НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ
НАВЧАЛЬНО-НАУКОВИЙ КОМПЛЕКС
«ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ»
НАЦІОНАЛЬНОГО ТЕХНІЧНОГО УНІВЕРСИТЕТУ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»

СИСТЕМНІ ДОСЛІДЖЕННЯ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

МІЖНАРОДНИЙ НАУКОВО–ТЕХНІЧНИЙ ЖУРНАЛ

№ 1 2024

ЗАСНОВАНО У ЛИПНІ 2001 р.

АДРЕСА РЕДАКЦІЇ:
03056, м. Київ,
просп. Перемоги, 37, корп. 35,
ННК «ІПСА» КПІ ім. Ігоря Сікорського
Тел.: **204-81-44**; факс: **204-81-44**
E-mail: journal.iasa@gmail.com
http://journal.iasa.kpi.ua

У номері:

NATIONAL ACADEMY OF SCIENCES OF UKRAINE
EDUCATIONAL AND SCIENTIFIC COMPLEX
«INSTITUTE FOR APPLIED SYSTEM ANALYSIS»
OF THE NATIONAL TECHNICAL UNIVERSITY OF UKRAINE
«IGOR SIKORSKY KYIV POLYTECHNIC INSTITUTE»

SYSTEM RESEARCH AND INFORMATION TECHNOLOGIES

INTERNATIONAL SCIENTIFIC AND TECHNICAL JOURNAL

№ 1 2024

IT IS FOUNDED IN JULY 2001

THE EDITION ADDRESS:

03056, Kyiv,
av. Peremogy, 37, building 35,
Institute for Applied System Analysis
at the Igor Sikorsky Kyiv Polytechnic Institute
Phone: **204-81-44**; Fax: **204-81-44**
E-mail: journal.iasa@gmail.com
http://journal.iasa.kpi.ua

In the issue:

Шановні читачі!

Навчально-науковий комплекс «Інститут прикладного системного аналізу» Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» видає міжнародний науково-технічний журнал

«СИСТЕМНІ ДОСЛІДЖЕННЯ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ».

Журнал публікує праці теоретичного та прикладного характеру в широкому спектрі проблем, що стосуються системних досліджень та інформаційних технологій.

Провідні тематичні розділи журналу:

Теоретичні та прикладні проблеми і методи системного аналізу; теоретичні та прикладні проблеми інформатики; автоматизовані системи управління; прогресивні інформаційні технології, високопродуктивні комп'ютерні системи; проблеми прийняття рішень і управління в економічних, технічних, екологічних і соціальних системах; теоретичні та прикладні проблеми інтелектуальних систем підтримання прийняття рішень; проблемно і функціонально орієнтовані комп'ютерні системи та мережі; методи оптимізації, оптимальне управління і теорія ігор; математичні методи, моделі, проблеми і технології дослідження складних систем; методи аналізу та управління системами в умовах ризику і невизначеності; евристичні методи та алгоритми в системному аналізі та управлінні; нові методи в системному аналізі, інформатиці та теорії прийняття рішень; науково-методичні проблеми в освіті.

Головний редактор журналу — ректор Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», академік НАН України Михайло Захарович Згуровський.

Журнал «Системні дослідження та інформаційні технології» включено до переліку фахових видань ВАК України.

Журнал «Системні дослідження та інформаційні технології» входить до таких наукометричних баз даних: Scopus, EBSCO, Google Scholar, DOAJ, Index Copernicus, реферативна база даних «Україніка наукова», український реферативний журнал «Джерело», наукова періодика України.

Статті публікуються українською та англійською мовами.

Журнал рекомендовано передплатити. **Наш індекс 23918.** Якщо ви не встигли передплатити журнал, його можна придбати безпосередньо в редакції за адресою: 03056, м. Київ, просп. Перемоги, 37, корп. 35.

Завідувачка редакції **С.М. Шевченко**
Редакторка **Р.М. Шульженко**
Молодша редакторка **Л.О. Тарин**
Комп'ютерна верстка, дизайн **А.А. Патіохи**
Свідоцтво про реєстрацію КВ № 23234–13074 ПР від 22.03.2018 р.

Підписано до друку 29.03.2024. Формат 70×108 1/16. Папір офс. Гарнітура Times.
Спосіб друку – цифровий. Ум. друк. арк. 14,411. Обл.-вид. арк. 28,56. Наклад 77 пр. Зам. № 11/04

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Свідоцтво про державну реєстрацію: ДК № 5354 від 25.05.2017 р.
просп. Перемоги, 37, м. Київ, 03056.
ФОП Пилипенко Н.М., вул. Мічуріна, б. 2/7, м. Київ, 01014.
Виписка з Єдиного державного реєстру № 2 070 000 0000 0214697 від 17.05.2019 р.,
тел.. (044) 361 78 68.

Dear Readers!

Educational and Scientific Complex «Institute for Applied System Analysis» of the National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute» is published of the international scientific and technical journal

«SYSTEM RESEARCH AND INFORMATION TECHNOLOGIES».

The Journal is printing works of a theoretical and applied character on a wide spectrum of problems, connected with system researches and information technologies.

The main thematic sections of the Journal are the following:

Theoretical and applied problems and methods of system analysis; theoretical and applied problems of computer science; automated control systems; progressive information technologies, high-efficiency computer systems; decision making and control in economic, technical, ecological and social systems; theoretical and applied problems of intellectual systems for decision making support; problem- and function-oriented computer systems and networks; methods of optimization, optimum control and theory of games; mathematical methods, models, problems and technologies for complex systems research; methods of system analysis and control in conditions of risk and uncertainty; heuristic methods and algorithms in system analysis and control; new methods in system analysis, computer science and theory of decision making; scientific and methodical problems in education.

The editor-in-chief of the Journal is rector of the National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», academician of the NASU Michael Zaharovich Zgurovsky.

The articles to be published in the Journal in Ukrainian and English languages are accepted. Information printed in the Journal is included in the Catalogue of periodicals of Ukraine.

СИСТЕМНІ ДОСЛІДЖЕННЯ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

1 • 2024

ЗМІСТ

SYSTEM RESEARCH AND INFORMATION TECHNOLOGIES

1 • 2024

CONTENT

ПРОГРЕСИВНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ, ВИСОКОПРОДУКТИВНІ КОМП'ЮТЕРНІ СИСТЕМИ

UDC 62-50
DOI: 10.20535/SRIT.2308-8893.2024.1.01

POTENTIAL APPLICATIONS OF INTERNET OF THINGS: A COMPREHENSIVE ANALYSIS

M. PUNITHA, P.M. REKHA

Internet of Things (IoT) is the amalgamation of hardware, like sensors and trackers, which monitor several parameters of the environment or physical objects, and software that processes all the data gathered by hardware. Globally, the IoT market is anticipated to reach 53.8 billion USD by 2025. This enhancing demand is due to its innate ability to automate, which drives several industries to adopt IoT. In addition, minimum memory cost, processing, and storage with an increase in Big Data (BD), cloud, and conjunction of industrial networks and the internet are the added factors for the increase in IoT development. Due to this significance, IoT has applications in numerous areas like medical management, farming, wearable technology, smart energy meters, smart cities, etc. The applications are not limited to the examples mentioned above. Considering this, existing studies have considered different applications and attempted to execute them. As different applications have been focused on by these studies, the present review intends to provide a compilation of potential applications of IoT as considered by conventional research between 2018 and 2022. The study also intends to explore the advantages and disadvantages of different IoT applications (deliberated by conventional studies) through tabular analysis. Further, this review emphasizes IoT's major key challenges and countermeasures to resolve its security issues. Finally, the study affords recommendations that will assist all IoT experts in bringing IoT products with enhanced security into the market.

Keywords: Internet of Things, automation, security, potential applications.

INTRODUCTION

The contemporary universe is experiencing silent smart evolution with enhancing technological progress touching all life aspects. Smart technologies positioned as the epi-centre of DT (Digital Transformation) possess drastic result leasing to innovation towards designing IoT (Internet of Things) as the main pillar of recent Industry 4.0 [1]. IoT is defined as the network of objects embedded with circuits, sensors, electronics, and connectivity, which permits the objects for gathering and transmitting data. In IoT, the term 'thing' might be a vehicle with in-built sensors or any manmade or natural objects for which an IP (Internet Protocol) address could be assigned by which data could be transferred on a network. Consequently, it has become easy for creating chances to directly incorporate the world into computer-oriented systems that lead to efficiency, enhancements, minimized human exertion, and economic benefits. IoT definition has aroused due to the con-

junction of several technologies like ML (Machine Learning), the internet, automation, micro-electromechanical systems, and wireless technology. This conjunction has enabled to bridge of the gap between information and operational technology permitting unstructured machine-retrieved data to remain examined for insights that will bring innovation. It permits objects to get sensed and managed remotely throughout the existing infrastructure of the network creating choices to directly integrate the physical environment into computer-oriented systems. IoT is also capable of interacting with no human intervention. Few primary IoT applications already exist in transportation, automotive industries, and healthcare. By a report produced by a business insider, nearly 24-billion IoT-based devices were used by 2020. It has also been predicted that the revenue of IoT will touch 300 billion dollars in the upcoming years which will lead to numerous jobs in various industries. This vast reach is due to the significance of IoT which includes real-time monitoring and tracking, optimal decision-making, automation, and the capability of IoT for affording sensor information and permitting communication between devices.

For instance, automatic driving technology demands huge data from several sensors embedded within vehicles. These embedded sensors gather the engine's behaviour, field data, and camera feed for enhancing the self-driving method for handling any circumstance which could happen while driving. IoT also possesses several other applications. IoT finds its application in smart grids for energy management where sensors are deployed on each customer outlet and transmission line. These sensors assist in notifying failures, and irregularities in line, realizing the behavior pattern and usage nature over time. The smart meters could also alert the customers regarding the cost of peak time and non-peak time based on which cost can be reduced. In addition, IoT is applicable in fleet management. IoT logistics have been a complicated task as goods have to be dealt with better efficiency and care. Despite transmission from one place to another, the service providers must confirm that correct conditions are maintained at the transportation time. To alleviate such manual efforts, smart sensors that can connect with IoT networks persistently monitor GPS location, container's tilt angle, shock, temperature, and humidity. Data gathered from such sensors are later processed and evaluated in a central cloud server. This information could be accessed by the logistics team from anywhere through the internet. Fleet movement could also be monitored in real-time and later conveyed to customers regarding the delivery progress. Besides, IoT finds its application in the manufacturing sector where the initial IoT adopters have altered various phases of product development. IoT (Industrial IoT) will assist in optimizing several manufacturing phases through monitoring of inventory management and supply chain, quality testing, product enhancement, etc. Further, IoT could be employed in agriculture to support researchers and farmers to discover several cost-efficient and optimized manners to enhance production. Individual agriculture stages can be improved through smart-sensor technology; automation supports to minimize manual labour. Additionally, IoT is applicable in the healthcare sector for saving the lives of individuals. Initiating from the gathering of essential data from bedside devices, accessing patient information and healthcare records throughout several departments and real-time processes in diagnosing, overall patient care could be enhanced with the execution of IoT. In addition to such useful applications of IoT, there also exist certain issues due to IoT usage [2]. Hence, the present work intends to summarize the significant challenges in IoT use and countermeasures to solve the drawbacks along with the analysis of existing works about potential applications of IoT.

Objectives

The major objectives of the present work are listed below:

- To comprehensively examine the potential IoT applications as considered by traditional works ranging from (2018–2022) to bring out the drawbacks faced by these works.
- To discuss the advantages and disadvantages of different IoT applications (considered by conventional studies) through tabular analysis.
- To emphasize the major key issues of IoT and countermeasures to resolve the security challenges of IoT along with recommendations that will assist IoT experts while designing products in the future.

Paper Organization

Section 2 explores the IoT evolution and its concepts. This is followed by IoT architecture, protocols, and significance of IoT in section 3. Subsequently, the potential applications of IoT are discussed in section 4. After this, a comparative analysis is presented in section 5. Following this, major key challenges are summarized in section 6 with the countermeasures for security problems in IoT in section 7. Finally, the entire study is concluded in section 8.

EVOLUTION OF IOT (INTERNET OF THINGS) AND ITS SIGNIFICANT CONCEPTS

The arrival of smart concepts has made the globe become completely connected. These concepts create a network of several devices. Its fundamental role involves the connection of several devices for transmitting and receiving data. Kevin Ashton was the one who invented the term “IoT” in 1999. Following this, LG established the first smart fridge in 2000. After 7 years, 1st iPhone was introduced. IoT has accomplished significant influence on the globe in its initial phase and will also persist to evolve with time. The concept of IoT has also bought numerous applications ranging from fiction to statistics permitting the 4IR (Fourth Industrial Revolution). This has caused a significant impact on social, technical, and economic aspects. Scientists have stated that probable merits attained from IoT technology will develop a predictable future where smart things sense, contemplate and act. It is a trending technology that embodies several concepts like edge computing, electronic devices, geo-location of the sensor, fog computing, etc. A basic IoT concept is the things that have aroused to encompass several device kinds from wireless sensors and RFID tags to intricate systems like consumer devices and many more basic facilities. IoT possesses diverse names which expand or refine its overall possibility. Examples include IoE (Internet of Everything-things like processes, people, data, and connection), and IIoT (Industrial IoT-explaining how IoT employs in the manufacturing and industrial sector) [3]. IoT comprises huge interconnected devices as a network. Such devices transmit and gather huge data amounts regarding how they function and describe the information stored by the devices. These devices also possess sensors embedded within them which continuously emit information about the environment along with the functionality of the devices. Thus, IoT acts as a medium for dumping all the information gathered by IoT devices.

This platform examines data completely for gathering significant information which is later sent back by the data afforded. Lastly, gathered data is shared among other devices to achieve better performance to enhance the experience of users [4]. In the past era, IoT has multiplied its attention in numerous areas. Accordingly, numerous researchers have tried to afford a glimpse of the IoT landscape. The study [5] intended to realize the evolvement of IoT and its diversified technologies, applications, services, and concepts. It has been explored that, AI (Artificial Intelligence), CC (Cloud Computing), and BDA (Big Data Analytics) have a crucial contribution as IoT has been advancing its vision of smart services by the use of connected devices. Though the notion of the IoT concept has been prevailing for a long time, numerous technologies have made this concept practical. Reliable and affordable sensors have been making this technology probable for several manufacturers. Different network IPs (Internet Protocols) have made it ease in connecting sensors to the cloud for effective transmission of data. With the progress in ML (Machine Learning), analytics, and access to huge data stored in the cloud, businesses could gain fast and easy insights. The advent of such allied technologies persists to drive IoT boundaries, the overall evolution of IoT is tabulated in Table 1.

Table 1. IoT Evolution

S. no	References	IoT-Paradigms	Ahead of 2010	2010 to 2015	2015 to 2020	After 2020
1	[6]	Network	Sensor-networks	Self-organized and self-aware networks Transparency in locating sensor network Delay tolerant network Power network and storage network Hybrid networking	Awareness of network context	Self-learning and self-restoring networks Cognitive network
2	[3]	Hardware	Some sensors and RFID tags Construction of sensors into mobile NFC (Near Field Communication) in mobile Cheap and small MEMs technology	Multi-standard and multi-protocol readers More actuators and sensors Low-cost and secure tags (For example-silent-tags)	Biochemical (smart sensors) Tiny sensors (actuators and numerous sensors)	New materials and nano-technology
3	[7]	Data-processing	Processing serial data Processing parallel data QoS (Quality of Services)	Energy, frequency, and spectrum-aware processing of data Context-adaptable data processing	Context-aware processing of data and responses	Cognitive processing and cognitive optimization
4	[8]	Algorithms and Software	Integrating relational database IoT concerned with RDBMS Event oriented platforms Sensor-middleware Sensor network-middleware Localization or proximity algorithms	Open and large-scale semantic software components Assembly algorithms Social software based on NextGen (Next Generation) IoT Enterprise applications based on NextGen IoT	Goal oriented software Problem-solving and distributed intelligence Things to Things collaborative environment	User-oriented software Invisible IoT Easy to deploy IoT Things to Human Collaboration IoT-for-all

IoT-ARCHITECTURE AND PROTOCOLS

No single agreement of IoT architecture exists which is approved universally. Varied architectures have been endorsed by different investigators. For instance, the study [9] suggests a decentralized framework relying on SDN (Software Defined Networking) integrated with block chain for IoT in a smart city. The recommended framework depends on three major SDN technologies namely mobile computing, fog computing, and edge computing for detecting the attack existence in IoT networks.

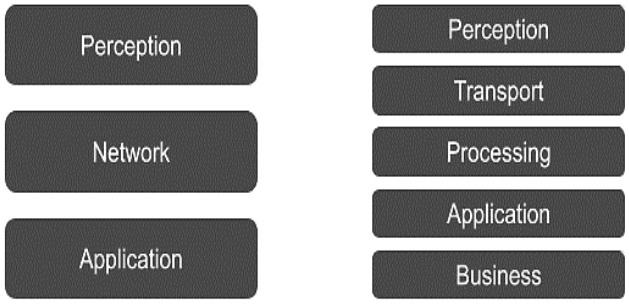
After the accomplishment of initial research on IoT, 3-layered architecture remained the central model for the applications based on IoT. These 3-layers include the perception layer, network layer, and application layer.

Perception layer: sensors stay in this layer and this is the area from where the data arrives. Data might be collected from numerous sensors on the connected device. The actuators which work on their surrounding also exist in this layer.

Network layer: this layer explores how huge data are moving in the application. It also connects several devices and transfers the data to suitable back-end services.

Application layer: it is the layer that the users view. It could be a dashboard exploring device status that is the system's part or an application for controlling a device.

The 3-layered model is a good manner of explaining an IoT-based project. However, it is restricted in possibility. Due to this reason, several proposed models possess additional or different layers and this renowned model is termed a 5-layered model. It includes the perception layer, transport layer, processing layer, application layer, and business layer. In this case, the role of the perception layer and application layer seems to be similar to 3-layered architecture. On contrary, the transport layer performs sensor data transformation from the perception layer to the processing layer and inversely through networks like 3G, Bluetooth, NFC, RFID, wireless, and LAN. Following this, the processing layer (also termed as middleware layer stores, evaluates, and processes many data which arrive from the transport layer. This layer could manage and afford diverse service sets to lower layers which apply numerous technologies like databases, CC, and BD processing modules. Lastly, the business layer maintains the entire IoT system which includes business, applications, and profit frameworks along with the privacy of users. The 3-layered and 5-layered models are shown in Figure taken from source [10].



3-layered architecture 5-layered architecture

3-layered and 5-layered architecture [10]

Further, IoT devices utilize network protocols and standards for permitting physical objects for interacting with one another and the cloud. Standards and network protocols are policies that include certain rules which explain the communication amongst many devices throughout the network. Moreover, single devices namely personal computers and smartphones also utilize network protocols to perform communication. However, general protocols which are utilized by these devices might seem to be unfit for specific necessities like latency, range, and bandwidth of solutions based on IoT. Thus, a few enhanced versions of a few prevailing protocols and new IoT protocols have arouse to satisfy the needs of IoT devices. These standards and IoT protocols are extensively categorized into two distinct divisions. This includes IoT data protocols and IoT network protocols.

IoT data protocols [11]: these are utilized for connecting IoT devices of low power. They afford communication with the hardware on the user's side without requiring an internet connection. Connectivity in these standards and protocols exists through cellular or wired networks. Some IoT data protocols are listed below.

HTTP (Hypertext Transfer Protocol). HTTP (Hypertext Transfer Protocol): is an application layered protocol that transmits hypermedia documents like HTML. It was framed for communication amongst web servers and browsers. However, it could also be utilized for several other purposes. It is a segment of the IP suite and states the services and commands utilized to transfer the data of the webpage. It exists as a basis for the exchange of any data on the web and is a client and server protocol that indicates that requests are started by the recipient (typically a web browser).

CoAP (Constrained Application Protocol). CoAP (Constrained Application Protocol): is an application layer protocol that is designed for addressing the requirements of IoT systems relying on HTTP. It is ideal for usage in devices having limited resources like WSN nodes or IoT microcontrollers.

AMQP (Advanced Message Queuing Protocol). AMQP (Advanced Message Queuing Protocol): is an open-standard application layered protocol utilized for the transaction of messages amongst the servers. It permits interoperable and encrypted messaging amongst applications and organizations. The main operation of this protocol involves: receiving and positioning messages (in queues), setting an interaction amongst the messages, and storing the messages. With its reliability and security level, it is generally applied in settings that need analytical environments based on the server like the banking sector. Nevertheless, it is not extensively utilized elsewhere.

MQTT (Message Queuing Telemetry Transport). MQTT (Message Queuing Telemetry Transport): is a lightweight protocol and features a publisher and subscriber model that permits the simple flow of data amongst varied devices. Despite the wide adoption of this protocol as an IoT standard especially with industrial applications, it seems to not assist defined device management and data representation.

DDS (Data Distribution Service). DDS (Data Distribution Service): is an ascendable IoT protocol that permits high-quality interaction in IoT. Alike MQTTT, it also operates with the publisher and subscriber model and could be installed in numerous settings from cloud to small devices which makes it ideal for embedded and real-time systems.

IoT network protocols [12]: are utilized for connecting devices to a network. Typically, these protocols are used over the internet. A few of the network protocols of IoT are listed below.

Bluetooth. It is a widely utilized protocol for communicating within a specific range. It remains a standard communication protocol of IoT which is optimal for low-powered, short-range, and low-cost wireless transmission amongst electronic devices. Besides, BLE (Bluetooth Low Energy) remains a version of Bluetooth which minimizes the consumption of power and has a crucial part in connecting IoT devices.

Zigbee. These networks are identical to Bluetooth in the aspect that, it already possesses a significant base for the user in the IoT world. Nevertheless, its specifications marginally conceal the use of Bluetooth. It consumes low power and possesses a low range of data, a long communication range, and high security. It is a simple protocol for exchanging data and is frequently executed in devices having minimum needs like sensors and microcontrollers.

RFID (Radio Frequency Identification). RFID (Radio Frequency Identification): utilizes radio waves for transmission of fewer data packets upon network within the low range. It seems to be easy for embedding the RFID chip into IoT devices. After embedding, the RFID readers could read the corresponding tags and afford information regarding the product which is attached to the tags. Common RFID applications include inventory management. Through the attachment of RFID tags to products and then connecting them with IoT devices, businesses could track the available product count.

Wi-Fi (Wireless Fidelity). It is a renowned IoT protocol utilized for connecting neighbouring devices existing within a particular range by broadcasting a signal or hotspot. Generally, this connection makes use of several radio waves which are meant in broadcasting information upon specific frequencies namely 2.4GHz, 5GHz, or 6GHz. Presently, 6GHZ is probable to evolve as the biggest novelty in the past twenty years. It remains the core of this DT (Digital Transformation) which will permit reliable and fast interactions from the next device generation.

