ratio number 2 depends on the values of the ratios with numbers 4 and 5.

When viewing the financial analysis system "B", we pointed out the opportunities for increasing profitability and economic efficiency. The same opportunities apply for optimising the values of the ratios included in the analysis system "C".

We believe that the proposed and approved financial analysis systems "B" and "C" in relation to Address Real Estate Ltd enhance the financial analysis of this agency and are useful for it. What has been said is also valid for the financial analysis system "A" from the position of the equity, which system, as we have already indicated, can only be used by analysts working in real estate companies. These systems can be used by all real estate companies.

All the calculations and comparisons we have made can be achieved faster and with less effort if quality software is used. This would mean faster work with less labour intensity compared to the situation where such software is not available.

## 5. Conclusion

Financial analysis of real estate companies by using the DuPont Advanced System provides a detailed view of various aspects of the company's financial status. This approach helps to identify the key drivers of profitability and provides valuable information for the management of the company, with the aim of improving financial results and strategic planning. We presented systems for financial analysis and approved them in relation to Address Real Estate Ltd – one of the largest real estate agencies in our country. This shows that the systems we have developed have a practical focus. They can be useful for real estate intermediary companies, but also for other companies in this field that provide services such as for example facilities and property companies (generally speaking, facilities companies carry out maintenance of the common parts of real estate, while property companies provide services in relation to individual properties, namely management and administration of residential and administrative properties). Analysts of a real estate company can track the development of ratio quantities included in financial analysis systems and identify ratio trends accordingly. In the presence of unfavourable trends, they can recommend measures that the management body could take to stop such trends. Analysts need to have in-depth knowledge in areas such as financial accounting and financial analysis. This will even allow them to develop systems for financial analysis meeting the needs of the companies in which they work. The presented research can be useful both to specialists in this practice as well as to experts in the field of scientific knowledge such as financial analysis.

## References

1. Galabov M. Financial analysis of real estate agencies, Publishing complex - UNWE, Sofia, 2013 (in Bulgarian).
2. Galabov M. Financial and accounting aspects of the activity of real estate companies, Publishing complex - UNWE, Sofia, 2016 (in Bulgarian).
3. Dimitrova R., I. Daneva, E. Kalchev, Cv. Marinova, K. Kostenarov, R. Dimitrova, I. Nikolova, N. Kanaryan, N. Nenkov. Introduction to Finance, New Bulgarian University, 2019.
4. Ivanova, R. Analysis of the financial status of the enterprise, Publishing complex - UNWE, 2015.
5. Kasarova, V. Financial analysis, publishing house - New Bulgarian University, Sofia, 2013 (in Bulgarian).
6. Milanova, E., D. Petrova, M. Nachkova, R. Ivanova, L. Todorov. Financial and accounting management, Publishing complex - UNWE, Sofia, 2018 (in Bulgarian).
7. Brealey R., St. Myers, Al. Marcus. Fundamentals of Corporate Finance, Tenth edition, McGraw-Hill Education, Ltd., 2020.
8. Mondello E. Corporate Finance: Theorie und Anwendungsbeispiele, Springer Gabler, 2022.
9. Watson D., A. Head. Corporate Finance: Principles and Practice, Eight edition, PLS, 2019.
10. Hillier D., St. Ross, R. Westerfield, J. Jaffe, Br. Jordan. Corporate Finance, 4. Edition, McGraw-Hill Education Ltd, 2020.
11. Asquith P., L. Weiss. Lessons in Corporate Finance: A Case Studies Approach to Financial Tools, Financial Policies, and Valuation, 2. Edition, John Wiley & Sons, Inc., 2019.
12. Berk J., P. Demarzo. Corporate Finance: The Core, 5. Edition, Pearson PLC, 2022.
13. Welch I. Corporate Finance, 5. Edition, IAW, 2022.
14. Ross St., R. Westerfield, J. Jaffe, B. Jordan. Corporate Finance, 13th edition, McGraw-Hill Education; 2022.