Significance of IoT

Tracking and Monitoring in Real-Time. The potential of a web-based system for monitoring and tracking is numerous. IoT tracking affords effective means for monitoring and tracking everything from stolen goods, and shipping containers to vehicle fleets. Distinct devices could also detect alterations in climatic conditions. Multiple industries exist where IoT trackers could immensely enhance the company's efficacy. For instance: IoT devices find their significance in determining COVID-19 by following individual infected persons and taking suitable actions for reducing its spread. Through the data gathered from several tracking devices, it is possible to find the area affected with the maximum number of victims. Additionally, the individuals who are absconding from the isolation wards or clinics could also be found. It is also probable to monitor the suspects. Conclusively, with the assistance of IoT devices, the exposure rate could be effectively controlled with the enhancement of patient health [13].

Optimal Decision Making. IoT devices possess numerous sensors based on which they could gain considerable data from several sources, affording them additional information for working with the data obtained. For instance, the study

[14] has explored an integrative agricultural monitoring system through the use of IoT and smartphone application. Through the use of this system, farmers could remotely observe the farm for determining the soil's pH level, wetness duration of the leaf, humidity, temperature, and soil moisture. The system quickly evaluates the soil and weather conditions in the specific region where the plant exists and affords new insights for manipulating the decision process. For instance, the article [15] intended to outline and progress control through the sensor modes in crop areas with the management of data through web applications and smartphones. This permits manual or automatic management by users. Automatic control utilizes data from the sensors used to monitor the soil moisture for watering. Nevertheless, the user could opt for manual management of crop watering in functional mode. The system could also send notifications by LINE API for line applications. Outcomes have explored the execution to be valuable in agriculture.

Automation. The main reason behind the IoT invention is convenience. The smart devices which automate day-to-day tasks permit humans for performing other activities. Such devices lessen the workload of people. For instance, the research [16] has explored that home automation methods are moving to mechanization techniques wherein machinery equipment manages varied systems in houses with less effort from humans. It also forces the automatic management of home appliances by several technologies. A Bluetooth kit used for automating homes is cheap and versatile. However, it could be used only within a limited range. Moreover, an automation solution utilized a zigbee-RF module for creating a wireless network that permits users for remote monitoring of the appliances at home. GSM-based automation outline is also under consideration wherein consumers could monitor and manage the home appliances by transferring messages from the corresponding devices.

POTENTIAL APPLICATIONS

Generally, day-to-day applications work smart. However, they are not able to interact with one another. To permit them for communicating with one another for sharing valuable information, IoT comes into origin with wide applications. These evolving applications with autonomous abilities would certainly enhance the lives of individuals. IoT is bringing technological modifications to the daily lives of people which assists in creating and comfortable and simple life through several applications and technologies. There are innumerable IoT applications in almost all domains inclusive of industry, medicine, education, mining, transportation, governance, manufacturing, etc. Some common and recent applications of IoT are discussed in this section.

IoT based smart-city

IoT possesses good importance in constructing smart cities. It also holds positive implications in enhancing the progress of urban trade which includes the industry throughout the city with commercial progress in the city's central area. In a huge sense, realizing IoT is of huge importance to re-alter the city's industrial structure that encompasses the features of centralized use of resources and optimizing a smart city's structure. The direct influence of executing IoT is to minimize the cost of information management. As IoT relies on a network, it quickly processes, stores and transfers huge information. As a new technology, the physical network

not only creates new opportunities for the economic development of urban areas but also improves the creation of new management and production techniques. An IoT model relying on fog-computing has been endorsed in the study [17] that efficiently resolves the network scalability and BD processing issues. This model has been suggested to create efficient, harmonious, and coordinated operations of the city through several information processing, network transmission and intelligent perception means. Analytical outcomes have explored that, the recommended scheme has been suitable for fog-environment having numerous computing resources. Additionally, IoT could be utilized in several ways for making the cities highly effective ranging from the management of traffic, air pollution control, waste management, early planning for environmental disasters, creation of smart buildings, etc. For providing traffic solutions, IoT makes use of varied sensor kinds and fetches GPS location from the smartphone of drivers to find the location, vehicle speed, and number. Concurrently, traffic lights based on IoT connected with cloud platforms permit the monitoring of timings of the green light and automatically modify the lights relying on the current circumstance of traffic to avoid congestion. In addition, the use of historical data and smart solutions to manage the traffic could forecast where traffic might go and undertake measures for preventing probable congestion. Further, for monitoring air pollution, an IoT-based system has been considered for observing the quality of air upon the server through the internet. This will prompt an alarm if the quality of air exceeds a certain level indicating that, there are enough quantity of detrimental gases existing in the air namely smoke, benzene, etc. general use of IoT is managing waste through route optimization. It minimizes the consumption of fuel thereby emptying the dustbins all over the city. Moreover, IoT also assists in determining natural disasters before their occurrence. The sensors and IoT devices gather real-time information on things namely volcanic activity, barometric readings, and water levels. The sensors could detect tornadoes, earthquakes, cloudbursts, etc, and alerts through initial warnings. Thus, with its minimum energy consumption, great connectivity, low cost, and strong coverage, IoT has turned out to be a key expertise in smart-city creation. Nevertheless, faced with numerous terminals, non-uniformity in smart buildings, rational assignment of restricted resources, and integration of heterogeneous data have become significant trends in the IoT study area. Consequently, the study [18] has suggested a method to process the heterogeneous data gathered by IoT networks in smart-building thereby converting them into standardized homogeneous data which could be taken as input to monitor and manage the procedures in smart buildings to optimize its performance.

IoT-based smart home

Using IoT for home automation has become a modern lifestyle to comfort the citizens of smart cities. The person who is employing applications for home automation at the time of construction of the home could manage, monitor as well as regulate the use of energy in all probable manners. It lessens the manual work. For instance, a lamp in a bedroom glows automatically once an individual enters the room. Moreover, lights in a specific area in the home could be programmed to automatically switch on or off. Such home automation comes with several advantages: they bring safety through the appliance, enables light control, enhances awareness through cameras, and improvises convenience through automatic temperature adjustments, alarm control, smartphone alerts, energy management, etc. Some significant applications of IoT in automating homes are emphasized in the study [19]. It has been claimed that IoT manages home automation devices from

any place throughout the globe by managing them from tablets, smartphones, and personal computers. Monitoring has been another factor of IoT for affording how things could be known in advance in terms of water distribution, security alarms, and energy management. User-friendliness has been a significant factor of IoT for managing home utilizations with less interface, limited range of wireless transmission, and easy opera table through tablets and smartphones. Similarly, the research [20] has suggested a system with interconnecting sensors, data sources, and actuators to accomplish multi-purpose home automation. This system has been termed a toggle which works by holding the ability for powerful and flexible API (Application Program Interface) that indicates the basis of a common and simple communication mechanism. Most devices utilized by qToggle rely on raspberry-pi boards or ESP8266 or ESP8285. Besides, an application has also been developed which permits users for managing the sequence of sensors and home appliances. In such cases, qToggle has been utilized for several purposes like controlling temperature and lights, security alarms, garden sprinklers, concurrent opening and closing of doors, and observing energy and power. Suggested qToggle has been flexible, and user-friendly and could be developed through the use of varied devices.

IoT for smart-energy

IoT has become a boundless ally in managing energy consumption and smart distribution in smart system cases. With the enhancement of IoT networks for optimal energy, a smart meter comes with additional operations namely bidirectional communication which permits the integration of networks and users, controlling smart equipment, etc. Smart meters remain the basic component of a smart grid. Moreover, meters utilized with a management system could be used to monitor and control the appliances of the home and other devices by the requirements of users. In a contemporary smart home, IoT and smart meters have been hugely deployed for altering the conventional analog meters which digitalize the meter readings and data gathered. Data could be transmitted wirelessly which significantly lessens the manual efforts. Nevertheless, the smart home community has been susceptible to the theft of energy. These attacks could not be detected effectively as the conventional methods need specific device installation to make it work. This levies an issue for such theft detection to be executed despite the deficiency in devices for energy monitoring. To resolve this, the study [21] has developed SETS (Smart Energy Theft System) which relies on statistical models and ML (Machine Learning). Three stages of decision taking modules exist. The initial stage involves the prediction framework that utilizes a multi-model prediction system. The such system integrates several ML models into a forecast system to determine the rate of power consumption. This has been followed by the primary decision-making model which makes use of SMA (Simple Moving Average) to filter the abnormality. The final phase includes a secondary decision-making framework which creates the final decision stage on energy stealing. Simulation outcomes have represented that; the endorsed system could perform successful detection at a rate of 99.6% which improves IoT security in a smart home for saving energy. It has also been found that SETS has improved IoT security from energy theft and could be further executed in industrial and commercial sectors. The solution of the single integrated system has to be economical and effective. Smart computation systems permit the monitoring of energy consumption thereby af-

fording valuable information regarding the quality of energy. The information afforded by such systems has been utilized by operators for enhancing the energy supply. Varied methodologies could also be employed on this end like charge scheduling, demand side management, and non-intrusive monitoring of load. The intention of the research [22] has been to design, construct, explore and validate the solution for a cheap smart meter to monitor energy consumption based on IoT. It transfers the gathered data by wireless communication utilizing IoT protocols. The collected data by IoT middleware has been capable of managing and affording users with information for energy use on the internet. The smart meter operates online where all the data has been attained in real-time. For easy integration with any tracking software solution, the meter possesses a multi-protocol link. Zigbee, 6LoWPAN, and Bluetooth have been permitted for this process. Wi-Fi has also been utilized for validating the ability of the smart meter to interact with IoT middleware. Lastly, the solution has been validated in real-world settings and is also used recently.

IoT in medical management

The influence of IoT on the medical industry has attained extensive attention in recent years. Digitalization has been rapidly occurring in hospitals. Numerous firms have been developing platforms to connect numerous devices in clinics. For instance, Philip's health suite (an open platform) permits medical devices for sharing data with the specific platform which could later process and evaluate these data which could then be generated by medical workers inclusive of nurses and physicians. This process assists the physicians in taking decisions. An eICU program has also been developed by Philips, which integrates audio technology with visual technology in addition to data visualization and predictive analytics. This lays a centrally observed intense care in clinics by use of connected devices which afford data in real-time [23]. Another research area in IoT in healthcare involves the networked data produced and retrieved from healthcare devices in serious care settings, analysis and observance of patients in the clinics from X-ray, MRI, and CT scanners and mammography with integrated EMR (Electronic Medical Records) with imaging results that could assist in fast medical decisions. Moreover, digital pathology represents a term utilized for explaining actionable information produced using AI (Artificial Intelligence) algorithms on diseased tissue images including tumours, other diseases, and wounds. Various primary MedTech companies like Philips have created products of digital pathology for the market. Likewise, other global organizations like Siemens possess products of digital pathology [24]. Another evolving sector where IoT has initiated to make an impact includes prosthetics and the implantation of medical devices like defibrillators, robotic surgeries, hip joints, etc. Measurement of heart rate, computation of intake or calorie burn, automatic patient monitoring, etc. includes a few tasks accomplished by the IoT devices integrated with healthcare sensors. IoT with fig-computing, mobile edge-computing, and cloud computing exists as promising technologies to build a digital, smart, and advanced medical management system. An enhanced block chain framework based on IoT has been suggested in the study [25] to access and maintain EMR with reliability, efficiency, transparency, and security. In this case, block chain resolves the privacy limitations of IoT through the use of cryptographic algorithms. Reliability issues have also been focussed through the use of tamper-resistant ledgers. Similarly, the research [26] has used SHA (Saskatchewan Health Authority) which includes

several medical areas as case-study. The concept of the study has been to execute IoT in SHA. It has been assumed that this will be sufficient in consolidating to assure the interoperability and interconnectivity among medical areas through network designs which will enable wide communication in a health area. It has been argued that using IoT will give huge merits like enhance workforce productivity, enhanced business models, and cost savings with enhanced cooperation with patients and medical practitioners in all the segments of medical delivery. The smart solution has also been accomplished by concentrating on certain medical services like cloud services, emergency services, operational services, and cancer-care services by IoT, particularly with WSNs and other devices through the full mesh-hierarchical network configuration.

IoT-wearable

IoT-based wearable technology has been associated with ubiquitous computing. It is a technology with smart devices and microcontrollers. The device could be worn on the human body as an accessory or an implant. Such wearable devices could perform numerous identical computing similar to laptops and mobile phones. Nevertheless, in certain conditions, wearable technology could perform better than handheld devices. For instance: smart-belt, smart-shoes, smart-ring, fitness trackers, smart jewellery, etc. Wearable devices based on IoT are not restricted to these instances alone. While wearable technology inclines to indicate the items that could be put on with ease and taken off with ease, there exist invasive types of concepts as in implanted device cases like smart tattoos or microchips. No matter if a device has been incorporated into a human body or is worn, the wearable device has intended to develop convenient, portable, handy, and constant free access for computers and electronics. The study [27] has presented a HAR (Human Activity Recognition) system relying on DL (Deep Learning) methods and a Wi-Fi sensor conceived to use the wearable and smart devices for recognizing the day-to-day activities of users prevailing within AAL (Ambient Assisted Living). Generally, the proposed model exploits neural networks and Wi-Fi connections to be utilized on the cloud to perform demand tasks and on embedding components or low-cost devices for regular activity recognition. By this, connection to cloud services has been vital only when any individual initiates to get monitored affording trivial training for creating an entire dataset to fit into the use case. The intention of the work has not been real-time, however, it exists as a personalized activity monitoring in the long-term for the activity accomplished during the day by old people to infer any unwanted behaviours often associated with emergency or unhealthy cases. Obtained outcomes have been positive in comparison with other research works. Furthermore, the study [28] has introduced a smartwatch and data pipeline based on the cloud to develop a user-friendly medicine intake observation system that could contribute for enhance medication adherence. The introduced smartwatch gathers sensor data through a gyroscope or accelerometer. With the suggested sensor data retrieval, preprocessing and ML algorithms, the research has accomplished a maximum F1-score of 0.977. Outcomes have revealed that spark-cluster with numerous storage, memory, and CPU could construct ML models quicker by the use of several computing resources simultaneously.

IoT in farming

IoT in farming utilizes drones, computer imaging, remote sensors, and robots integrated with persistent progress of analytical tools and ML to monitor the crops, surveying as well as field mapping with the provision of suitable data to the farmers for better management plans of a farm for saving money and time. Employing IoT in farming targets traditional farming functionalities to satisfy the enhancing demands with limited production losses. Accordingly, the research [29] has developed an algorithm based on image processing to detect and observe the infected fruits from cultivation to harvesting. To accomplish this, ANN (Artificial Neural Network) has been employed. Four tomato crop diseases have been chosen for the research. Two databases have been used for training infected images and execution of query images. Weight has been adjusted for training through the backpropagation concept. Empirical outcomes have presented the mapping and classification of respective image categories. Images have been categorized as texture, morphology, and colour. Practical execution of methodology has been accomplished using MATLAB. Morphology has afforded results at a rate of 93%. The suggested algorithm is better at determining the disease's spread. Similarly, the article [30] has executed services based on IoT for the farming sector. The main intention has been to gather data from several spots in farmland. This data later is accessed by farmers in a mobile application through a cloud platform. Data gets represented in graphical form. The mobile application also affords several beneficial services for farmers. Application users could also manage the fundamental functions of environmental, irrigation, fertilization, and soil data. These data have been automatically correlated with the invalid data filtered out from a view of evaluating the crop performance. Suggested application has also forecasted crops and recommended crops based on a specific farm. A farm could possess numerous crops in its fields. Thus, individual crops will possess varied parameters that have to be controlled. For this, a cluster has been needed that will gather data individually. To undertake this, nodes have been installed on several field areas relying on parameters. The individual node includes a sensor and raspberry pi connected with it. The sensors might be humidity sensors, soil moisture, or temperature sensor. As the soil moisture sensor has been analog, it needs an analog-to-digital converter. Data from the sensor remains in an analog form that has to be transmitted in digital format. Thus, raw data has been supplied to the analog-to-digital converter which gets converted to digital (in voltage value format). Based on this value, the percentage of soil moisture has been considered. Thus, through clustering technology, accurate decisions can be made by farmers about detecting the particular area in which soil moisture gets reduced, the particular period in which motor pumps can be turned on/off and device parameters to be managed. All this data transferred to the cloud has been stored in a cloud database. These data could be viewed by farmers after logging into their corresponding accounts. Data from the cloud has been afforded to the mobile application by which farmers could easily control several devices and maintain all the readings retrieved from sensors.

COMPARATIVE ANALYSIS – ADVANTAGES AND DISADVANTAGES

The reviewed articles have been comparatively evaluated to bring out their advantages and disadvantages. The tabular analysis is shown in Table 2.

Table 2. Tabular analysis – advantages and disadvantages

S.no	Reference	IoT applications	Objective	Outcome/ Advantages	Disa dvantages
1	[17]	Smart-city	An IoT model relying on fog-computing has been endorsed to create coordinated operations of the city through intelligent perception means	Minimize processing delay, and reduce running time and violation rate. Resource allocation maintains stability	Competition for several resources might possess a negative impact on resource allocation due to limited communication and storage resources
2	[19]	Smart home	A significant application of IoT in automating homes is emphasized in the study	The researcher has claimed that IoT manages home automation devices from any place throughout the globe by managing them from tablets, smartphones, and personal computers	IoT has been limited in privacy, confidentiality, and over-dependence on technology
3	[22]	Energy meters based on IoT	The study has aimed to design, construct, explore and validate the solution for a cheap smart meter to monitor energy consumption based on IoT	For easy integration with any tracking software solution, the meter possesses a multi-protocol link	Limited with respect to standardization of communication protocols, lack of plug & play support, and inefficiency in BD management.
4	[25]	Medical management	An enhanced block chain framework based on IoT has been suggested to access and maintain EMR with reliability, efficiency, transparency, and security	Block chain has resolved the privacy limitations of IoT through the use of cryptographic algorithms. Reliability issues have also been focussed on by considering the tamper-resistant ledgers	A highly compact system has to be considered in the future to alleviate the security issues in EMR
5	[27]	IoT wearables	The research has presented a HAR (Human Activity Recognition) system relying on DL (Deep Learning) methods and Wi-Fi sensors conceived to use wearable and smart devices for recognizing the day-to-day activities of users	Obtained results have been positive	Real-time execution has not been undertaken
6	[30]	Agriculture	The main intention has been to gather data from several spots in farmland based on IoT	Application users could also manage the fundamental functions of environmental, irrigation, fertilization, and soil data	Data has to be centralized in the future
7	[31]	Industry	The basic objective has to predict the motor vibration measurements	The working of system compared with other ML techniques has stated better results	A unified system for maintenance planning in industry has to be built

Continued table 1

S.no	Reference	IoT applications	Objective	Outcome/ Advantages	Disa dvantages
8	[32]	Industry	To develop an IoT architecture to detect motor faults	The results has projected better accuracy in monitoring induction machines using IoT system with ML algorithms	The study has to emphasize by applying the methodology in other machine types
9	[33]	Industry	The main objective of the study has to remove noise and preserving anomalies in IIoT data	The study has detected the noise in presence of significant anomalies and has distinguished the abnormal patterns and sensor noise caused by equipment failure	It is not completely automatic detection and temporary change in noise has resulted in sensor alteration

From comparative analysis, a few disadvantages faced by conventional works have been explored which include limitations in terms of privacy, confidentiality, over-dependence on technology, lack of real-time execution, data to be centralized, need of high compact system, and limited communication, lack of plug & play support, inefficiency in BD management and storage resources. Major Key challenges of IoT are discussed below,

MAJOR KEY ISSUES AND CHALLENGES OF IOT

IoT has interconnected several physical devices through the internet for exchanging data amongst them. Data is maintained in the cloud. Despite its extensive popularity, it is facing several issues as listed below [34]:

Regarding technological challenges, IoT lacks standards, and a deficiency in intelligent analysis, and connectivity.

Standards: IoT organizations lack in adopting a standard which is the reason for their deficiency in planning, implementing, and managing the IoT devices.

Deficiency in intelligent analysis: Sometimes the data gathered by the sensor might be inaccurate which might lead to wrong results.

Connectivity: With the increase in the evolution of IoT devices, connecting numerous devices have become a challenge.

Regarding societal challenges, IoT is encountering numerous issues to meet up with customer demands that change frequently. New devices also grow rapidly, hence, time is required. As the users have partial knowledge regarding IoT devices, they might get concerned when the interface seems to be complex which might lead to averting the product use. Inventing new IoT devices and integrating them with previous ones requires time and money.

IoT also faces challenges in design where battery life seems to be a limitation. Challenges are also involved in packing and including microchips with minimum power consumption and weight. Designers have to resolve the design timing issue and deliver the device to market at the correct time. In addition, IoT faces issues due to deployment concerning connectivity, data collection and processing, and capability cross-platform. In IoT, scalability has become another concern which is of 2 kinds-vertical and horizontal scalabilities. In this case, vertical

scalability indicates the removal or inclusion of computing-resources corresponding to the IoT node, while, horizontal scalability indicates the removal or inclusion of the IoT node. Though the study [35] has attempted to solve scalability issues of IoT using a cloud-based model, it still faces several challenges IoT nodes demanding enhancing services like functional scalability, data storage, privacy, security, etc. Moreover, IoT has lacked end-to-end security solutions and privacy standards which is a prevailing concern to deploy IoT. Deficiencies in encryption, scarce testing, and updating are a few instances of lack of security in IoT. All these challenges have to be solved for maintaining IoT devices reliable and secure.

COUNTERMEASURES FOR SECURITY ISSUES IN IOT

Few countermeasures exist using which IoT security challenges could be mitigated [36]. This includes data encryption, access control, certification, communication security, etc.

Data encryption strategies. Encryption involves the procedure of converting PT (Plaintext) into CT (Cipher text). The IoT network layer accepts the hop-by-hop encryption strategy for securing nodes at the network layer. In this way, information gets encrypted during transmission. However, it has to maintain PT in the individual node by encryption process and decryption process. On contrary, IoT's application layer accepts end-to-end encryption for the secure transfer of information from the sender to the receiver. By business needs, one could select any encryption approach. In addition, with security management and the exchange of keys, one could avoid attacks namely fabrication records, eavesdropping, etc.

CC (Cloud Computing). Huge data are stored in the cloud and its performance seems to be more with minimum cost. IoT could adopt CC for storing, processing, and gathering data from numerous sensor nodes that are also capable of affording 3rd party security for IoT systems.

Communication Security in IoT. IoT devices include small devices having low power that results in weak communication security. Hence, a secure and strong communication protocol is needed for accomplishing communication security.

Access Control and Certification. Through the use of PKI (Public Key Infrastructure), one could accomplish authentication with public key certification to preserve the confidentiality and authenticity of IoT. It also seems to be a secure manner of determining the parties involved in transferring information. Identifying parties could also be performed by a trusted 3rd party termed notarization. Besides, access control affords secure IoT through the limitation of device access and person or things that are not legal for IoT resource access. For correct, IoT access control, the system has to afford a certification system for security.

Recommendations

The efficient approach to secure IoT is to concentrate on fundamentals. IoT device manufacturers, architects, developers, application developers, service developers, and experience designers have to work collaboratively to bring security in the initial phases of designing and ensure that it seems to be consistent throughout the entire IoT phase. It is vital for individuals contributing to the development of IoT for including security features at the design stage of solution development for IoT. Efforts for preventing attacks involve designing security, embedding features of the firewall to integrate additional defence layers, affording encryption abili-

ties, and including capabilities of tamper detection. When manufacturers fail to completely test their IoT devices, safety and consumer trust might be at risk. Hence, it is significant to assure that security has a purpose, constructed in all the ecosystem aspects which are implementing specific IoT devices, services, or products. When constructing IoT products, the vendors must always apply optimal strategy and intend for integrity, availability, and confidentiality.

CONCLUSION

The study reviewed the potential applications of IoT from existing works that ranged from 2018 to 2022. It also discussed the IoT evolution, its concepts, architecture, and protocols. Additionally, the significance of IoT was highlighted. The advantages and disadvantages of different IoT applications (considered by conventional studies) were also discussed through tabular comparative analysis. Through this comparative assessment, a few issues in considering IoT was determined which comprised of limitations in terms of privacy, confidentiality, overdependence on technology, lack of real-time execution, data to be centralized, lack of plug & play support, inefficiency in BD management and storage resources, need of high compact system and limited communication. After this, the study summarized the major key challenges of IoT technological challenges, societal challenges, and design and deployment challenges. Issues due to scalability and security were also emphasized. Finally, the study provided countermeasures for resolving the security challenges of IoT with recommendations that will act as a guideline for researchers and IoT experts in resolving the unsolved gaps to attain a better vision in adopting IoT products with enhanced security.

REFERENCES

1. P. Balaganesh, M. Vasudevan, R. Rameswari, and N. Natarajan, “Recent Trends in IOT-Enabled Composter for Organic Wastes,” in *Sustainable Cities and Resilience*, ed: Springer, 2022, pp. 445–457.
2. P. Gokhale, O. Bhat, and S. Bhat, “Introduction to IOT,” *International Advanced Research Journal in Science, Engineering and Technology*, vol. 5, pp. 41–44, 2018.
3. N. Sharma, M. Shamkuwar, and I. Singh, “The history, present and future with IoT,” in *Internet of Things and Big Data Analytics for Smart Generation*, ed: Springer, 2019, pp. 27–51.
4. R. Román-Castro, J. López, and S. Gritzalis, “Evolution and trends in IoT security,” *Computer*, vol. 51, pp. 16–25, 2018.
5. B.K. Chae, “The evolution of the Internet of Things (IoT): A computational text analysis,” *Telecommunications Policy*, vol. 43, p. 101848, 2019.
6. J. Wang, M.K. Lim, C. Wang, and M.-L. Tseng, “The evolution of the Internet of Things (IoT) over the past 20 years,” *Computers & Industrial Engineering*, vol. 155, p. 107174, 2021.
7. R. Krishnamurthi, A. Kumar, D. Gopinathan, A. Nayyar, and B. Qureshi, “An overview of IoT sensor data processing, fusion, and analysis techniques,” *Sensors*, vol. 20, p. 6076, 2020.
8. R. Nawaratne, D. Alahakoon, D. De Silva, P. Chhetri, and N. Chilamkurti, “Self-evolving intelligent algorithms for facilitating data interoperability in IoT environments,” *Future Generation Computer Systems*, vol. 86, pp. 421–432, 2018.
9. S. Rathore, B.W. Kwon, and J.H. Park, “BlockSecIoTNet: Blockchain-based decentralized security architecture for IoT network,” *Journal of Network and Computer Applications*, vol. 143, pp. 167–177, 2019.
10. M. El-Hajj, A. Fadlallah, M. Chamoun, and A.J.S. Serhrouchni, “A survey of internet of things (IoT) authentication schemes,” *Sensors*, vol. 19, p. 1141, 2019.

11. O. Mnushka, *IOT architecture patterns and data protocols*, 2018.
12. A. Triantafyllou, P. Sarigiannidis, and T.D. Lagkas, "Network protocols, schemes, and mechanisms for internet of things (iot): Features, open challenges, and trends," *Wireless Communications and Mobile Computing*, vol. 2018, 2018. doi: 10.1155/2018/5349894.
13. S. Ketu, P.K. Mishra, "Enhanced Gaussian process regression-based forecasting model for COVID-19 outbreak and significance of IoT for its detection," *Applied Intelligence*, vol. 51, pp. 1492–1512, 2021.
14. M.A. Patil, A.C. Adamuthe, and A. Umbarkar, "Smartphone and IoT based system for integrated farm monitoring," in *Techno-Societal 2018*, ed: Springer, 2020, pp. 471–478.
15. J. Muangprathub, N. Boonnam, S. Kajornkasirat, N. Lekbangpong, A. Wanichsombat, and P. Nillaor, "IoT and agriculture data analysis for smart farm," *Computers and Electronics in Agriculture*, vol. 156, pp. 467–474, 2019.
16. G. Arun Francis, M. S. Manikandan, V. Sundar, and E. Gowtham, "Home Automation Using Iot," *Annals of the Romanian Society for Cell Biology*, pp. 9902–9908, 2021.
17. C. Zhang, "Design and application of fog computing and Internet of Things service platform for smart city," *Future Generation Computer Systems*, vol. 112, pp. 630–640, 2020.
18. R. Casado-Vara, A. Martin-del Rey, S. Affes, J. Prieto, and J. M. Corchado, "IoT network slicing on virtual layers of homogeneous data for improved algorithm operation in smart buildings," *Future Generation Computer Systems*, vol. 102, pp. 965–977, 2020.
19. T. Alsharari, S. Alresheedi, A. Fatani, and I. Maolood, "Significant role of internet of things (IoT) for designing smart home automation and privacy issues," *International Journal of Engineering & Technology*, vol. 9, pp. 515–519, 2020.
20. C. Stolojescu-Crisan, C. Crisan, and B.-P. Butunoi, "An IoT-based smart home automation system," *Sensors*, vol. 21, p. 3784, 2021.
21. W. Li, T. Logenthiran, V.-T. Phan, and W. L. Woo, "A novel smart energy theft system (SETS) for IoT-based smart home," *IEEE Internet of Things Journal*, vol. 6, pp. 5531–5539, 2019.
22. D.B. Avancini, J.J. Rodrigues, R.A. Rabêlo, A.K. Das, S. Kozlov, and P. Solic, "A snew IoT based smart energy meter for smart grids", *International Journal of Energy Reserch*, vol. 45, pp. 189–202, 2021.
23. V. Herasevich, S. Subramanian, "Tele-ICU technologies," *Critical Care Clinics*, vol. 35, pp. 427–438, 2019.
24. *Medtech and the Internet of Medical Things*. 2018. Available: https: //www2.deloitte.com/content/dam/Deloitte/global/Documents/Life-Sciences-Health-Care/gx-lshc-medtech-iomt-brochure.pdf
25. P.P. Ray, D. Dash, K. Salah, and N. Kumar, "Blockchain for IoT-based healthcare: background, consensus, platforms, and use cases," *IEEE Systems Journal*, vol. 15, pp. 85–94, 2020.
26. A. Onasanya, S. Lakkis, and M. Elshakankiri, "Implementing IoT/WSN based smart Saskatchewan healthcare system," *Wireless Networks*, vol. 25, pp. 3999–4020, 2019.
27. V. Bianchi, M. Bassoli, G. Lombardo, P. Fornacciari, M. Mordonini, and I. De Munari, "IoT wearable sensor and deep learning: An integrated approach for personalized human activity recognition in a smart home environment," *IEEE Internet of Things Journal*, vol. 6, pp. 8553–8562, 2019.
28. D. Fozoonmayeh et al., "A scalable smartwatch-based medication intake detection system using distributed machine learning," *Journal of Medical Systems*, vol. 44, pp. 1–14, 2020.
29. H. Pang, Z. Zheng, T. Zhen, and A. Sharma, "Smart farming: An approach for disease detection implementing IoT and image processing," *International Journal of Agricultural and Environmental Information Systems (IJAEIS)*, vol. 12, pp. 55–67, 2021.
30. A. Aher, J. Kasar, P. Ahuja, and V. Jadhav, "Smart agriculture using clustering and IOT," *International Research Journal of Engineering and Technology (IRJET)*, vol. 5, pp. 2395–0056, 2018.

31. G. Scalabrini Sampaio, A. R. de A. Vallim Filho, L. Santos da Silva, and L. Augusto da Silva, "Prediction of motor failure time using an artificial neural network," *Sensors*, vol. 19, p. 4342, 2019. doi: 10.3390/s19194342.
32. M.-Q. Tran, M. Elsisi, K. Mahmoud, M.-K. Liu, M. Lehtonen, and M.M. Darwish, "Experimental setup for online fault diagnosis of induction machines via promising IoT and machine learning: Towards industry 4.0 empowerment," *IEEE Access*, vol. 9, pp. 115429–115441, 2021.
33. Y. Liu, T. Dillon, W. Yu, W. Rahayu, and F. Mostafa, "Noise removal in the presence of significant anomalies for industrial IoT sensor data in manufacturing," *IEEE Internet of Things Journal*, vol. 7, pp. 7084–7096, 2020.
34. A. Čolaković, M. Hadžialić, "Internet of Things (IoT): A review of enabling technologies, challenges, and open research issues," *Computer Networks*, vol. 144, pp. 17–39, 2018.
35. C. MacGillivray, D. Reinsel, and M. Shirer*, The growth in connected iot devices is expected to generate 79.4 zb of data in 2025, According to a New IDC Forecast*. Available: https://www.businesswire.com/news/home/20190618005012/en/The-Growth-in-Connected-IoT-Devices-is-Expected-to-Generate-79.4ZB-of-Data-in-2025-According-to-a-New-IDC-Forecast
36. .A. Abdul-Ghani, D. Konstantas, "A comprehensive study of security and privacy guidelines, threats, and countermeasures: An IoT perspective," *Journal of Sensor and Actuator Networks*, vol. 8, p. 22, 2019.

Received 05.12.2022

INFORMATION ON THE ARTICLE

Punitha Mahadevappa, ORCID: 0000-0003-3567-5537, JSS Academy of Technical Education, India

Rekha Puranic Math, ORCID: 0000-0003-0866-9502, JSS Academy of Technical Education, India, e-mail: rpmresearch22@gmail.com

ПОТЕНЦІЙНЕ ЗАСТОСУВАННЯ ІНТЕРНЕТУ РЕЧЕЙ: ВСЕБІЧНИЙ АНАЛІЗ / Пуніта Махадеваппа, Ріка Пуранік Мет

Анотація. Інтернет речей (IoT) — це об'єднання апаратного забезпечення, наприклад датчиків і трекерів, які відстежують кілька параметрів середовища або фізичних об'єктів, і програмного забезпечення, яке обробляє всі дані, зібрані апаратним забезпеченням. Очікується, що глобальний ринок IoT зросте на 53,8 мільярдів доларів США до 2025 року. Такий зростальний попит пояснюється його вродженою здатністю до автоматизації, яка спонукає кілька галузей до впровадження IoT. Крім того, мінімальна вартість пам'яті, оброблення та зберігання зі збільшенням великих даних (BD), хмари та поєднання промислових мереж та Інтернету є додатковими факторами для збільшення розвитку IoT. Завдяки цій важливості IoT застосовується у багатьох сферах, таких як управління медициною, сільське господарство, носимі технології, розумний лічильник енергії, розумне місто тощо. Застосування не обмежуються наведеними прикладами. Із урахуванням цього існуючі дослідження розглядали різні програми та намагалися їх реалізувати. Оскільки ці дослідження зосереджувалися на різних додатках, метою цього огляду є надання компіляції потенційних застосувань IoT за результатами звичайних досліджень у період з 2018 по 2022 рік. Дослідження також має на меті вивчити переваги та недоліки різних IoT додатків (розглянутих традиційними дослідженнями) за допомогою табличного аналізу. Крім того, у праці наголошується на основних ключових проблемах IoT включно з контрзаходами для вирішення проблем безпеки IoT. Дослідження дає рекомендації, які допоможуть усім експертам з Інтернету речей вивести на ринок продукти Інтернету речей із підвищеною безпекою.

Ключові слова: Інтернет речей, автоматизація, безпека, потенційні застосування.

UDC 62-50
DOI: 10.20535/SRIT.2308-8893.2024.1.02

DIGITAL MEDICAL IMAGE ENCRYPTION APPROACH IN REAL-TIME APPLICATIONS

IZZ K. ABBOUD, MUAAYED F. AL-RAWI, NASIR A. AL-AWAD

Abstract. Patient information and medical imaging data are now subject to stringent data security and confidentiality standards due to the proliferation of telemedicine techniques and medical imaging instruments. Because of the problems described above, as well as the possibility of data or information being stolen, this brings up the dilemma of transmitting data on medical images via an open network. In the past, potential solutions included the utilization of methods such as information concealment and image encryption. Nevertheless, attempting to reconstruct the original image utilizing these approaches may result in complications. In the process of this paper, an algorithm for safeguarding medical images based on the pixels of interest was established. Detection of image histogram peaks for the purpose of calculating peaks in medical images pixels of interest in medical image that have had their threshold values processed. The threshold is shown by taking the average of all the peaks in the histogram. After that, a Sudoku matrix is used to assign values of interest to each of these pixels. The proposed method will be assessed by a variety of statistical procedures, and the outcomes of these analyses will be compared to previously established standards. According to the findings, the suggested method has superior security performance compared to other image encryption methods already in use.

Keywords: real time applications, medical images, encryption, security, peak detection.

INTRODUCTION

Because of increased and better investment in multimedia techniques, research pertaining to medical imaging has made great strides forward in recent years. The vital personal information that the patient wishes to keep private is included in the medical image. In order to safeguard sensitive information, medical images are often encrypted. Textual data is often encrypted via one of many standard algorithms, including Advanced Encryption Standard (AES), Data Encryption Standard (DES), International Data Encryption Algorithm (IDEA), or Triple DES. Other frequent encryption techniques include the pixels in medical images are not evenly distributed, and the data has a high resolution. There are also distinct geographical patterns. Due to the slowness of bulk data, traditional encryption is not an appropriate method for safeguarding images from digital imaging and communication systems (DICOM) in the field of medical care.

RELATED WORK

When medical professionals require more information to diagnose a patient, medical imaging provides a secondary source of information that is both vital and effective [1]. Unfortunately, the fastest way to transmit medical images (and the one that is often considered to be the most effective) is typically over open net-

works like sharing files and email. Images that have been sent in this manner are at risk of being subject to actions such as content alteration, illicit duplication, and the loss of copyright [2]. As a direct consequence of this, there has been an increase in the number of studies into medical image security that focus on image encryption and information concealment [3]. The article [4] provided an explanation of an approach that was not only simple but also effective. It included utilizing matrix multiplication to change the pixel values in an image, which resulted in a fairly straightforward process but made it very difficult for unauthorized individuals to extract the information contained in the images. A combination of a logistic map and a 3D Lorenz, both of which displayed multiple operating modes, was essentially what produced the 5-D hyper chaotic map that was mentioned in the study [5]. While one of the modes concentrates only on the pixels that are derived from images with clear text, the other option diffuses the light a total of two times in order to produce images that are secured. Because of the study that solved the security problem [6], it is now possible for online users' sensitive data to be exchanged on web apps without the users' needing to worry about their privacy being compromised. Article [7] developed a 1-D chaotic map in order to get additional security by recognizing its shortcomings. This was followed by the presentation of a modified version of the plain text attack. Paper [8] revealed the hidden data that was hiding in a portion of the image by picking an essential section of a medical image, which is something that is often done by choosing the portions of the image that are utilized more frequently. In article [9], a method for partially encrypting secret data contained in photographs was presented, with FF1 and FF3-1 serving as key components.

The sensitive information will be encrypted without leading to an increase in the file size, which might result in a loss of memory. The gray scale encryption method based on Image Region of Interest (ROI) with chaos is presented in article [10]. To begin, the portion of the ROI that has to be recognized must be done so utilizing the Sobel edge detection technique. The edges of the blocks must then be used to sort the components of the image into those that are significant and those that are not essential. Sine maps are utilized to encrypt the irrelevant area, whereas the Lorenz method is utilized to encrypt the region that contains the ROI. Paper [11] provides a self-generating area of interest (ROI) approach for watermarking applications in biological images. The most significant benefit that this approach has over others is that it is secure enough to avoid a wide variety of attacks, including those using Gaussian, median, sharpening, and wiener filters. Research [12] addressed a new approach in which he showed how to identify the ROI with perfect precision, how to prevent information leakage in the ROI section, and how to retrieve the information lossless from encryption in the transform domain. This approach was presented as part of the discussion of the new approach. Therefore, here we come across a unique lossless game theory based medical image encryption approach with optimal ROI parameters in addition to ROI concealed locations. This method was developed in this work. In order to retrieve the medical picture without losing any data, the process of encryption must first entail a transformation at the pixel level of the ROI. This is done to safeguard the loss of information contained within the medical image. The chaos-based encryption approaches covered in the article [13] make use of a variety of different encryption algorithms. The article [14] suggested an enhanced histogram shifting (HS) reversible watermark technique for medical images and others' work in order to increase the hidden capacity of the algorithm. For the purpose of embedding information utilizing the HS technique, an image should be cut up into

smaller parts. The authors of article [15] place a high priority on integrating information into texture regions via the use of HS and contrast enhancement, with the goals of enhancing contrast in texture areas and improving how the image is subjectively perceived. In the study by [16], the authors preserved patient information by using a reversible image masking strategy for HS. After that, they enhance image quality by using two parameters of linear prediction: weight and threshold. The economy and the interest rate merged. Research [17] suggests using an HS technique to look into the lossless data that high-resolution medical images conceal. Employ a strong correlation in the image's local block pixels for the purpose of rendering the smooth surface of the medical imaging anatomy. It is not difficult at all to modify the capacity and the signal-to-noise ratio (PSNR) in accordance with the block size, the partition level, and the number of embedded bits. Many of the objectives of the aforementioned approaches include ensuring that the image is protected from infringement on its copyright and minimizing the amount of distortion in the visual quality of the embedded image. Image histogram peak detection is a basic approach for digital image processing that may be used directly and efficiently for image segmentation, quality evaluation, enhancement, decrease in data, and other purposes. It is also one of the most important aspects of digital image processing.

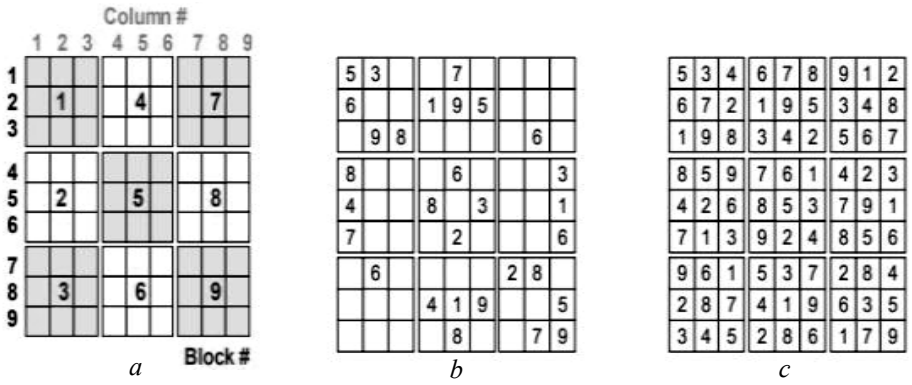


Fig. 1. A sample Sudoku puzzle and its solution: *a* — Row#, Column # and Block# notation; *b* — a sample Sudoku puzzle; *c* — the solution to Sudoku puzzle (*b*)

SUDOKU MATRIX

A Sudoku matrix is denoted by the notation $X \times X$ and may include any number between 1 and N. However, given that X is the square of the number and N equals X, each number can only appear once in each row of the matrix. Only the first value in each column and the first value in each block will be increased. The following illustration in Fig. 1 provides a sample of a Sudoku problem as well as the answer for $X = 9$. The result of successfully solving a Sudoku problem is referred to as the "Sudoku matrix".

Sudoku Typical Sudoku problems are derived from the Sudoku Matrix by omitting some cells, but each problem also includes hints on how to solve it on its own. The researchers have made an effort to come up with a number of different solutions. In this piece, we will construct a Sudoku matrix via the use of a technique called the Latin square. The downside of this rapid and systematic technique is that the set of Sudoku matrices that it generates is just a subset of the universal set of all possible Sudoku matrices. This is a limitation of the method, but it does not prevent it from being useful.

PROPOSED ENCRYPTION SELECTIVE METHOD

Fig. 2 presents the block diagram of the proposed selective image encryption method. The proposed method is comprised of a number of processes that, in order to identify and encrypt the area of interest in the medical image, are necessary. The first step is to calculate the histogram peak of the original image using the formula presented in Fig. 3. The peak detection method uses the image histogram to first create a peak detection signal. The extrema that are between the zero and zero intersections of the peak detection signal are then used in order to locate the peaks that are present in the histogram. A close approximation of the first derivative may be achieved using convolution by utilizing a differentiator. The peak may be identified in a histogram that has ideal smoothness by locating the point where the sign and zero intersection of the signal that was produced by the h and S convolutions occur. The extrema of the histogram and the location of the turning point may be estimated using the zero intersection method. The peak values of the original medical imaging are shown by the symbol "*" in Fig. 3. The threshold value for separating the relevant pixels in a medical image may be derived by taking the average of all the peak values that are acquired via the use of the histogram peak detection function. The next step is to examine each pixel in the original medical image against the predetermined threshold value; if the value is higher, the pixels must be grouped together to form a meaningful pixel block. The diffusing procedure is performed on a Sudoku matrix consisting of numerous random 16*16 grids. Perform an XOR operation on the significant pixel block using the pixels in the Sudoku matrix to generate a random encryption of the block.

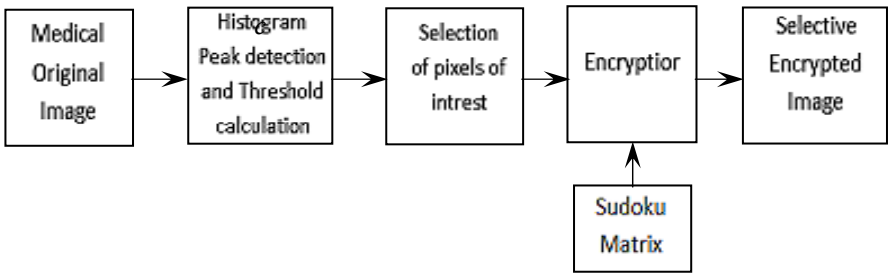
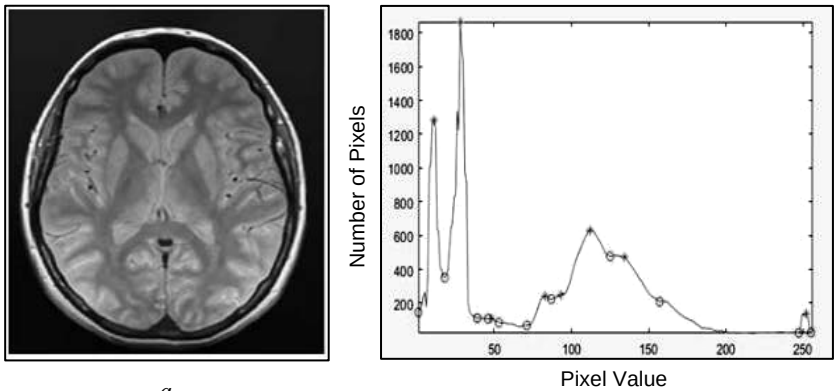


Fig. 2. The architecture of proposed visible image encryption method



a *b*

Fig. 3. Original MRI image — *a*; Peak detection using histogram — *b*

SIMULATION AND RESULT DISCUSS OF THE PROPOSED METHOD

By simulating the proposed encryption selective method on a PC using MatLab, images are transferred after the medical image has been encrypted. When transferring the medical image, we transfer the encrypted image to protect the original medical image. This is possible because the only person who will be able to view the original after this process is the person to whom we want to transfer the image. After the decryption procedure of the encrypted medical image, the original image is only sent to that specific person. Fig. 4 displays the original Magnetic Resonance Imaging (MRI) with its encryption image. Once an image has been encrypted, the original image cannot be reconstructed until the encryption process has been completed correctly. In a similar fashion, Fig. 5 and Fig. 6 display the original image as well as the encrypted version of the hand and leg images respectively.

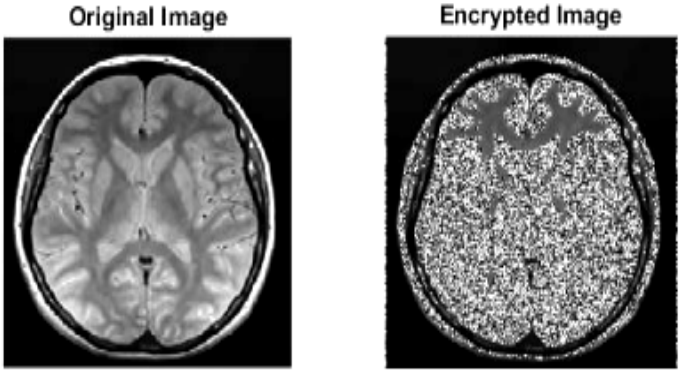


Fig. 4. The input MRI image and corresponding ROI encrypted MRI image

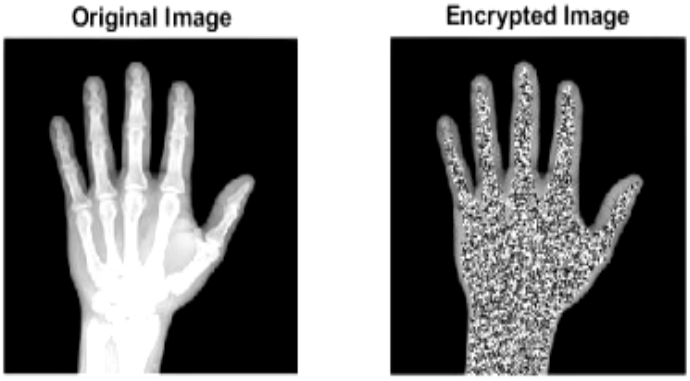


Fig. 5. The input hand image and corresponding ROI encrypted hand image

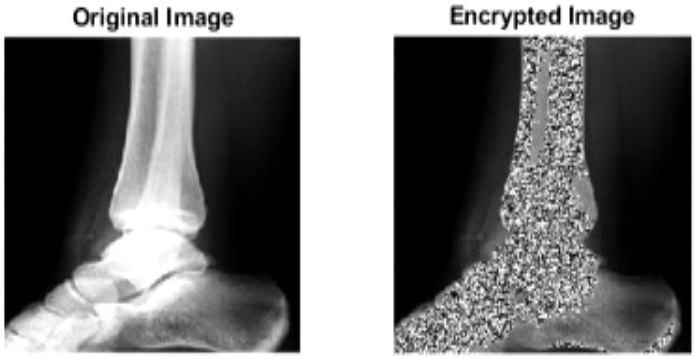


Fig. 6. The input leg image and corresponding ROI encrypted leg image

CONCLUSION

In this paper, we present a method for partially encrypting personal data, such as tumors in the brain, hand parts, and so on. Padding and a rise in data volume as a result of wasted storage space over time are challenges that are inherent to traditional image protection systems. Additionally, since the whole image is encrypted, it cannot be recognized before it is decrypted, and when it is decrypted, critical information is revealed. The difficulty with conventional sub-image encryption is that it encrypts superfluous sections by first encrypting a rectangular region that covers information that has to be kept private. This issue can now be resolved via the technique that was suggested. The suggested approach encrypts the data using a Sudoku matrix after it has been used to determine the important pixels using a histogram peak detection approach.

REFERENCES

1. B. Thomee, D.A. Shamma, and G. Friedland, "YFCC100M: The New Data in Multimedia Research," *Communications of the ACM*, vol. 59, no. 2, pp. 64–73, 2015.
2. S. Pouyanfar, Y. Yang, and S.C. Chen, "Multimedia Big Data Analytics: A Survey," *ACM Computing Surveys*, vol. 51, no. 1, pp. 1–34, 2018.
3. Y. Sun, S. Fang, and Y. Hwang, "Investigating Privacy and Information Disclosure Behavior in Social Electronic Commerce," *Sustainability*, vol. 11, no. 12, pp. 1–27, 2019.
4. M. Han, L. Li, and Y. Xie, "Cognitive Approach for Location Privacy Protection," *IEEE Access*, vol. 6, pp. 13466–13477, 2018.
5. S. Liu, C. Guo, and J.T. Sheridan, "A review of optical image encryption techniques," *Optics & Laser Technology*, vol. 57, pp. 327–342, 2014.
6. Y. Dai, H. Wang, and Z. Zhou, "Research on medical image encryption in telemedicine systems," *Technology and Health Care*, vol. 24, no. s2, pp. S435–S445, 2016.
7. G. Chen, Y. Mao, and C. Chui, "A symmetric image encryption based on 3d chaotic cat maps," *Chaos, Solitons Fractals*, vol. 21, pp. 749–761, 2018.
8. R. Enayatifar, A.H. Abdullah, and I.F. Isnin, "Chaos-based image encryption using a hybrid genetic algorithm and a DNA sequence," *Optics and Lasers in Engineering*, vol. 56, pp. 83–93, 2014.
9. Ü. Çavuşoğlu, S. Kaçar, A. Zengin, and I. Pehlivan, "A novel hybrid encryption algorithm based on chaos and S-AES algorithm," *Nonlinear Dynamics*, vol. 92, pp. 1745–1759, 2018.
10. J. Wu, X. Liao, and B. Yang, "Color image encryption based on chaotic systems and elliptic curve EIGamal scheme," *Signal Processing*, vol. 141, pp. 109–124, 2017.
11. C. Paar, J. Pelzl, *Understanding cryptography: a textbook for students and practitioners*. Springer Science & Business Media, 2009.
12. Z. Su, G. Zhang, and J. Jiang, "Multimedia security: a survey of chaos based encryption technology," in *Multimedia-A Multidisciplinary Approach to Complex Issues*, *InTech*, pp. 99–124, 2012.
13. T. Xiang, S. Guo, and X. Li, "Perceptual visual security index based on edge and texture similarities," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 5, pp. 951–963, 2016.
14. M. Noura, H. Noura, A. Chehab, M.M. Mansour, L. Sleem, and R. Couturier, "A dynamic approach for a lightweight and secure cipher for medical images," *Multimedia Tools and Applications*, vol. 77, no. 23, pp. 31397–31426, 2018.

15. M. Li, D. Lu, and W. Wen, “Cryptanalyzing a color image encryption scheme based on hybrid hyper-chaotic system and cellular automata,” *IEEE Access*, vol. 6, pp. 47102–47111, 2018.
16. L. Xu, Z. Li, and J. Li, “A novel bit-level image encryption algorithm based on chaotic maps,” *Optics and Lasers in Engineering*, vol.78, pp. 17–25, 2016.
17. L. Teng, X. Wang, and J. Meng, “A chaotic color image encryption using integrated bit-level permutation,” *Multimedia Tools and Applications*, vol. 77, no. 6, pp. 6883–6896, 2018.

Received 16.06.2023

INFORMATION ON THE ARTICLE

Izz K. Abboud, ORCID: 0000-0002-8344-8585, Mustansiriyah University, Iraq, e-mail: izz_kadhum@uomustansiriyah.edu.iq

Muaayed F. Al-Rawi, ORCID: 0000-0003-1841-1222, Mustansiriyah University, Iraq, e-mail: muaayed@uomustansiriyah.edu.iq

Nasir A. Al-Awad, ORCID: 0000-0003-3059-4375, Mustansiriyah University, Iraq, e-mail: nasir.awad@uomustansiriyah.edu.iq

ШИФРУВАННЯ ЦИФРОВИХ МЕДИЧНИХ ЗОБРАЖЕНЬ У ПРОГРАМАХ РЕАЛЬНОГО ЧАСУ / Ізз К. Аббуд, Муаайед Ф. Аль-Раві, Насір А. Аль-Авад

Анотація. Інформація про пацієнтів і дані медичних зображень тепер підпадають під дію суворих стандартів безпеки даних і конфіденційності, що є прямим результатом поширення телемедичних методів й інструментів для медичних зображень. Через зазначені проблеми, а також через можливість викрадення даних або інформації, виникає дилема передавання даних на медичні зображення через відкриту мережу. У минулому потенційні рішення включали в себе використання таких методів, як приховування інформації та шифрування зображень. Тим не менше спроба реконструювати оригінальне зображення за допомогою цих підходів може призвести до ускладнень. У ході роботи створено алгоритм для захисту медичних зображень на основі пікселів інтересу. Виявлення піків гістограми з метою обчислення піків у медичних зображеннях пікселів інтересу медичних зображеннях, для яких оброблені порогові значення. Порогове значення відображається як середнє значення усіх піків на гістограмі. Після цього застосовується матриця Sudoku для призначення значень інтересу кожному з цих пікселів. Запропонований метод оцінено за допомогою різноманітних статистичних процедур, а результати цих аналізів порівняно з раніше встановленими стандартами. Згідно з висновками, запропонований метод має кращу ефективність безпеки порівняно з іншими вже використовуваними методами шифрування зображень.

Ключові слова: додатки реального часу, медичні зображення, шифрування, безпека, виявлення піків.

UDC 004.852 + 616.12-07
DOI: 10.20535/SRIT.2308-8893.2024.1.03

CARDIOMYOPATHY PREDICTION IN PATIENTS WITH PERMANENT VENTRICULAR PACING USING MACHINE LEARNING METHODS

E.O. PEREPEKA, V.V. LAZORYSHYNETS, V.O. BABENKO, I.V. DAVYDOVYCH, I.A. NASTENKO

Abstract. Pacing-induced cardiomyopathy is a notable issue in patients needing permanent ventricular pacing. Identifying risk groups early and swiftly preventing the ailment can reduce patient harm. However, current prognostic methods require clarity. We employed machine learning to develop predictive models using medical data. Three algorithms — decision tree, group method of data handling, and logistic regression — formed models that forecast pacing-induced cardiomyopathy. These models displayed high accuracy in predicting development, signifying soundness. Factors like age, paced QRS width, pacing mode, and ventricular index during implantation significantly influenced predictions. Machine learning can enhance pacing-induced cardiomyopathy prediction in ventricular pacing patients, aiding medical practice and preventive strategies.

Keywords: permanent ventricular pacing, risk factors, artificial intelligence, forecasting, machine learning.

INTRODUCTION

Right ventricular myocardial pacing remains dominating method in providing medical care to patients with various potentially fatal bradyarrhythmias, even though at the beginning of the 21st century, a relation between this form of cardiac pacing and the left ventricular contractility impairment [1], as well as deterioration of clinical outcomes in the distant period [2; 3].

According to data from various sources, the incidence of pacing-induced cardiomyopathy (PICM) in patients with conventional right ventricular pacing and with preserved initial left ventricle ejection fraction (LVEF) ranges from 7.5 to 26% [4–10].

The risk of heart failure hospitalizations (HFH) and overall mortality are significantly higher among patients with PICM, as was shown in a large retrospective study by Sung Woo Cho et al. [10]. Though in patients with initially reduced systolic function of the left ventricle and high burden of ventricular pacing, the factors of deterioration of the clinical outcomes are well established [3], in patients with preserved LVEF, they have not yet been fully studied. Along with the wide availability and significant global experience of using this method of cardiac pacing in clinical practice, there is a growing number of publications focusing on the adverse effects of right ventricular myocardial pacing (and investigating risk factors that led to them), one of which is the development of the so-called pacing-induced cardiomyopathy, which is characterized by a decrease in the left ventricle contractility and negative remodeling of the heart chambers,

The identification of risk factors and prediction of PICM development in patients with an implanted pacemaker is an objective of significant importance for modern medicine, considering the appearance of modern physiological methods of cardiac pacing (such as conduction system pacing) which allow preventing or minimizing the negative consequences of right ventricular myocardial pacing [11–14]. It is important to note that machine learning and artificial intelligence are becoming more prevalent in healthcare, particularly in cardiology. These technologies have successfully predicted disease cases and identified pathologies [15]. However, studies that apply machine learning to indicate PICM were not found after analyzing various literature sources.

The primary focus of research in the intersection of cardiology and machine learning is centered around the prediction and diagnosis of diseases, including ischemic heart disease (IHD) [16; 17], HF [18], atrial arrhythmias [19; 20], and others, using data from patients' medical records, imaging, and biosignals. In the context of PICM, the scientific community focuses on studying risk factors and developing preventive measures [21; 22]. Thus, the use of machine learning can contribute to identifying patients at considerable risk of PICM, which will allow the introduction of prompt and effective therapeutic interventions or other invasive strategies. This study focuses on figuring out the possibilities of using the machine learning methodology to predict the development of PICM in patients with permanent ventricular pacing.

Specific tasks due to the urgency of the problem are determined by the following aspects:

1. Development of PICM prediction models based on various machine learning algorithms using the available medical dataset.

2. Comprehensive evaluation of constructed models using classification metrics including (but not limited to) accuracy, sensitivity, and specificity.

3. A detailed study of the importance of individual factors included in the model in the context of their influence on predicting PICM.

The objective of the study is the construction of detailed prognostic models for the development of PICM and the identification of critical factors that contribute to the occurrence of this complication.

MATERIALS AND METHODS

In this research, we used anonymized data from patient examinations performed at the State Institution "M. Amosov National Institute of Cardiovascular Surgery" of the National Academy of Medical Sciences of Ukraine within the framework of the cooperation agreement with the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute".

Before initiating the study, the M. Amosov National Institute of Cardiovascular Surgery performed a bioethical evaluation of the research protocol. We analyzed data on thirty-four patients, of which nine (26.5%) were diagnosed with PICM, which was determined with an LVEF of less than 45%. Left ventricle ejection fraction was within normal limits in the remaining twenty-five patients (73.5%).

The study included only those patients who met the following criteria: availability of echocardiographic data at the time of pacemaker (PM) implantation;

total percentage of ventricular pacing at the time of examination is not less than 90%; preserved LVEF at the time of implantation (≥ 50%); age restrictions patients (18–80 years at the time of implantation and control examination, respectively); this was to be a primary PM implantation without previous endocardial lead extractions or power source replacements.

For the study, the M. Amosov National Institute of Cardiovascular Surgery systematically collected and documented data, which included gender, age, the period from PM implantation to follow-up, as well as the main and concomitant diagnoses of the patients.

In addition, the data from echocardiographic and electrocardiographic studies were collected, as well as cardiac pacing parameters at two stages: at the time of hospitalization and during the control examination.

It is important to note that the used database has seventeen attributes, described in detail in Table 1.

T a b l e 1 . Attributes of the selected database

Attribute[1]	Data type	Symbolic notation
PICM	Binary	y
Gender	Binary	x_1
Age	Continuous integer	x_2
Time from pacemaker implantation to follow-up	Continuous integer	x_3
LVEF at the time of implantation	Continuous integer	x_4
LA diameter at the time of PM implantation	Continuous integer	x_5
Width of native QRS complex	Continuous integer	x_6
Width of paced QRS complex	Continuous integer	x_7
Presence of atrial arrhythmias (including AF)	Binary	x_8
Right ventricle pacing site	Binary	x_9
Structural heart diseases	Binary	x_{10}
Diabetes mellitus	Binary	x_{11}
Hypertension	Binary	x_{12}
Ischemic heart disease	Binary	x_{13}
Pacemaker type (single-chamber/dual-chamber)	Binary	x_{14}
Rate-adaptive pacing mode	Binary	x_{15}
Left ventricle EDI at the time of PM implantation	Continuous	x_{16}

The purpose of applying machine learning technologies was to find key input (independent) variables x that correlate with the presence of cardiomyopathy, represented as an output (dependent) variable y. Machine learning aims to identify patterns and relationships between variables through data processing. Machine learning algorithms are designed to explore dependencies in data and show trends that may not be clear. A substantial number of scientific developments confirmed this hypothesis, where authors considered tasks from various subject areas, including medicine [23–25].

[1] Accepted abbreviations: PICM — stimulation-induced cardiomyopathy; AP — artificial pacemaker; LVEF — left ventricular ejection fraction; LA — left atrium; AF — atrial fibrillation; EDI — end-diastolic index.

As can be seen from Table 1, the output variable *y* is dichotomous, which writes down the need to solve the classification problem.

Taking this into account, we decided to use three simple classification algorithms: decision tree [26], group method of data handling (GMDH) [27], and logistic regression [28].

Decision trees are one of the most convenient algorithms because of their visual interpretation and ability to oversee numerical and categorical data. They work by partitioning the space of input variables into regions corresponding to different classes of the output variable. However, they can be prone to overfitting, especially with complex data.

GMDH is an algorithm that creates a model based on a pairwise comparison of objects. Its main advantage is the high interpretability of the results, which supplies the possibility of a clear understanding of the classification mechanisms. However, due to high computational complexity, GMDH may only be effective for a small volume of data.

Logistic regression is a statistical algorithm commonly used to predict the probability of an event occurring by applying a logistic function. This method works well on two-class problems but can run into issues with non-linear relationships or many categorical variables.

RESULTS

Before building PICM prediction models, we divided the patient sample into a train (80%, or twenty-seven patients) and a test (20%, or seven patients) using a stratification method, which preserves the class ratio between subjects in each sample. We measured the performance of each algorithm by its accuracy (proportion of correctly classified patients), sensitivity (proportion of correctly classified patients with pathology), and specificity (proportion of correctly classified healthy patients) [29]. The performance of the selected classification algorithms is presented in Table 2.

Table 2. Evaluation of the constructed PICM prediction models by classification metrics

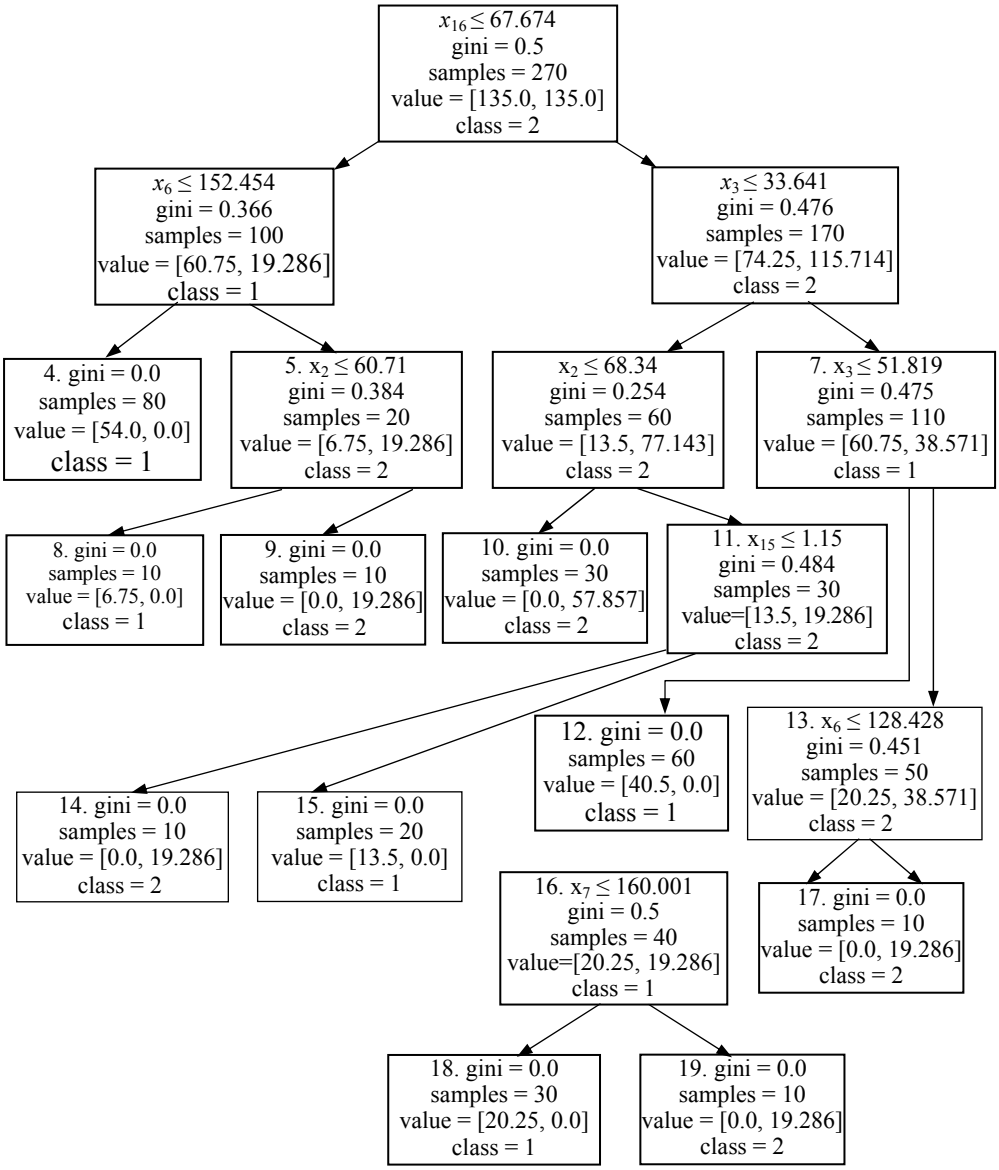
Classifier	Train (80%)			Test (20%)		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
Decision tree	1.000	1.000	1.000	1.000	1.000	1.000
GMDH	1.000	1.000	1.000	1.000	1.000	1.000
Logistic regression	0.852	0.857	0.850	1.000	1.000	1.000

Using the *scikit-learn* (the Python library), we implemented the classification model (Figure) based on the decision tree method. The tree has a depth of five and consists of nine leaves.

According to the data presented in Table 2, this model shows 100% accuracy on the test sample, indicating its reliability and the absence of overfitting phenomena.

Six input variables: x_3, x_6, x_7, x_2, x_{15}, and x_{16}, were used in the model and are illustrated in the tree (Fig. 1). We expressed each variable's impact through the tree's weights: $x_3 = 0.303$, $x_6 = 0.193$, $x_7 = 0.146$, $x_2 = 0.127$, $x_{15} = 0.118$, $x_{16} = 0.113$. The weighting coefficients were figured out using the Gini index [26]. It

considers the number of times each variable was used to split the data and how effective this split was. The final weighting of each variable is a normalized value based on the overall reduction in the Gini index caused by each variable. A variable with a higher importance value is considered more "important" to the model.



Decsion tree model

Applying the GMDH, we generated a model for which the formula below is given:

$$y = 0.001x_{16}^2 - 0.0009x_7x_{16} - 0.643x_{13}x_{14} - 0.006x_2x_{12} - 0.006x_{12}x_{16} -$$

$$- 0.031x_9x_{15} + 0.0003x_7^2 + 0.018x_4x_{12} - 0.519x_{11}x_{12} - 0.736$$

We conducted the training process by using *GMDH Shell DS* software. As shown in Table 2, the results show that the GMDH model is completely accurate

in testing, indicating no model overfitting. This model includes ten independent variables, namely: $x_{16}, x_{13}, x_7, x_{14}, x_{12}, x_{11}, x_4, x_2, x_9$, and x_{15}. The weight of each of these variables is found based on the change in the model's predicted values when replacing the variable's actual values with its average value. As a result, the following variable weights were obtained: $x_{16} = 83.7\%$, $x_{13} = 53.9\%$, $x_7 = 53.2\%$, $x_{14} = 39.7\%$, $x_{12} = 33.4\%$, $x_{11} = 30.8\%$, $x_4 = 17.6\%$, $x_2 = 6.4\%$, $x_9 = 1.6\%$, $x_{15} = -0.2\%$.

In the third step, we applied a logistic regression model. The general form of the logistic model is defined by formula:

$$p = \frac{1}{1 + e^{-y}},$$

where p is the calculated probability of occurrence of a given event (PICM in this context); e is the basis of natural logarithms (2.713); y is the linear regression equation. The following logistic regression model was obtained:

$$y = -0.065x_2 + 0.044x_7 - 1.758x_{15} + 0.119x_{16} - 8.179.$$

We conducted the training procedure using the *scikit-learn* package of the *Python* programming language. The complexity of this model is four. Interestingly, this model includes variables also used in the earlier models: x_2, x_7, x_{15}, and x_{16}.

DISCUSSION

While analyzing constructed models for predicting pacing-induced cardiomyopathy based on the data in Table 2, it was found that only the logistic regression model failed to present an ideal result for the entire sample (with a classification accuracy of 85.2% in the train, despite 100% accuracy in the test). The observed phenomenon can be explained by the intrinsic simplicity of the logistic model in contrast to the other comparable models utilized in the research.

The developed decision tree model was structurally simple and included only six independent variables. While the results of the classification estimation are excellent, this model may be prone to misprediction of new data due to the limited initial sample size. The GMDH model wins here by incorporating ten independent variables for prediction. Additionally, the algorithm for constructing such a model allows non-linear combinations of variables, which sensitively increases their predictive power.

The identified combinations of factors influencing the PICM development align with the latest global publications. The three prediction models include the following independent variables: x_2 (patient's age), x_7 (width of paced QRS complex), x_{15} (rate-adaptive pacing mode), and x_{16} (left ventricular EDI at the time of PM implantation).

Among them, variable x_7 has a significant impact, especially in the decision tree (0.147) and GMDH (53.2%), with one of the highest weighting coefficients. There are also independent variables that were not included in any of the models, such as x_1 (patient gender), x_5 (LA diameter at the time of PM implantation), x_8 (presence of atrial arrhythmias), and x_{10} (structural heart diseases).

The modeling results obtained during the study open the possibility of predicting undesirable clinical consequences of right ventricular pacing based on combinations of the most informative factors. That makes it possible to prevent the influence of these factors or intervene at the stage of medical care provided, choosing more physiological cardiac pacing methods.

CONCLUSION

The study successfully developed models for predicting pacing-induced cardiomyopathy (PICM) based on various machine learning algorithms using an available medical dataset of thirty-four patients.

Methods used — including decision tree, group method of data handling (GMDH), and logistic regression — allowed robust predictive models to be created. On the test sample, all of them showed 100% prediction accuracy.

Obtained results demonstrated the high efficiency of the used machine learning algorithms in terms of the accuracy of the PICM prediction, the absence of overfitting, and the ability of the models to classify adequately normal and pathological states of patients.

A detailed study of values included in the models allows an understanding of their role in developing PICM.

The most significant data included in the models were patient age, paced QRS complex width, rate-adaptive pacing mode, and left ventricular end-diastolic index (EDI) at the time of pacemaker implantation.

The developed models can serve as a basis for further improving diagnostic and treatment technologies for PICM prevention strategies.

REFERENCES

1. J.C. Nielsen, L. Kristensen, H.R. Andersen, P.T. Mortensen, O.L. Pedersen, and A.K. Pedersen, “A randomized comparison of atrial and dual-chamber pacing in 177 consecutive patients with sick sinus syndrome,” *Journal of the American College of Cardiology*, 42(4), pp. 614–23, 2003. doi: 10.1016/S0735-1097(03)00757-5.
2. M.O. Sweeney et al., “Adverse Effect of Ventricular Pacing on Heart Failure and Atrial Fibrillation Among Patients With Normal Baseline QRS Duration in a Clinical Trial of Pacemaker Therapy for Sinus Node Dysfunction,” *Circulation*, 107(23), pp. 2932–2937, 2003. doi: 10.1161/01.CIR.0000072769.17295.B1.
3. The DAVID Trial Investigators*, “Dual-Chamber Pacing or Ventricular Backup Pacing in Patients With an Implantable Defibrillator,” *JAMA*, 288(24), pp. 3115–3123, 2002. doi: 10.1001/jama.288.24.3115.
4. S. Khurshid et al., “Incidence and predictors of right ventricular pacing-induced cardiomyopathy,” *Hear Rhythm*, 11(9), pp. 1619–1625, 2014. doi: 10.1016/j.hrthm.2014.05.040.
5. E.L. Kiehl et al., “Incidence and predictors of right ventricular pacing-induced cardiomyopathy in patients with complete atrioventricular block and preserved left ventricular systolic function,” *Hear Rhythm*, 13(12), pp. 2272–2278, 2016. doi: 10.1016/j.hrthm.2016.09.027.
6. A. Abdin et al., “Incidence and predictors of pacemaker induced cardiomyopathy: A single-center experience,” *Journal of Electrocardiology*, 57, pp. 31–34, 2019. doi: 10.1016/j.jelectrocard.2019.08.016.
7. Mohamed Abdelmohsen Sayed, Haitham Abd El Fatah Badran, Said Khaled, and Emad Effat Fakhry, **“**Predictors of right ventricular pacing-induced left ventricular dysfunction in pacemaker recipients with preserved ejection fraction,” *Herzschrittmachertherapie + Elektrophysiologie*, 33(3), pp. 312–318, 2022. doi: 10.1007/s00399-022-00880-w.

8. H.T. Perla, S. Chandra Srinath Patloori, A. Manickavasagam, D. Chase, and J. Roshan, "Do the predictors of right ventricular pacing-induced cardiomyopathy add up?" *Indian Heart Journal*, 73(5), pp. 582–587, 2021. doi: 10.1016/j.ihj.2021.07.011.
9. J.H. Kim, K.-W. Kang, J.Y. Chin, T.-S. Kim, J.-H. Park, and Y.J. Choi, "Major determinant of the occurrence of pacing-induced cardiomyopathy in complete atrioventricular block: a multicentre, retrospective analysis over a 15-year period in South Korea," *BMJ Open*, 8(2), pp. e019048, 2018. doi: 10.1136/bmjopen-2017-019048.
10. S.W. Cho et al., "Clinical features, predictors, and long-term prognosis of pacing-induced cardiomyopathy," *European Journal of Heart Failure*, 21(5), pp. 643–651, 2019. doi: 10.1002/ejhf.1427.
11. M. Abdelrahman et al., "Clinical Outcomes of His Bundle Pacing Compared to Right Ventricular Pacing," *Journal of the American College of Cardiology*, 71(20), pp. 2319–2330, 2018. doi: 10.1016/j.jacc.2018.02.048.
12. P.S. Sharma et al., "Clinical outcomes of left bundle branch area pacing compared to right ventricular pacing: Results from the Geisinger-Rush Conduction System Pacing Registry," *Hear Rhythm*, 19(1), pp. 3–11, 2022. doi: 10.1016/j.hrthm.2021.08.033.
13. P. Vijayaraman et al., "Outcomes of His-bundle pacing upgrade after long-term right ventricular pacing and/or pacing-induced cardiomyopathy: Insights into disease progression," *Hear Rhythm*, 16(10), pp. 1554–1561, 2019. doi: 10.1016/j.hrthm.2019.03.026.
14. L.M. Rademakers, S. Bouwmeester, T.P. Mast, L. Dekker, P. Houthuizen, and F.A. Bracke, "Feasibility, safety and outcomes of upgrading to left bundle branch pacing in patients with right ventricular pacing induced cardiomyopathy," *Pacing and Clinical Electrophysiology*, 45(6), pp. 726–732, 2022. doi: 10.1111/pace.14515.
15. K. Basu, R. Sinha, A. Ong, and T. Basu, "Artificial intelligence: How is it changing medical sciences and its future?" *Indian Journal of Dermatology*, 65(5), pp. 365–370, 2020. doi: 10.4103/ijd.IJD_421_20.
16. Ch Anwar Ul Hassan et al., "Effectively Predicting the Presence of Coronary Heart Disease Using Machine Learning Classifiers," *Sensors*, 22(19), 7227, 2022. doi: 10.3390/s22197227.
17. O. Petrunina, D. Shevaga, V. Babenko, V. Pavlov, S. Rysin, and I. Nastenko, "Comparative Analysis of Classification Algorithms in the Analysis of Medical Images From Speckle Tracking Echocardiography Video Data," *Innovative Biosystems and Bioengineering*, 5(3), pp. 153–166, 2021. doi: 10.20535/ibb.2021.5.3.234990.
18. J. Wang, "Heart Failure Prediction with Machine Learning: A Comparative Study," *Journal of Physics: Conference Series*, 2031(1), 12068, 2021. doi: 10.1088/1742-6596/2031/1/012068.
19. M. Varshneya, X. Mei, and E.A. Sobie, "Prediction of arrhythmia susceptibility through mathematical modeling and machine learning," *Proceedings of the National Academy of Sciences*, 118(37), 12068, 2021. doi: 10.1073/pnas.2104019118.
20. E.R. Pujadas et al., "Atrial fibrillation prediction by combining ECG markers and CMR radiomics," *Scientific Reports*, 12(1), 18876, 2022. doi: 10.1038/s41598-022-21663-w.
21. P.Z. Fruelund et al., "Risk of Pacing-Induced Cardiomyopathy in Patients with High-Degree Atrioventricular Block—Impact of Right Ventricular Lead Position Confirmed by Computed Tomography," *Journal of Clinical Medicine*, 11(23), 7228, 2022. doi: 10.3390/jcm11237228.
22. J. Mizner, P. Jurak, H. Linkova, R. Smisek, and K. Curila, "Ventricular Dyssynchrony and Pacing-induced Cardiomyopathy in Patients with Pacemakers, the Utility of Ultra-high-frequency ECG and Other Dyssynchrony Assessment Tools," *Arrhythmia & Electrophysiology Review*, 11, e.17, 2022. doi: 10.15420/aer.2022.01.
23. I.H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Computer Science*, 2(3), 160, 2021. doi: 10.1007/s42979-021-00592-x.
24. A. Rajkomar, J. Dean, and I. Kohane, "Machine Learning in Medicine," *The New England Journal of Medicine*, 380(14), pp. 1347–1358, 2019. doi: 10.1056/nejmra1814259.
25. V. Babenko et al., "Classification of Pathologies on Medical Images Using the Algorithm of Random Forest of Optimal-Complexity Trees," *Cybernetics and Systems Analysis*, 59(2), pp. 346–358, 2023. doi: 10.1007/s10559-023-00569-z.

26. B. Charbuty, A. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," *Journal of Applied Science and Technology Trends*, 2(1), pp. 20–28, 2021. doi: 10.38094/jastt20165.
27. V. Vaishnav, J. Vajpai, "Assessment of impact of relaxation in lockdown and forecast of preparation for combating COVID-19 pandemic in India using Group Method of Data Handling," Chaos, Solitons & Fractals, 140, 110191, 2020. doi: 10.1016/j.chaos.2020.110191.
28. C. El Morr, M. Jammal, H. Ali-Hassan, and W. El-Hallak, "Logistic Regression," in book *"Machine Learning for Practical Decision Making,"* pp. 231–249, 2022. doi: 10.1007/978-3-031-16990-8_7.
29. Ž.Đ. Vujovic, "Classification Model Evaluation Metrics," *International Journal of Advanced Science and Computer Applications*, 12(6), 2021. doi: 10.14569/IJACSA.2021.0120670.

Received 15.08.2023

INFORMATION ON THE ARTICLE

Eugene O. Perepeka, ORCID: 0000-0001-9755-8825, Amosov National Institute of Cardiovascular Surgery, Ukraine, e-mail: eugeneperepeka@gmail.com

Vasyl V. Lazoryshynets, ORCID: 0000-0002-1748-561X, Amosov National Institute of Cardiovascular Surgery, Ukraine, e-mail: lazorch@ukr.net

Vitalii O. Babenko, ORCID: 0000-0002-8433-3878, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: vbabenko2191@gmail.com

Illia V. Davydovych, ORCID: 0000-0001-9987-8267, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: bkmzbkmz6@gmail.com

Ievgen A. Nastenko, ORCID: 0000-0002-1076-9337, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: nastenko.e@gmail.com

ПРОГНОЗУВАННЯ КАРДІОМІОПАТІЇ У ПАЦІЄНТІВ З ПОСТІЙНОЮ ШЛУНОЧКОВОЮ ЕЛЕКТРОКАРДІОСТИМУЛЯЦІЄЮ ЗА ДОПОМОГОЮ МЕТОДІВ МАШИННОГО НАВЧАННЯ / Є.О. Перепека, В.В. Лазоришинець, В.О. Бабенко, І.В. Давидович, Є.А. Настенко

Анотація. Кардіоміопатія, спричинена кардіостимуляцією, є важливою проблемою для пацієнтів, які потребують постійної шлуночкової кардіостимуляції. Раннє виявлення груп ризику та швидка профілактика недуги можуть зменшити шкоду для пацієнтів. Однак сучасні методи прогнозування потребують доопрацювання. Застосовано машинне навчання для розроблення прогностичних моделей на основі медичних даних. Три алгоритми — дерево рішень, група оброблення даних та логістична регресія — сформували моделі, які прогнозують кардіоміопатію, спричинену кардіостимуляцією. Ці моделі показали високу точність у прогнозуванні розвитку, що свідчить про їх надійність. Ключові фактори, такі як вік, ширина QRS, режим кардіостимуляції та шлуночковий індекс під час імплантації, суттєво впливали на прогнози. Машинне навчання може покращити прогнозування кардіоміопатії, спричиненої кардіостимуляцією, у пацієнтів, які перебувають на шлуночковій електрокардіостимуляції, допомагаючи медичній практиці та профілактичним стратегіям.

Ключові слова: постійне ритмоведення шлуночків, фактори ризику, штучний інтелект, прогнозування, машинне навчання.

ПРОБЛЕМИ ПРИЙНЯТТЯ РІШЕНЬ ТА УПРАВЛІННЯ В ЕКОНОМІЧНИХ, ТЕХНІЧНИХ, ЕКОЛОГІЧНИХ І СОЦІАЛЬНИХ СИСТЕМАХ

UDC 519.766.4:004.942
DOI: 10.20535/SRIT.2308-8893.2024.1.04

OPERATIONAL RISK ESTIMATION USING SYSTEM ANALYSIS METHODOLOGY

P.I. BIDYUK, O.L. TYMOSHCHUK, L.B. LEVENCHUK

Abstract. Financial risks are considered today as popular research topics due to the existing practical necessity for the use of their mathematical models, estimates of possible loss in many areas of human activities, forecasting, and respective managerial decisions in financial and other spheres where capital, obligations, stocks, bonds, and other activities are circulating successfully. Financial processes today exhibit sophisticated forms of evolution in time that require the application of sophisticated modeling, risk estimating, forecasting, and decision-making/support methods, techniques, and procedures. The system analysis approach is applied to solving such problems as a unique and universal research methodology. The financial risks, specifically the operational ones in the study considered, are classified as nonlinear and nonstationary processes that require appropriate methods for analysis and a rather sophisticated analytical description to estimate and forecast possible loss. The results of operational risk analysis are achieved in the form of systemic methodology, models constructed with statistical data, regression analysis, and Bayesian techniques, and estimated loss with the models. The models and system analysis approach proposed for analyzing financial processes are suitable for practical applications, provided the users have appropriate statistical data and expert estimates.

Keywords: financial operational risk, mathematical model, statistical data, system analysis methodology, loss estimation, decision support system.

INTRODUCTION

Financial risks are related to very popular research topics due to existing practical necessity for the use of their mathematical models, estimates of possible loss in many areas of human activities, forecasting and respective managerial decisions in financial and many other spheres where capital, obligations, stocks, bonds and other activities are circulating successfully [1]. Generally financial processes often exhibit rather sophisticated forms of evolution in time what requires application of modern modeling, forecasting and decision making/support methods, technics and procedures. The effective system analysis approach should be applied to solving such problems successfully [2]. The most known financial risks can be classified as nonlinear and nonstationary processes (NNP) that require appropriate attention for analysis and rather sophisticated analytical description to estimate and forecast possible loss.

The possible process nonstationarities appear due to availability of deterministic and stochastic trends, changing in time variance, and possible nonlinearities can be divided into two wide classes: nonlinearities with respect to variables and nonlinearities with respect to parameters. The nonlinearities with respect to variables are easier to cope with because the models (parameters) containing them can be estimated easier than models nonlinear with respect to parameters. For example, the model describing process with nonlinear trend (say, polynomial trend of second or higher order) can be correctly estimated using ordinary or nonlinear LS (NLS) or maximum likelihood (ML) methods. On the other side, the models nonlinear with respect to parameters may require for their estimation application of sophisticated optimization procedures like NLS, including Bayesian techniques such as Monte Carlo for Markov Chain (MCMC) or Bayesian optimization approach.

Thus, most of the financial processes that researchers have to cope with today are nonlinear and nonstationary what requires paying special attention to modeling their structure and parameter estimation. Generally, most of the processes around us (in economy, finances, industrial production, weather forecasting, description of natural phenomena) are related to NNPs, and they should be modeled, forecasted and governed correctly using respective theoretical basis provided by the estimation theory [3]. Usually specialized intellectual decision support systems (IDSS), based upon system analysis principles, are constructed to analyze the processes that help substantially to improve constructed model adequacy, quality of forecasts, respective decisions, and increase general understanding of practical situations related to the modeling and forecast estimation problems. This is especially true regarding financial risk analysis where different model types are to be applied to reach high quality results of possible loss estimation/forecasting.

Among often met financial risks that researchers today have to cope with the following ones should be mentioned: market risk, operational risk, risk of a country, credit and liquidity risks, risk of structured products sale, Exchange-traded products (ETP) risk, Over-the-counter (OTC) derivatives risk and some others. In this study we provide for a short review of modern methods for modeling selected types of financial risks, and estimation of possible loss with special attention to active usage of IDSS. Comparison of some approaches to risk estimation will also be given. The studies [4, 5, 6] provide for grounded notion of “risk”, consider possible instruments for operational risk analysis, and describe basic problems met on the way of development and practical implementation of risk estimation and prediction systems in financial institutions. The study [7] touches upon analysis of key risks inherent to Ukrainian banking system. Many studies are devoted to analysis and use of loss distributions approach (LDA) because it provides the possibility for estimating the volume of capital necessary for covering operational loss. The paper [8] is completely dedicated to analysis and practical use of the popular LDA approach.

PROBLEM STATEMENT

The purposes of the study are as follows: to provide a short review of operational financial risk modeling methods and respective loss estimation; to consider possibility of financial risk estimation by making use of regression techniques (generalized linear models) and Bayesian approach to data analysis; to compare quality

of risk estimation by the methods selected; to give necessary illustrative examples of financial operational risk estimation; to stress necessity of constructing and practical application of IDSS to estimation/prediction of possible financial loss.

SOME GENERAL APPROACHES TO FINANCIAL RISK ESTIMATION

Here we start analysis with operational risk that is available practically in every activity when we consider capital circulation problems on different levels of economy and finances. The term "operational risk" does not have clearly established definition. Some financial institutions define it as non-measurable risk (what is also correct). In 2001 the Basel Committee on Banking Supervision formulated general supervision standards and governing principles for banking system, and gave definition for operational risk: "operational is the risk of loss due to inadequate or erroneous internal processes, actions of co-workers or systems, or influence of external events" [1].

The Basel Committee also proposed some methods for operational risk estimation, namely: the basic indicator approach (BIA); standardized indicator approach (SIA); internal estimation approach (IMA); an approach based upon loss distribution curve (LDA); scoring cards, and scenario analysis techniques. The first three methods are relatively simple, and can be used in various financial organizations though they do not take into consideration special features of an enterprise, its size and principles if risk management implemented within its structure. Using the methods mentioned "good" and "bad" financial company will be keeping the same amount of capital to cover operational risks. The last three methods are considered as advanced approaches to operational risk analysis and estimation. They are based upon mathematical models and advanced information technologies, and provide the possibility for decreasing the capital necessary for covering operational risk loss. Now consider some types of models used in practice of operational and some other risks analysis.

GENERALIZED LINEAR MODELS

Generalized linear models (GLM) create to some extent universal approach to modeling various processes including linear and nonlinear, stationary and nonstationary ones. This is possible thanks to the fact that this group of models includes the models of different structures: pure linear Bayesian multiple regression; Poisson regression; variance and covariance analysis; nonlinear structures including nonlinear polynomial elements; nonlinear structures like logit and probit models that are capable to capture nonlinear dependences in the processes under study. Combination of linear and nonlinear structural elements in one model is directed to describing complex nonlinear nonstationary dependences inherent to many financial processes. For example, nonlinear logit and probit models can be combined with paired or multiple linear regression, Bayesian networks (static and/or dynamic), Bayesian data filtering procedures that can be linear or nonlinear.

ESTIMATION OF REGRESSION MODEL STRUCTURE

The basics of regression model structure estimation were proposed by Box and Junkins in the 1960s. Today the modeling methodology proposed, based on system analysis approach, includes the following steps:

- systemic analysis of a process under study on the basis of available basic theory and expert knowledge, analysis of inputs and outputs behavior, represented by the time series information, study of earlier existing model structures, and other accessible information;
- preliminary processing of available data that is based upon application of appropriate normalizing and filtering procedures, appropriate analysis of possible extreme values, filling in missing values, structuring the data etc;
- analysis of statistical data for stationarity and nonlinearity using available statistical tests;
- generation and estimation of candidate model structures with taking into consideration possible structure of GLM: identification of stochastic and systemic components and link function; computing descriptive characteristics; determining statistical significance of independent variables using Wald statistic; estimate statistical characteristics of other structural elements of constructed mathematical model (for example, residuals);
- selection/development and application of model parameter estimation method (or methods); it can be linear or nonlinear least squares (NLS) method, maximum likelihood (ML) or Monte Carlo based Bayesian approach;
- selection of the best fitting model in the set of estimated candidates using statistical quality criteria.

Now consider some details of the stages mentioned.

PRELIMINARY DATA PROCESSING

The time series theory considers observations as random variables containing deterministic component. Generally we have practically always process the data measurements influenced by some random external disturbances and measurement errors. The purpose of preliminary data processing is in eliminating possible errors from available measurements and improving conditions for determining their distribution and further modeling. According to current requirements when GLMs are hired for formal describing the processes under study the data should correspond to the family of exponential distributions [9; 10].

Preliminary data processing usually includes the following operations:

- normalization and possible correcting of measurements; normalization means application of logarithmic operator or transforming the data to acceptable and convenient amplitude interval;
- possible data correction may include imputation of missing values, application of filtering techniques, and processing extreme values;
- computing, when necessary, first and higher order differences that are necessary for analysis of corresponding effects of a time series under study.

Sample mean is subtracted sometimes from the observations to get a possibility for constructing a model for deviations. It is clear that application of specific data preprocessing technique depends on specific features of the data available and established purpose of modeling (forecasting, deeper analysis of a process or control).

Nonlinear deterministic trend available in data can be identified by estimating the following time polynomial:

$$y(k) = a_0 + c_1 k + c_2 k^2 + ... + c_m k^m,$$

where $k=0,1,2,...$ is discrete time $(t=kT_s)$, where t is continuous time, and T_s is sampling interval for measurements. If one of the coefficients c_i, $i=2,...,m$ is statistically significant, then hypothesis on trend linearity is rejected. In the case when trend is quickly changing it time and its functional description is not adequate enough, then model of stochastic trend should be constructed that is based upon combinations of random processes.

ESTIMATION OF OTHER ELEMENTS OF GLM MODEL STRUCTURE

Taking into consideration possible GLM distributions there are several GLM types presented in Table 1.

Table 1. GLM types considered in the study

Type of model	Link function	Distribution of dependent variable
GLM	$g(\mu)=\mu$	Normal
Log-linear model	$g(\mu)=\ln(\mu)$	Poisson
Logistic model	$g(\mu)=\ln\left(\frac{\mu}{1-\mu}\right)$	Binomial
Probit-analysis	$g(\mu)=\Phi^{-1}\mu$	Binomial
"Survival" analysis	$g(\mu)=\mu^{-1}$	Gamma-distribution, exponential

Considering GLM, model structure is estimated from the point of view of its components and includes the following elements:

- stochastic component: this is independent variable that is characterized by distribution related to the exponential family;
- systematic component that includes p independent variables creating so called "linear predictor":

$$\eta = X \cdot \beta;$$

- estimating features of the link function with taking into consideration their classification.

Besides the definition mentioned the model structure includes the following elements: model order (maximum order of equations creating the model); model dimension or number of equations creating the model; selection of a link function, variance function and elements of the linear predictor.

MODEL PARAMETER ESTIMATION

After estimating model structure the next problem is to compute model parameters using available statistical/experimental data. Usually, the model structure estimated provides information on a number of parameters to be estimated and the estimation method to be applied in concrete case. In some cases, it is reasonable to use the saving principle that supposes the following: number of parameters to be estimated should not exceed their necessary quantity. Here "necessity" can be viewed as necessity to preserve in the model constructed basic statistical characteristics of the process being studied: mean value, variance and covariance.

Very often parameters of a generalized linear model can be estimated using ordinary LS (OLS) method. In a case of normal data distribution this is equivalent to application of maximum likelihood (ML) method. Generally, widely used methods for estimation of GLM parameters are the following: OLS, weighted LS (WLS), ML, method of moments and Monte Carlo techniques (iterative and non-iterative). As far as the OLS in some cases results in biased parameter estimates due to its sensitivity to data outliers an alternative procedure for estimation is ML.

In the case of normal residuals logarithmic likelihood function, l, can be represented for n measurements as follows:

$$-2l = n\log(2\pi\sigma^2) + \sum_{i=1}^{n} \frac{(y_i - \mu_i)^2}{\sigma^2}.$$

For a fixed, σ^2, maximization of l is equivalent to minimization of quadratic deviations from mean: $\sum[y(k) - \mu]^2$; thus, for linear model we have:

$$\eta_i = \mu_i = \sum_{j=1}^{p} x_{ij}\beta_j.$$

Very popular today method for GLM parameter estimation is Markov Chain Monte Carlo (MCMC) procedure that can be applied to linear and nonlinear models. The estimation algorithm based on this method is functioning as follows [11]:

- sampling X_u^{i+1} from the distribution $P(X_u \mid X_0, \theta^i)$, where X_0 is experimental data;
- sampling parameter estimate, θ^{i+1}, from distribution $P(\theta \mid X_0, X_u^i)$.

The statistical series generated this way under weak regularity conditions has marginal stationary distribution, $P(X_u, \theta \mid X_0)$, that can be trivially transformed into $P(\theta \mid X_0)$. The MCMC techniques require rather high computational resources but they are very popular due to their universality, good scaling characteristics, capability to take into account non-observable variables, low estimation errors, and possibilities for parallel computing. Correctness of the conditions necessary for parameter estimation can be analyzed after computing the parameters. After computing the parameters, we get estimates of the random process influencing the process under study as follows:

$$\hat{\varepsilon}(k) = e(k) = y(k) - \hat{y}(k),$$

and analyze its statistical characteristics indicating correctness of the parameter values.

MODEL DIAGNOSTICS: SELECTION OF THE BEST MODEL OF THE CANDIDATES CONSTRUCTED

At this stage model adequacy is analyzed that includes the steps given below.

a) Visual study of residual graph, $e(k) = y(k) - \hat{y}(k)$, where $\hat{y}(k)$ is dependent variable estimate computed with the model constructed. The graph should not exhibit outliers and long periods of time with large values of residuals (i.e., long periods of model inadequacy). In the case of constructing GLM such model characteristics can result from random selection of distribution law for dependent variable without substantial argumentation. It can result in substantial

deviations of forecast estimates computed with the model from actual observations.

b) The estimated model residuals should not be auto-correlated. To analyze the possible autocorrelation level it is necessary to compute autocorrelation function (ACF) and partial ACF (PACF) for the series, $\{e(k)\}$, and analyze statistical significance of the functions with Q-statistics. Besides, the level of residuals autocorrelation can be additionally tested with Durbin–Watson statistic, $DW = 2 - 2\rho$, where, $\rho = E[e(k)e(k-1)]/\sigma_e^2$, is correlation between neighboring values of residuals; σ_e^2 is variance of the residual process; if $\rho = 0$, autocorrelation between residuals is absent, and ideal value for the DW statistic is 2.

c) Statistical significance of the model parameter estimates can be tested with Student $t-$ statistic, $t = [\hat{a} - a^0]/SE_{\hat{a}}$, where $\hat{a}$ is parameter estimate; a^0 is zero-hypothesis regarding the estimate; $SE_{\hat{a}}$ is standard error for the estimate. Usually all the computer systems that include functionality for time series analysis provide all necessary information for a user regarding analysis of statistical significance of parameter estimates.

d) Determination coefficient, $R^2 = \text{var}[\hat{y}]/\text{var}[y]$, where, $\text{var}(\hat{y})$ is variance of dependent variable estimated via the model constructed; $\text{var}(y)$ is actual sample variance of dependent variable. The ideal value of the coefficient is, $R^2 = 1$, when the values of variance in nominator and denominator are the same. This statistical parameter can be interpreted as a measure of information contained in a sample and in a model. From this point of view R^2 compares volume of information represented by a model to the volume of information represented by data sample used for model constructing.

e) The sum of squared errors should take minimum value for the best candidate model constructed:

$$\sum_{k=1}^{N} e^2(k) = \sum_{k=1}^{N} [\hat{y}(k) - y(k)]^2 \to \min_{\hat{\theta}}.$$

f) For estimating constructed model adequacy Akaike information criterion (AIC) is used:

$$AIC = N \ln\left(\sum_{k=1}^{N} e^2(k)\right) + 2n,$$

and Bayes–Schwarz criterion:

$$BSC = N \ln\left(\sum_{k=1}^{N} e^2(k)\right) + n \ln(N),$$

where $n = p + q + 1$, is a number of estimated model parameters (p is a number of parameters for auto-regression part; q is a number of parameters for the moving average part; 1 is added in a case when the bias, a_0, is added). Both criteria exhibit minimum values for the best model among the candidates estimated.

g) Besides the statistics mentioned, Fisher statistic $F \sim R^2/(1-R^2)$ is applied that may show adequacy of a model as a whole after testing it for statistical significance.

To formulate statistical inference regarding the model constructed Bayesian factor, $BF(i, j)$, is also used that is represented by the ratio of posterior probabilities to prior [12; 13; 14]. The criteria of maximum marginal density of the distribution, $p(x \mid M_i)$, is used that corresponds to the following condition: $BF(i, j) > 1$.

Correct application of the methodology presented provides a possibility for constructing adequate mathematical model in the class of generalized linear models if the statistical/experimental data hired corresponds to the requirement of system representation and contain necessary information.

EXAMPLE 1. GLM MODEL CONSTRUCTING

Structure of statistical data hired for model constructing is shown in Table 2. The data shows loss due to car insurance for three selected regions of Ukraine.

Table 2. The structure of statistical data

N	Characteristic of data	Value of the characteristic
1	Power of sample	9546
2	Dependent variable	Loss due to insurance
3	Region where police was sold	Kyiv, Crimea, Odesa
4	Year of a car production	Starting from 2006
5	Car brend	Mitsubishi, Toyota, VAZ

For the data described in Table 2, suppose that dependent variable is normally distributed, and accept as a link log-function. The histogram for dependent variable "loss" is presented in Fig. 1.

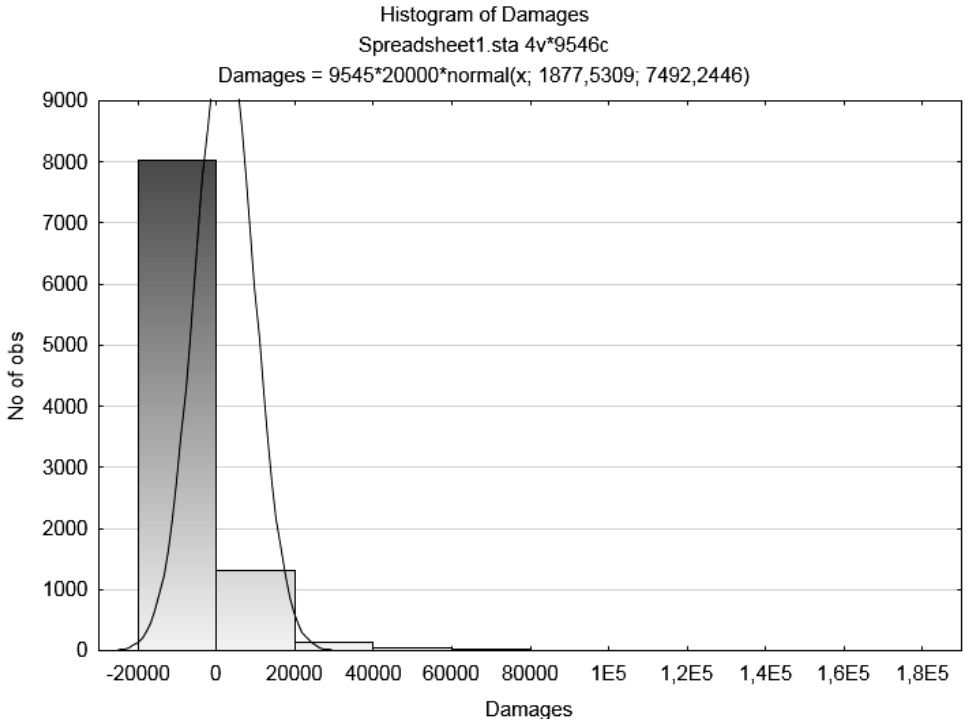


Fig. 1. The histogram for normally distributed dependent variable

The mean value of dependent variable "Loss" is 1877.531. To select the best model from the set of possible models constructed Akaike information criterion was used. Its minimum value corresponds to the best model. For log-normal model value of the criterion is 20.666. The drawback of the criterion is that its estimate asymptotically overestimates true value with non-zero probability. An alternative is Hannan–Quinn criterion that is based upon minimizing of corresponding sum instead of value itself.

Information Schwarz criterion usually selects the best model with a number of parameters that does not exceed number of parameters in the model selected by the Akaike criterion. The Schwarz criterion is asymptotically more reliable, and the Akaike criterion has a tendency to bias closer to selection of parametric models. However, for the example under consideration the values of Akaike, Hannan–Quinn, and Schwarz criteria are practically equal; that is why any of them is suitable for use. The standard deviation for dependent variable, "Loss", is 7492.245.

The Likelihood Ratio Test is used for testing restrictions regarding parameters of statistical model, estimated with sample statistical data. If the value of *LR*-statistic exceeds critical value from χ-squared distribution at given significance level then the restrictions are neglected, and the model without restrictions is hired; otherwise, the model with restrictions is used.

The variance calculated shows how much the mean random value deviates from mathematical expectation; here it is important that this is quadratic value. At the same time the variance itself is not very convenient for practical analysis because it has quadratic dimensionality of a random value. That is why standard deviation is used further on as a measure of risk. From the point of view of financial analysis, the standard deviation is more important because the mean deviation of insurance results from expected return shows actual financial results and insurance risk. As a risk measure can also be used mean absolute deviation (mad). In practice the value of standard deviation is greater than mean absolute deviation but these values have the same order, and here the following relation holds: $mad = 0.7979 \cdot S$. The result of forecasting loss and risk is presented in Table 3. The relative error of forecasting with log-normal model amounted to 1.06%, what is high quality result for the model hired. The total forecasted loss amounts to 18111231.380, and actual loss was 17921032.581, what supports the fact that the variable "Loss" is normally distributed. The proposal to use logarithmic link function is acceptable for subsequent analysis of alternative models.

Table 3. Forecasting results with log-normal model

Value	Total loss	Mean	Std. deviance	Max	Min	Variance
Actual	17921032.58	18577.53	7492.24	151771.4	0.00	
Forecasted	18111231.38	18976.45	9379.910	14010.97	634.0	49.535

Thus, the log-normal model is acceptable but it is not the best for the dataset used. That is why the search for the better model was continued. Other modeling results are given in Table 4.

Table 4. Other modeling results with application of alternative data distributions and link functions

N	Model characteristics		Total forecasted loss	Actual total loss	Deviation of actual data from forecasts	Risk of loss
	Dependent variable distribution	Link function				
1	Gamma	LOG	102008320.905	17921032.581	84087288.32	1.003
2	Normal	LOG	18111231.380		190198.799	0.495
3	Poisson	LOG	17921032.574		0.007	0.547
4	Normal	identity	17921032.589		0.009	0.532

Results of model parameters estimation with the use of classic and Bayesian approach are presented in Table 5.

Table 5. Results of model parameters estimation with the use of classic and Bayesian approach

N	Classic approach			Bayesian approach			
	Mean	Std. deviance	Variance, %	Mean	Std. deviance	Variance, %	R-squared
1	11805.7	15358.1	130.091	11804.34	15247.23	128.669	0.897
2	1897.45	939.91	49.535	1897.294	939.94	49.4	0.998
3	1877.53	1027.56	54.73	1877.301	1027.552	55.679	0.998
4	1877.53	999.302	53.224	1876.909	999.751	53.809	1.0

Thus, after constructing models using various proposals regarding initial distributions of dependent variable and link function the following results of computing were achieved:

1) the best data approximating model turned out to be the model with initial Poisson distribution of dependent variable and logarithmic link function what is proved by the forecasted total loss of 17921032.574 with practically zero errors of forecasting;

2) the value at risk for the models constructed was in the range 40–60%, what requires extra measures for minimizing this value;

3) the model based on normal distribution of data showed the risk about 49–50 %, but there were observed substantial deviations of forecasts from actual data;

4) comparing the models based upon normal data distribution with logarithmic and identity link functions it was established that Akaike criterion accepts the same value of about 20.66; that is why the model selection should be based upon forecasts of total loss;

5) the model based upon gamma-distribution with logarithmic link function showed maximum deviation from actual data (84087288.324) and maximum error: about 100%;

6) it is clear from results in Table 4 that the quality of model estimation with Bayesian approach are close to classic estimation with the maximum likelihood method but with better estimates of variance and standard deviation.

Thus, it can be concluded that the model based upon Poisson distribution and exponential link function is better for practical use from the point of view of forecasting, risk estimation, and parameter estimation.

BAYESIAN NETWORK APPLICATION TO OPERATIONAL RISK ANALYSIS

The next approach to operational risk estimation that we consider in the study is based upon application of Bayesian networks (BN) that is constructed for a commercial enterprise. The bases for the model create expert estimates though statistical data can also be hired successfully. Comparing to regression approach, BN provide a possibility for taking into consideration not only risk dependence on risk factors, but also the dependences of between risk factors [15]. The probabilistic inference also can be computed using incomplete data.

Mathematically, BN is directed acyclic graph nodes of which correspond to risk factors and environment variables, and edges correspond to possible relations between nodes. Each BN is also described by the set of conditional distributions characterizing risk factors and environment variables.

Formally, BN is a triple $\mathbf{N} = <\mathbf{V}, \mathbf{G}, \mathbf{J}>$, where $\mathbf{V}$ stands for a set of network variables; $\mathbf{G}$ is directed acyclic graph nodes of which represent variables of the process being modeled; $\mathbf{J}$ represents joint probability distribution for the network variables, $\mathbf{V} = \{X_1, X_2, ..., X_n\}$ [15; 16]. Here also Markov condition should be fulfilled: each network variable does not depend on all other variables except for the parent nodes of the variable. To construct the BN model first mutual information between all nodes is computed to discover possible dependences, then optimal network structure is estimated using appropriate criterion, say minimum description length (MDL) that is analyzed and re-computed at each iteration of the structure learning algorithm.

For two independent discrete events D and S Bayes theorem can be formulated as follows:

$$p(D \mid S) = \frac{p(D)\, p(S \mid D)}{p(S)}.$$

Suppose that state variable D can accept two values: D_t is true probability that system under study accepted one of possible state; D_f its opposite state. These two probabilities sum to 1 independently on the value that accepts variable S:

$$p(D_t \mid S) + p(D_f \mid S) = 1.$$

Apply to this equality Bayes theorem:

$$\frac{p(D_t)\, p(S \mid D_t)}{p(S)} + \frac{p(D_f)\, p(S \mid D_f)}{p(S)} = 1,$$

or

$$p(S) = p(D_t)\, p(S \mid D_t) + p(D_f)\, p(S \mid D_f).$$

Thus, known estimate $p(S)$ can be eliminated from consideration. In this example variable D has two states only, however, it is clear that $p(S)$ can be eliminated with arbitrary number of states for D.

Procedure of constructing Bayesian network. To estimate the degree of dependency between two random variables x^i and x^j the following expression was proposed in [17]:

$$MI(x^i, x^j) = \sum_{x^i, x^j} p(x^i, x^j) \cdot \log\left(\frac{p(x^i, x^j)}{p(x^i) \cdot P(x^j)} \right).$$

This is mutual information that shows volume of information contained in the variable x^i about variable x^j. It accepts nonnegative values only, $MI(x^i, x^j) \geq 0$, and in the case of complete independence between the variables it accepts zero value $MI(x^i, x^j) = 0$, because $p(x^i, x^j) = p(x^i) \cdot P(x^j)$ and

$$\log\left(\frac{p(x^i, x^j)}{p(x^i) \cdot P(x^j)} \right) = \log\left(\frac{p(x^i) \cdot P(x^j)}{p(x^i) \cdot P(x^j)} \right) = \log(1) = 0.$$

In the case BN contains N nodes, estimation of $MI(x^i, x^j)$ for all possible pairs of x^i and x^j it is necessary to perform $\frac{N \cdot (N-1)}{2}$ computations under condition $MI(x^i, x^j) = MI(x^j, x^i)$.

The use of minimum description length (MDL) principle to estimate BN structure. According to the Shannon coding theory, for known distribution $P(X)$ for variable X the optimal length of code for transmitting specific value of random variable X tends to the value of $L(x) = -\log P(x)$ [18]. The source entropy $S(P) = -\sum_x P(x) \cdot \log P(x)$ is minimum expected length of the coded message. Any other code that is based upon incorrect representation of information source will result in longer expected length of message. In other words, better model of message source results in more compact code of data.

In the problem of BN learning as a data source is functioning some unknown distribution function $P(D \mid h_0)$, where $D = \{d_1, ..., d_N\}$ is dataset; h is hypothesis regarding probabilistic origin of the data; $L(D \mid h) = -\log P(D \mid h)$ is empirical risk that is additive regarding the number of observations and proportional to empirical errors [19]. The difference between $P(D \mid h_0)$ and modeled distribution, $P(D \mid h)$, according to the Kullbak–Leibler measure is determined as follows:

$$\left| P(D \mid h) - P(D \mid h_0) \right| = \sum_D P(D \mid h_0) \cdot \log \frac{P(D \mid h_0)}{P(D \mid h)} =$$

$$= \sum_D P(D \mid h_0) \cdot \left| L(D \mid h) - L(D \mid h_0) \right| \geq 0.$$

In fact, this is the difference between hypothetically expected data code length and minimum possible length. This difference is always nonnegative and it equals to zero in the case of complete equality between distributions only. The MDL principle in its general formulation declares: from the set of possible candidate models it is necessary to select the one that describes data in short and completely (without loss of information) [19].

Thus, in general form the problem of forming MDL is formulated as follows: for the set of learning data $D = \{d_1, ..., d_n\}$, $d_i = \{x_i^{(1)} x_i^{(2)} ... x_i^{(N)}\}$ (here upper in-

dex is variable number); n is a number of observations; each observation includes N ($N \geq 2$) variables $X^{(1)}, X^{(2)}, ..., X^{(N)}$. Each j-th variable ($j = 1,...,n$) has $A^{(j)} = \{0,1,..., \alpha^{(j)} - 1\}$ ($\alpha^{(j)} \geq 2$) states, and each structure, $g \in G$, of BN is represented by N sets of parents, ($\Pi^{(1)}, ..., \Pi^{(N)}$), i.e., for each node, $j = 1,...,n$, $\Pi^{(j)}$, this is a set of parent variables such, that $\Pi^{(j)} \subseteq \{X^{(1)}, ..., X^{(N)}\} \setminus \{X^{(j)}\}$ (any node cannot be a parent for itself, what means that graph does contain cycles). Thus, MDL of the structure $g \in G$ for n observations, $x^n = d_1 d_2 ... d_n$, is computed as follows:

$$L(g, x^n) = H(g, x^n) + \frac{k(g)}{2} \cdot \log(n),$$

where $k(g)$ is a number of independent conditional probabilities in the net structure, g, and $H(g, x^n)$ is empirical entropy:

$$H(g, x^n) = \sum_{j \in J} H(j, g, x^n), \quad k(g) = \sum_{j \in J} k(j, g),$$

where MDL for j-th node is computed as follows:

$$L(j, g, x^n) = H(j, g, x^n) + \frac{k(j, g)}{2} \cdot \log(n);$$

$k(j, g)$ is a number of conditional probabilities for j-th node:

$$k(j, g) = (\alpha^{(j)} - 1) \cdot \prod_{k \in \phi(j)} \alpha^k,$$

where $\phi(j) \subseteq \{1,.., j-1, j+1,..., N\}$ is such a set that $\Pi^{(j)} = \{X^{(k)} : k \in \phi^{(j)}\}$.

The empirical entropy of j-th node is computed as follows:

$$H(j, g, x^n) = \sum_{s \in S(j,g)} \sum_{q \in A^{(j)}} -n[q, s, j, g] \cdot \log \frac{n[q, s, j, g]}{n[s, j, g]},$$

$$n(s, j, g) = \sum_{i=1}^{n} I(\pi_i^{(j)} = s); \quad n[q, s, j, g] = \sum_{i=1}^{n} I(x_i = q, \pi_i^{(j)} = s),$$

where $\pi^{(j)} = \Pi^{(j)}$ means that $X^{(k)} = x^{(k)}, \forall k \in \phi^{(j)}$; the function $I(E) = 1$, if the predicate, $E = true$, otherwise $I(E) = 0$. The simplified BN structure learning algorithm is shown in Fig. 2; it is performing in a cycle analysis of all possible acyclic network structures. The structure g^* contains optimal network structure. The optimal structure accepts minimum value for the function $L(g, x^n)$.

BN learning algorithm using MDL principle

1. $g^* \leftarrow g_0 (\in G)$;
2. for $\forall g \in G - \{g_0\}$: if $L(g, x^n) < L(g^*, x^n)$ then $g^* \leftarrow g$;
3. solution is contained in g^*.

EXAMPLE 2. BN MODEL CONSTRUCTION FOR OPERATIONAL RISK ESTIMATION AT COMMERCIAL ENTERPRISE

Consider commercial enterprise that provides for the services in entertainment sphere, and has obligations to other commercial enterprises for established possible services. These enterprises are banks, owners of cloud technologies and servers, and outsourcing companies that perform some separate functions for the entertainment enterprise. The obligations suppose transfer of money to other enterprises. That is why the main enterprise income should cover all obligations. Suppose that the main enterprise income is equal to one million euros.

Certainly, during the period of the main enterprise functioning various risk may arise that influence negatively its income. A substantial part of the risks create operational risks. To have a possibility for covering the risks the enterprise should direct a part of its capital (income) on covering operational risks, this is so called Capital at Risk (CaR). The use of statistical data in such case is practically impossible due to impossibility of separating operational loss from other risks. That is why the basic source of information for modeling operational risks provide the expert estimates.

According to recommendations provided for researchers by Basel there exist four basic factors influencing generation of operational risks:

- Risk related to working personnel: this type of risk is adhered to errors and incorrect actions of personnel, their insufficient qualification, sometimes overload at work place, possible irrational organization of working activities etc.
- Risk related to systems and work technologies. This type of risk is related to insufficiently high information technologies used at enterprise, the hardware system characteristics can be inadequate to performing operations, data processing techniques can be inadequate or data itself may exhibit unaccepted characteristics.
- Processes risk: this is risk of loss due to errors in the processes of performing various financial operations and corresponding calculations, forming reports, price forming operations etc.
- Risk induced by external influences: this type of risk is generated by events from external environment; for example, by legislative changes, general policy of state, economy, as well as the risks of external physical influences into activities of financial organization.

Each of the four named factors is also influenced by some other reasons of coming to being operational risk that can be analyzed by the experts of specific commercial enterprise. Table 6 illustrates the nodes (variables) selected for estimating the structure of Bayesian network. The nodes and respective values were selected using expert estimates.

The central point regarding the Bayesian network is touching upon variable Loss that is described by the quantitative values in euros as follows: {0, 1000, 10000, 50000, 100000}. The specific numbers are used to define the variable because of necessity to perform linear interpolation. The numbers were selected by an expert taking into consideration the loss at enterprise due to operational risks that took place earlier. Using the analysis being performed the enterprise managers can estimate the loss say within a week/month and accumulate necessary capital to cover the loss.

Table 6. Nodes of Bayesian network and their possible values

N	Variable name	Name of node in network	Possible values
1	Possible loss	Loss	0, 1000, 10000, 50000, 10000 ($)
2	Risk of external environment	Risk of External Environment	Low, Medium, High
3	Risk of working personnel	Risk of Staff	Low, Medium, High
4	Risk of hardware system used	Risk of System	Low, Medium, High
5	Risk of information processes	Risk of Process	Low, Medium, High
6	Staff qualification level	Qualified staff	Low, Medium, High
7	Compliance of	Compliance to staff	Low, Medium, High
8	Possible changes in legislation	Changes in legislation	Low, Medium, High
9	Reliability of payment systems used and partner banks	Reliability of payment systems and partner banks	Low, Medium, High
10	Possible hacker attacks and viruses	Hacker attacks and viruses	Low, Medium, High
11	Loss of information	Loss of information	0, 50, 100 (%)
12	Using workflow automation	Using workflow automation	Low, Medium, High
13	Possible crashes of servers	Server crashes	Low, Medium, High
14	Network failures	Network failures	Low, Medium, High
15	Using of firewall in the system	Using of Firewall	Yes, No
16	Using of UPS to support system	Using of UPS	Yes, No

The Loss variable is influenced by the four basic factors mentioned above: *Risk of Staff, Risk of System, Risk of Processes* and *Risk of External Environment.* These factors have the following levels: *Low*, *Medium*, and *High*, that can be interpreted as probability of realizing specific type of risk.

All other nodes and connections between them were identified on the basis of expert data, and reflect real reasons that influence specific type of risk:

- *Qualified staff* — describes the level of competence and qualification of personnel working at the enterprise. This node influences Risk of Staff, and Loss of information.
- *Compliance staff* — is related to correspondence of actually working staff to the necessary staff for functioning of an enterprise. This node has substantial influence on the variable Risk of Staff.
- *Changes in legislation* — describes possible level of legislative changes in a country. This is substantial reason for possible material loss because the possibilities for entertainment at any moment can be restricted or forbidden at all. This node influences the Risk of External Environment.
- *Reliability of payment systems and partner banks* — this factor can be a substantial reason for possible loss due to low quality of financial transactions (prolonged transaction time and possible errors). The factor also influences the Risk of External Environment.
- *Hacker attacks and viruses* — describes the level of hacker and viruses attacks to the enterprise servers. This influence can result in server reloading and system rejects. The node influences the Risk of External Environment.
- *Loss of information* — the loss is usually described in percentage form and reflects the portion of information that could be lost in the process of functioning of an enterprise. The node influences the Risk of Processes.

- *Using workflow automation* — this factor describes the level of usage specialized software that provides the possibility for performing workflow automation. This approach influences the Risk of Processes.
- *Network failures* — describes the level of computer network reliability that usually does not depend on enterprise (under study) itself. The level can be changed by changing network provider. This factor influences the Risk of External Environment.
- *Using of Firewall* — this factor has two states: Yes and No, and is described by probabilities that are related to the time intervals when firewall is used or not. It influences on Hacker attacks and viruses and Server crashes.
- *Using of UPS* — this factor is also characterized by two states and it is also described by the probabilities related to time intervals when uninterrupted power source was used or not. In its turn the factor influences on possible Server crashes.
- *Server Crashes* — the factor describes possible level of server crashes of an enterprise and depends on the Use of firewall and Uninterrupted power supplies. The reason is that loss of power or virus/hacker/ddos attack may result in server turnoff. The factor influences on the Risk of External Environment.

The next step of estimating Bayesian network structure is filling in probability tables for nodes. All probabilities in this example were estimated using expert estimates. On the other side, availability of necessary statistical information at the enterprise (i.e., storing the information during its functioning) will give the possibility for the fast replacing expert estimates by actual data. This way work adaptive risk analyzing systems.

For the nodes that do not have parents it will be unconditional probabilities for each possible state. The nodes that have parents require estimating conditional probabilities. The problem is that number of required conditional probabilities is growing fast with growing number of parents and their possible states. In case of discrete variables the number of values in conditional probability tables (CPT) can be estimated via the expression:

$$Count = NodeStates \cdot \prod_{i}^{Parents} ParentStates_i ,$$

where *Count* is the number of values necessary for CPTs; *NodeStates* is number of states for a specific node under consideration; *Parents* is the number of parents for a specific node being analyzed; $ParentStates_i$ represents the number of states for –*i*-th parent. Using this formula it was estimated that for the node Loss it were required about 400 values. Example of CPT for the node of "Loss of information" is given below in Table 7.

Table 7. Conditional probabilities for the node "Loss of Information"

Using workflow automation	**Low**	**Low**	**Low**	**Medium**	**Medium**	**Medium**	**High**	**High**	**High**
Loss of information	0%	50%	100%	0%	50%	100%	0%	50%	100%
Low	0.25	0.1	0.03	0.55	0.42	0.31	0.95	0.8	0.6
Medium	0.3	0.22	0.1	0.25	0.3	0.38	0.04	0.12	0.23
High	0.45	0.68	0.87	0.2	0.28	0.31	0.01	0.08	0.17

As a result of the model parameter estimation the BN was constructed for operational risk analysis for commercial enterprise given in Fig. 2.

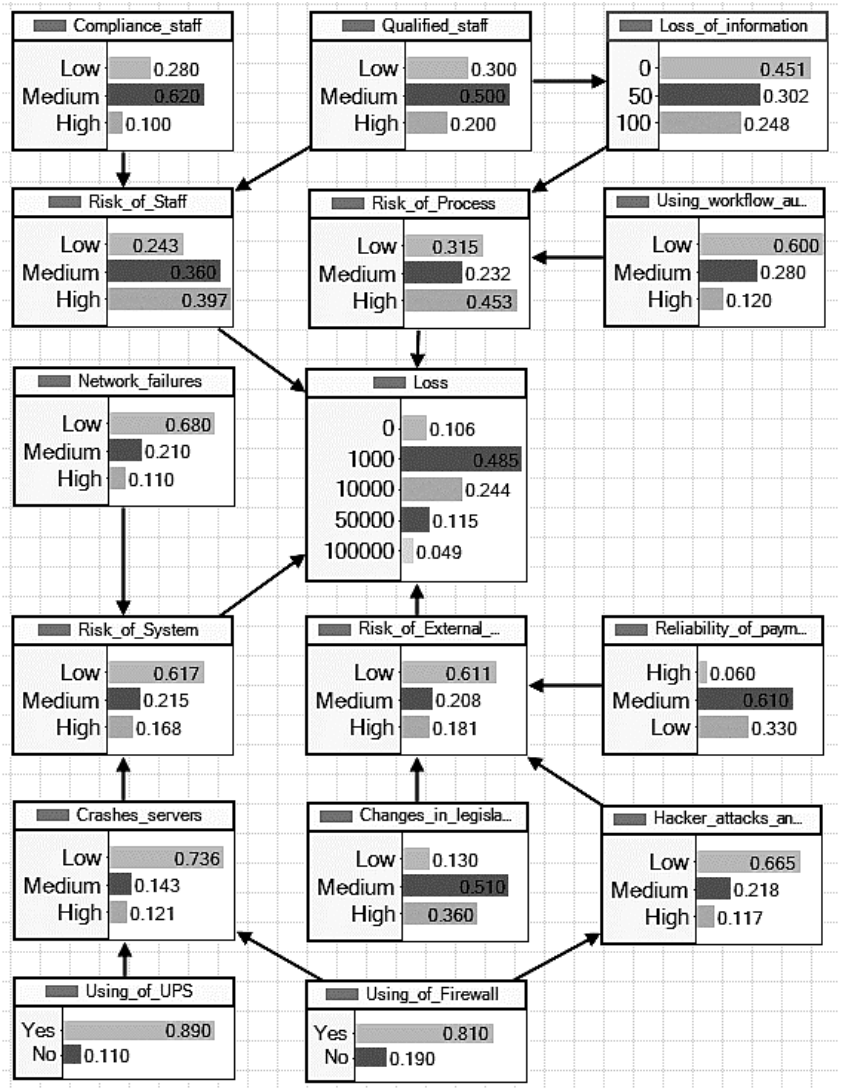


Fig. 2. Possible probabilistic inference estimation using the BN constructed

ANALYSIS OF THE RESULTS

Table 8 contains the capital necessary for covering the operational risks: the 90, 95 i 99 quintiles reflect the probabilities 0.9, 0.95 and 0.99, correspondingly.

T a b l e 8. Capital required for covering the risks at selected level of confidence

Quintile, %	90	95	99
CaR, $	38476	56833	97554

It can be easily tracked that Capital-at-Risk (CaR) is substantially growing with growing confidence level from 90 to 99. According to the proposal by the Basel Basic Indicator Approach (BIA) acceptable level of risk for banking sphere

is 15%. Thus, the enterprise with monthly income of about one million euros should keep CaR of 150000 euros. This is about three times as much than the result achieved in the example above. This result stresses that enterprises in banking sphere should use the Advanced Measurement Approaches (AMA) for analysis of operational risks. The volume of CaR can be reduced substantially this way and available extra capital could be used more rationally to achieve higher income.

CONCLUSIONS

The reasons for emerging financial operational risks in financial organizations were presented. It was shown that an urgent task for the organizations is development and practical use of the risk analysis and management systems on the basis of modern system analysis approach, providing the possibility for constructing mathematical models, more specifically probabilistic models of Bayesian type that provide the possibility for identification and taking into consideration uncertainties related to the random nature of multiple events in financial analysis.

Improved systemic multistep methodology of constructing models for financial processes and financial risks of arbitrary origin was proposed. An example of the methodology application is given that shows effectiveness of the methodology in application to operational risk analysis. Application of the methodology proposed for modeling financial processes with the use of generalized linear models and Bayesian parameter estimation guaranties high quality of risk estimation with minimum errors.

Also simplified methodology was proposed for constructing models of risk in the form of Bayesian networks using the notion of mutual information between variables of network, and criteria of estimation the quality of model structure on the basis of minimum description length. An example of BN model constructing was given for a commercial enterprise on the basis of expert estimates. The model was used for estimating distribution of loss in the case of operational risk availability. Analysis of estimation results provided by Bayesian network proves that models of the type selected can be successfully applied for estimating possible loss in financial organizations.

In future studies it is reasonable to create specialized commercial intellectual decision support system for performing mathematical modeling and financial risks estimation. The system should provide a possibility for the use of statistical data, expert estimates and generated (simulated) variables of continuous type convenient for the use in Bayesian framework. Also possible data uncertainties should be identified and taken into consideration to improve the results of all the stages of expert estimation and statistical data processing. All the stages of data and expert estimates processing should be controlled by appropriate sets of statistical quality analysis criteria to reach high quality of intermediate and final results of computing.

REFERENCES

1. *International Convergence of Capital Measurement and Capital Standards. A Revised Framework. Comprehensive Version*. Basel: Basel Committee on Banking Supervision, Bank for International Settlements, 2006, 158 p.

2. M.Z. Zgurovsky, N.D. Pankratova, *System Analysis: Theory and Applications*. Berlin: Springer-Verlag, 2007, 446 p.
3. M. Denuit, J. Dhaene, M. Goovaerts, and R. Kaas, *Actuarial Theory for Dependent Risks: Measures, Orders and Models*. Chichester (West Sussex): John Wiley & Sons, Inc., 2005, 440 p.
4. A. Cruz, *Modeling, Measuring and Hedging Operational Risk*. London: Wiley, 2002, 346 p.
5. *Operational Risk Regulation, Analysis and Management*; edited by Carol Alexander. New York: Pearson Education Limited, 2003, 369 p.
6. Pavel Shevchenko, *Modelling Operational Risk Using Bayesian Inference*. New York: Springer, 2011, 302 p.
7. S.O. Dmitrov, *Modeling operational risk of commercial bank*. Sumy: The State Higher Educational Institution "UABS NBU", 2010, 264 p.
8. A. Frachot, P. Georges, and T. Roncalli, *Loss Distribution Approach for operational risk*. Available: http://thierry-roncalli.com/download/lda.pdf
9. P.I. Bidyuk, S.V. Trukhan, "Development and Usage of Information System for Analysis and Forecasting Financial Processes," *System Sciences and Cybernetics*, no. 1, pp. 26–37, 2013.
10. P.I. Bidyuk, M.S. Rubets, "Information System for Modeling and Estimation of Operational Risks Using Artificial Intelligence Methods," *System Sciences and Cybernetics*, no. 1, pp. 5–29, 2015.
11. W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*. Boca Raton (Florida): Chapman & Hall/CRC, 1998, 486 p.
12. C. Alexander, *Bayesian Methods for Measuring Operational Risk*. Available: http://www.icmacentre.ac.uk/pdf/bayesian.pdf
13. X. Hao, *Operational Risk Control of Commercial Banks Based on Bayesian Network*. Atlantis Press, 2013, pp. 913–918.
14. Y.K. Yoon, *Modelling operational risk in financial institutions using bayesian networks*. London: University of London, 2003, 83 p.
15. M. Neil, N.E. Fenton, and M. Tailor, "Using bayesian networks to model expected and unexpected operational losses," *Risk Analysis*, pp. 34–57, 2005.
16. Alan Stuart, Keith Ord, *Kendall's Advanced Theory of Statistics: Volume 1, Distribution Theory*. Wiley, 1994, 700 p.
17. C.K. Chow, C.N. Liu, "Approximating discrete probability distributions with dependence trees," *IEEE Transactions on information theory*, vol. 14, issue 3, 6 p., May 1968.
18. T. Minka et al., "Infer.NET 2.6.," *Microsoft Research*, Cambridge, 2014. Available: http://research.microsoft.com/internet
19. M.Z. Zgurovsky, P.I. Bidyuk, O.M. Terentiev, and T.I. Prosyankina-Zharova, *Bayesian Networks in Decision Support Systems*. Kyiv: Publishing Company "Edelweiss" LLC, 2015, 300 p.

Received 15.10.2023

INFORMATION ON THE ARTICLE

Petro I. Bidyuk, ORCID: 0000-0002-7421-3565, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: pbidyuke_00@ukr.net

Oxana L. Tymoshchuk, ORCID: 0000-0003-1863-3095, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: oxana.tim@gmail.com

Liudmyla B. Levenchuk, ORCID: 0000-0002-8600-0890, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: lusi.levenchuk@gmail.com

ОЦІНЮВАННЯ ОПЕРАЦІЙНОГО РИЗИКУ З ВИКОРИСТАННЯМ МЕТОДОЛОГІЇ СИСТЕМНОГО АНАЛІЗУ / П.І. Бідюк, О.Л. Тимощук, Л.Б. Левенчук

Анотація. Фінансові ризики — популярна тема досліджень практичної необхідності використання їх математичних моделей, оцінювання можливих втрат у багатьох напрямах людської діяльності, прогнозуванні і відповідних управлінських рішеннях у фінансовій та інших сферах, де капітал, облігації, біржові акції та інші активи успішно використовуються. Фінансові процеси характеризуються складною формою розвитку у часі, що потребує застосування відповідних методів, прийомів та процедур моделювання, оцінювання ризиків, прогнозування і підтримання/прийняття рішень. Для вирішення цих проблем застосовано підхід на основі системного аналізу як унікальну та універсальну методологію дослідження. Фінансові ризики, зокрема операційні, які розглядаються у роботі, класифікуються як нелінійні нестаціонарні процеси, що потребують застосування належних методів для аналізу і досить складного аналітичного опису для оцінювання і прогнозування можливих втрат. Результати аналізу операційних ризиків отримано у формі системної методології, нових моделей, побудованих з використанням статистичних даних, регресійного аналізу та байєсівських методів, а також оцінок втрат, отриманих за допомогою створених моделей. Побудовано моделі і запропоновано системний підхід до аналізу фінансових процесів, придатний для практичного використання за умови, що користувачі матимуть належні статистичні дані та експертні оцінки.

Ключові слова: фінансовий операційний ризик, математична модель, статистичні дані, методологія системного аналізу, оцінювання втрат, система підтримання прийняття рішень.

UDC 338.3:330.46
DOI: 10.20535/SRIT.2308-8893.2024.1.05

STUDY ON THE PROFITABILITY OF AGRICULTURAL ENTERPRISES IN UKRAINE DURING THE RUSSIAN MILITARY INVASION OF UKRAINE

OLGA TSESLIV, TAMARA DUNAIEVA, JULIA YERESHKO, OLEKSANDR TSESLIV

Abstract. This paper examines the effectiveness of grouping agricultural enterprises according to the wheat harvested area and assesses their profitability. We have developed linear and non-linear regression equations to predict the income for said groups of enterprises. The methodology is designed for cases when future market prices are probabilistic in nature. With the help of the developed methodology, it is possible to calculate the necessary production volumes in the conditions of price fluctuations. We have used the Goldfeld–Quandt parametric test to test the model for heteroscedasticity. Calculations show that agricultural holdings are indeed inefficient, and preference should be given to enterprises with medium crop areas. Application of the Lagrange multipliers method when solving the problem of agricultural enterprise optimization makes it possible to increase profitability. The case of price risk, when future market prices are not deterministic, is considered. Therefore, it is necessary to be guided by two criteria when making managerial decisions: to maximize the expected total net income and to minimize the variance of the total net income.

Keywords: economic and mathematical models, heteroscedasticity, models of regression analysis, profitability, income, linear regression, nonlinear model, full-scale russian invasion of Ukraine.

INTRODUCTION

Wheat is one of the most important crops for food security worldwide. Growing wheat is also a source of income for the considerable part of Ukraine's population. Among the agricultural crops in Ukraine, wheat occupies more than half of the sown area. In the recent years, nation had entered the top ten major grain producing countries and became one of the world's leading exporters of wheat (Fig. 1). Moreover, wheat exports to Africa, Southeast Asia and the Western Hemisphere

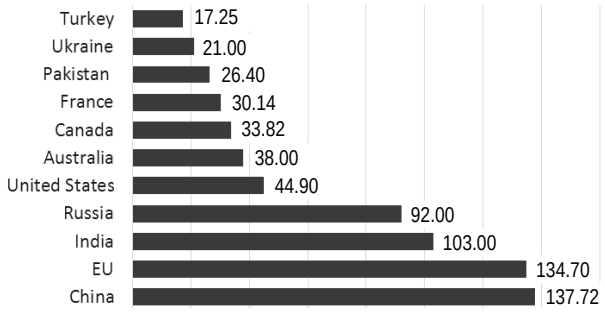


Fig. 1. Leading 10 wheat producers worldwide in 2022/2023 (in 1.000 metric tons)

were expected to increase in 2022/23. Unfortunately, as of May 2023 we can observe negative effects of war, as the economy exported 41.6 million tons of grain in the 2022/23 July-June season within the Black Sea Grain Initiative and overall wheat export is predicted to peak at 24 million tons. The negative trend is expected to remain, as the prognosis for the following period 2023/24 season stands at the threshold of 26 million tons [1].

A starting point in understanding the Ukraine's grain producing sector, is to analyse and group the enterprises by the size of the harvested area of wheat in 2021. In percentage terms, the ratio is as follows (Fig. 2). There are 24.016 wheat-growing enterprises in total. 61.6% of those are the small enterprises, in particular those with harvesting area of up to 100 hectares. Their aggregated volume of production is 1.986 thousand tons of wheat, which is accordingly 7.7% of the total volume of wheat production in Ukraine. There are only 123 enterprises with a total area of more than 3.000 hectares, the volume of production of which is 2.852.1 thousand tons, which is 11.1% of the total volume of wheat production in Ukraine [2].

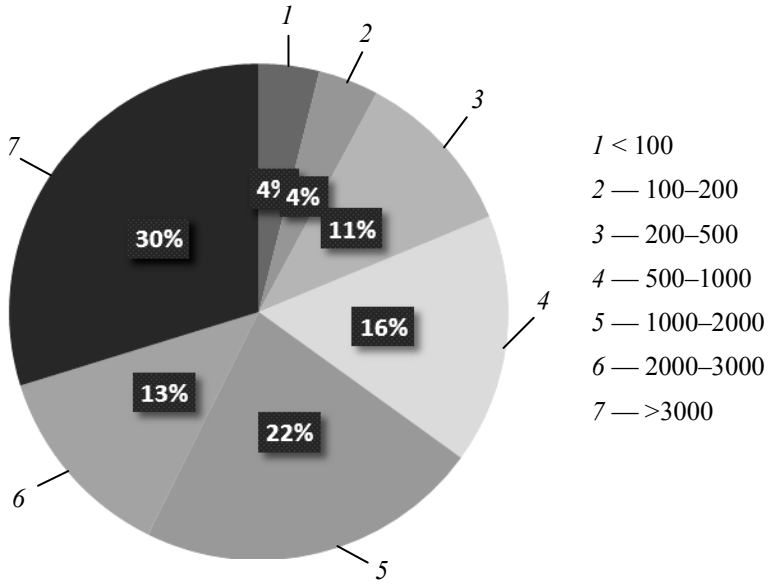


Fig. 2. Enterprises grouping by wheat production (thousand tons) in 2021
Source: Source: authors' elaboration on the data of [3]

The yield of wheat differs according to various groups of enterprises as follows: enterprises with an area of up to 100 hectares have a yield of 35 t/ha; with an area of 500-1000, respectively, 50.3 cwt/ha; for the enterprises with more than 3.000 hectares, the yield is 65.4 t/ha. By volume of harvested wheat to the total volume of production: enterprises with an area of up to 100 hectares collect 4%; with an area of 500–1000, respectively, 16.6%; with an area of 1000–2000 — 23.4%; with an area of 2.000–3.000 hectares — 12.8%; enterprises over 3.000 hectares 28.7%.

And the trend of overall wheat production in the previous years shows constant increase up to 2022 (Fig. 3), that is explained by the consequences of russian aggression towards Ukraine. As war escalates, under constant bombing and shelling, Ukrainian farmers are not able to harvest the grain. Moreover, fields, that are

under russian occurrence are unreachable and inaccessible to them. Those crops, if being harvested at all, are at the disposal of occupiers and are being expropriated by them and consequently, their sale bypasses Ukraine. Hence, 25687.2 thousand tons of wheat were grown in 2021. Ukrainians consume about 20% of local wheat — the rest is being exported. That is, 20549.76 thousand tons were exported in 2021.

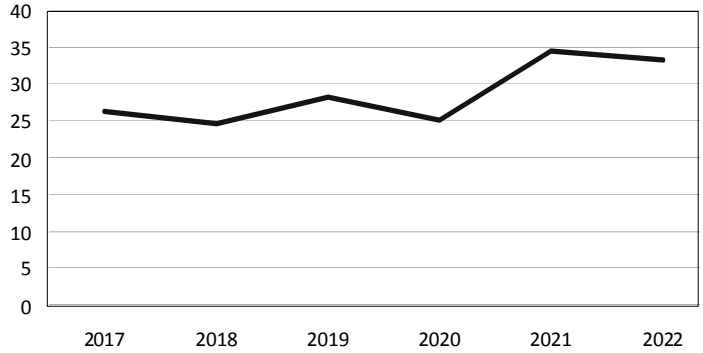


Fig. 3. Wheat harvest in Ukraine in previous years (1 cwt per hectare)
Source: authors' elaboration on the data of [4]

Due to the use of the higher-quality seeds, modern technology and plant protection products, this year, despite the war, farmers managed to harvest a quite proficient harvest. All this contributed to the increase in productivity. For the first time in 20 years, there was no drought in Ukraine. The best harvesting results were observed in Vinnytsia (6.7 million tons), Chernihiv (6.2 million tons), and Poltava regions (5.7 million tons) [4]. According to the groupings of agricultural producers, the productivity is as follows (Fig. 4): with an area of 100 hectares, as a percentage of the total 61%: with an area of 200–-500 hectares — 11.8%; with an area of 500–1000, respectively, 8%; with an area of 1000-2000 — 5.5%; with an area of 2.000–3.000 hectares — 1.6%; enterprises over 3.000 hectares 1.5%.

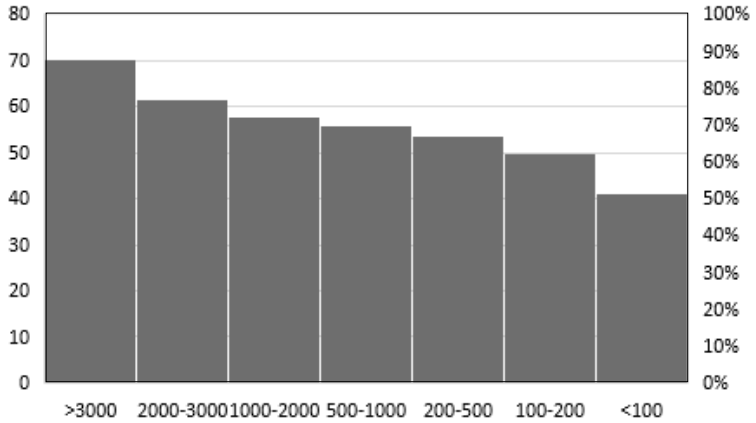


Fig. 4. Enterprises grouping by wheat yield in 2021 (centner per 1 hectare)
Source: authors' elaboration on the data of [3]

Compared to 2019, in 2022 the overall grain production trends were as follows: soybean harvest increased by 17; 40 million tons of corn were harvested; as well as 32 million tons of wheat (8 million tons more than in 2020). According to the press service of the Ukrainian Grain Association, the record wheat harvest guarantees the country's food security. By the end of 2021, all countries of the

world expect to harvest a record 2.8 billion tons of grain, the Food and Agriculture Organization of the United Nations (FAO) predicts.

Exports of agricultural products, during the 11 months of the war, amounted to $24.4 billion, which is 22% higher than last year, according to customs data. Ukrainian products are mostly imported to China, India and the Netherlands [5]. For 2023, the National Bank of Ukraine (NBU) forecasts consumer inflation (December to December) at the level of 18.7% and an average annual rate of 20.3% (in 2022 it was 26.6% and 20.2%, respectively, in 2021 — 10% and 9.4%, in 2020 — 5% and 2.7%). According to NBU estimates, the average unemployment rate in Ukraine in 2023 will be 26.1% (in 2022 — 25.8%, in 2021 — 9.8%, in 2020 — 9.5%) [21].

Optimizing in the war time

Theoretical issues of land management were developed in the works of: Schluter G. [6, p. 747], who investigated the extent to which an increase in the minimum wage will affect food prices; Sánchez-Fung J. [7], who deals with the stability of macro indicators; Rasool H. [11, p. 87], who investigates the relationship between rural wages and food inflation; Dorward A. [12, p. 633], food safety specialist; as well as Sitikantha P. [10, p. 244], Bhattacharya R. [11, p.144], Duckett T. [12, p. 17], Bhattacharya R. [13, p. 146], Nwaolisa E. [17] Wilson I. [19, p. 505].

During the war, the problem of optimizing the groupings of agricultural producers is relevant both at the micro and macro level, since a significant part of the Ukraine's agrarian-industrial complex products is being exported. Moreover, agricultural production is characterized by seasonality. The manufacturer periodically builds up stocks of products that will possibly be sold in the future. The proposed methodology is designed for cases when the price of agricultural products is probabilistic. Its use will allow the owner of agricultural products to protect his economic interests to the maximum possible extent. The methods of substantiating management decisions in conditions of risk and uncertainty are to be considered from Taha H. [10].

Literature review. Models and metods

Since enterprises growing wheat have different areas, and accordingly will receive different income, it is advisable to monitor their income. To do this, to begin with, the variance of the residuals for different groups of enterprises was calculated, that is, it was checked for heteroskedasticity. The presence of heteroskedasticity causes a violation of the properties of model parameter estimates when calculating them using the least squares method. Therefore, it is always necessary to study this phenomenon and, if it exists, to use the generalized least squares method (Aitken's method) to estimate the model parameters. Here, in order to determine heteroskedasticity we have used the Goldfeld–Quandt parametric test.

In an econometric model that characterizes the dependence of consumption costs on income, the variance of the residuals may change for observations that belong to different groups of the population in terms of income.

We have developed a technique for optimizing the grouping of enterprises in wartime. Since future market prices are not deterministic, the decision must be guided by two criteria: maximizing expected total net income and minimizing the

variance of total net income. The Lagrange method is used to solve the problem. The developed methodology is designed for the probabilistic nature of future market prices for products and makes it possible to take inflation into account.

Let us consider the method for optimizing the grouping of enterprises in the conditions of the price risk and price fluctuations, in a wartime. Since future market prices are not deterministic, the decision must be guided by two criteria: maximizing expected total net income and minimizing the variance of total net income. To solve the problem, the Lagrange method is used. Hence, using the Lagrange method we develop a methodise, designed for cases when the price of agricultural products has a probabilistic nature. We find the maximum income according to formula:

$$\bar{z}_{\max} = \sum_{t=1}^{T} (p_{t^-} c_t) x_t ,$$

$\bar{z}_{\max}$ — maximum income; $p_{\bar{t}}$ — the sale price of a product unit at the given time t; c_t — costs associated with the storage of a product unit until the time $t = \overline{1,T}$; $\sigma(z)_{\min}$ — minimal dispersion; a — production volume; T — planning period. Then the maximum dispersion is:

$$\sigma(z)_{\max} = a\sigma_t^* .$$

Next, the worst values of the criterion indicators are calculated for a set of effective variants of the calendar plan:

$$\sigma(z)_{\min} = \frac{a}{\sqrt{\sum_{t=1}^{T} \frac{1}{\sigma_t^2}}};$$

$$\bar{z}_{\min} = \frac{a}{\sqrt{\sum_{t=1}^{T} \frac{1}{\sigma_t^2}}} \sum_{t=1}^{T} \frac{(\bar{p}_t - c_t)}{\sigma_t^2} .$$

Therefore, we calculate the range of variation of criterion indicators:

$$\bar{z}_{\min} \le \bar{z}_0 < \bar{z}_{\max};\ \ \sigma(z)_{\min} < \sigma_0 \le \sigma_{\max} .$$

The optimal calendar plan for the sale of stocks of agricultural products is determined. This plan $x^* = (x_1^*, \ldots, x_T^*)$ is computed by solving the convex programming problem, given that s — development criteria:

$$s \to \max ,$$

$$\sum_{t=1}^{T} (\bar{p}_t - c_t) x_t \ge \bar{z}_0 + s(\bar{z}_{\max} - \bar{z}_0), \tag{1}$$

$$\sum_{t=1} \sigma_t^2 x_t^2 \le \sigma_0^2 - s(\sigma_0^2 - \sigma^2(z)_{\min}), \tag{2}$$

$$\sum_{t=1}^{T} x_t = a, \tag{3}$$

$$x_t \geq 0, \quad t = \overline{1,T}. \tag{4}$$

It should be added that the optimal value s^* will show whether the acceptable levels of criterion indicators chosen by the product owner were true (at $s^* \geq 0$) or not (at $s^* < 0$).

MODELLING THE DEPENDENCE OF ENTERPRISE'S INCOME ON THE HARVESTED AREA

To build this model, the original data set, which includes 7 observations, is used. These data and calculations based on them are given in Table 1. Based on the nature of the relationship between the value of income of enterprises from the harvested area, it can be assumed that the variance of the residuals is not constant for each observation, that is, there may be a phenomenon of heteroscedasticity. Therefore, in order to choose the right method for estimating the parameters of the model, it is necessary to check whether heteroscedasticity is inherent in the given input data.

T a b l e 1. Input data and calculations

Enterprises by area, thousand hectares	Volume of production, thousand centners	Crop yields, 1 centner per hectare	Production expenditures per ton, UAH	Price per ton including costs, UAH	Revenue, UAH billion
up to 100.00	1986.0	39.2	630.37	7299.63	14.498
10001–200.00	2010.3	45.2	539.21	7390.78	14.858
200.01–500.00	5044.5	47.0	508.08	7421.92	37.44
500.01–1000.00	6000.8	48.3	494.38	7435.62	44.62
1000.01–2000.00	5806.4	49.4	493.27	7436.73	43.181
2000.01–3000.00	1987.1	49.6	489.98	7440.02	14.785
over 3000.00	2852.1	49.1	464.14	7465.86	21.294

Source: authors' developments.

We consider the price per ton of wheat to be a constant value, which in 2022 is 7930 UAH. The cost of wheat production on one hectare of land is 22 thousand UAH. Thus, with an average yield of 5.37 centners/hectare (Fig. 4), the cost of 1 ton of production will be equal to 4.1 thousand UAH. At the selling price of wheat of 7931 UAH/ton, the profit per hectare will be 20 thousand UAH, which will provide the profitability of more than 90%. Data for the Goldfeld–Quandt parametric test is given in Table 2.

T a b l e 2. Data for the Goldfeld–Quandt parametric test

Y	X	X^2	XY	$\hat{Y}$	$(Y-\hat{Y})$	$(Y-\hat{Y})^2$
14.498	1986	3944196	28793.028	13.57	0.928	0.861184
14.858	2010.3	4041306.09	29869.0374	13.74	1.118	1.249924
37.44	5044.5	25446980.25	188866.081	34.98	2.46	6.0516
44.62	6000.8	36009600.64	267755.696	41.67	2.95	8.7025
43.181	5806.4	33714280.96	250726.158	40.31	2.871	8.242641
14.785	1987.1	3948566.41	29379.2735	13.57	1.215	1.476225
21.294	2852.1	8134474.41	60732.6174	19.63	1.664	2.768896

Source: authors' developments.

Identification of variables: $Y = f(X,u)$, where Y — dependent variable (revenue); X — independent variable (area size); u — stochastic component.

Model specification:

$$Y = a_0 + a_1 X + u\,,\quad \hat{Y} = \hat{a}_0 + \hat{a}_1 X\,,\quad u = Y - \hat{Y}\,.$$

Using the Goldfeld–Kwandt algorithm, we determine the presence of heteroscedasticity. We find C observations that are in the middle of the population:

$$\frac{C}{n} = \frac{4}{15},\quad \frac{C}{7} = \frac{4}{15},\quad C = \frac{4*7}{15},\quad C \approx 2\,.$$

Then $n_1 = 3$, $n_2 = 3 = 3$. Let us estimate an econometric model for the population $n_1 = 3$. Let us quantitatively estimate the model parameters based on OLS:

$$\begin{cases} n\hat{a}_0 + \hat{a}_1 \sum x = \sum y, \\ \hat{a}_0 \sum x + \hat{a}_1 \sum x^2 = \sum xy. \end{cases}$$

For each model we find the sum of squares of residuals:

$$S_1 = u_1' u_1 = \sum (Y_1 - \hat{Y}_1)^2\,,\quad S_2 = u_2' u_2 = \sum (Y_2 - \hat{Y}_2)^2\,,\quad S_1 = 8.16,\quad S_2 = 12.48.$$

Finding the criterion R:

$$R = \frac{S_1}{S_2};\quad R = \frac{12.48}{8.16} = 1{,}53;\quad F_{\text{tabl}} = 6.67\,.$$

Because of $R < F_{\text{tabl}}$ grouping of enterprises by the size of the harvested area of wheat heteroscedasticity is absent. If there is no heteroscedasticity, the least squares method can be applied (Table 2).

Calculating the coefficients of linear pairwise correlation (r_{xy}) and determination (R_1^2): $R_1^2 = r_{xy}^2 = 1^2 \approx 0.9999$.

Most often, in case of a system of linear equations, the linear method of least squares is used. For our case, the linear regression formula is obtained: $Y = = 0.0075 * C_3 - 0.1215$. Using this formula, we can calculate exactly how much the income will increase when the area increases by one unit. In our case, an increase in area by 1 hectare will lead to an increase in income by 0.13 (billion UAH).

Further on, we find and analyse the power regression equation $\hat{y} = ax^b$, for the inputs x_i and y_i from Table 3.

Table 3. Additional values for calculating linear pairwise correlation (r_{xy}) and determination (R_1^2) coefficient

i	x_i	y_i	$\hat{y}_i$	$(x_i - (\bar{x}))$	$(x_i - (\bar{x}))^2$	$\exists_i$	$\exists_i^2$	A_i	***c***	$\Delta\exists_i^2$
1	1986	14.4	14.6	-1683.5	2834364.65	-0.188	0.035	0.001	–	–
2	2010.3	14.8	14.8	-1659.5	2754129.91	-0.007	0.0001	0.0005	0.1811	0.0328
3	5044.5	37.4	37.4	1374.9	1890467.86	-0.051	0.002	0.0014	0.0438	0.0019
4	6000.8	44.6	44.6	2331.2	5434693.25	-0.001	0	0	0.0496	0.0025
5	5806.4	43.1	43.1	2136.8	4566097.39	0.008	0.0001	0.0002	0.0105	0.0001
6	1987.1	14.7	14.6	-1682.4	280662.037	0.090	0.008	0.006	0.081	0.006
7	2852.1	21.2	21.1	-817.45	668236.18	–	0.068	0.028	–	0.047

Source: authors' developments.

Assessing the significance of regression and correlation parameters

In order to estimate the significance of regression and correlation parameters, let's find the average value $\bar{x} = 3669{,}56$; make a table of additional values, where $\varepsilon_i = y_i - \hat{y}_i$; $\varepsilon_i = y_i - \hat{y}_i$; $A_i = \left| \frac{y_i - \hat{y}_i}{y_i} \right|$.

The average approximation error (Table 3):

$$\overline{\mathrm{A}} = n \sum \frac{(y_i - \hat{y}_i)}{y_i} 100\% = 0{,}43\%,$$

F — Fisher criteria $F_{\text{tabl}} = 6.67$.

As it is known, the least squares method is a method of finding an approximate solution of an over determined system, which is used in regression analysis. The most commonly linear least squares method is used in the case of a system of linear equations. For our case, we obtain the linear regression formula: $Y = 0.0075 * C3 - 0.1215$.

Using this formula, we can calculate, how much income will increase with an increase in area per unit. In our case, an increase in area by 1 hectare will lead to an increase in income by 0.13(billion UAH).

Let us find and analyse the power regression equation $\hat{y} = ax^b$, for data x_i and y_i for Table 4.

T a b l e 4. Auxiliary variables for calculating power regression

i	x_i	y_i	$\ln x_i$	$\ln^2 x_i$	$\ln y_i$	$\ln x_i \ln y_i$
1	1986	14.498	7.59	57.66	2.67	20.30
2	2010.3	14.858	7.61	57.84	2.69	20.52
3	5044.5	37.44	8.52	72.69	3.62	30.88
4	6000.8	44.62	8.69	75.68	3.79	33.04
5	5806.4	43.181	8.66	75.11	3.76	32.63
6	1987.1	14.785	7.59	57.67	2.69	20.45
7	2852.1	21.294	7.95	63.29	3.06	24.33
Total	25686.9	190.676	56.64	459.97	22.31	182.18

Source: authors' developments.

Let us calculate the coefficients a and b of the power regression equation by the known formulas:

$$b = \frac{n \sum (\ln x_i * \ln y_i) - \sum \ln x_i * \sum \ln y_i}{n \sum \ln^2 x_i - (\sum \ln x_i)} = \frac{7 * 182.19 - 56.64 * 22.31}{7 * 459.97 - 56.64^2} \approx 1.0064;$$

$$a = \exp\left(\frac{1}{n} \sum \ln y_i - \frac{b}{n} \sum \ln x_i \right) = \exp\left(\frac{1}{7} * 22.31 - \frac{1.0064}{7} * 55.64 \right) \approx 0.007.$$

Nonlinear regression equation $\hat{Y} = 0{,}007 * x^{1{,}0064}$.

Let's compare the calculations with linear and power regression.

The data obtained by calculations, linear and power regression actually match. Thus, both relational models can be used.

Needless to say, that in order to create economic and mathematical models of agricultural enterprises in market conditions, it is necessary to take into account all factors, in particular: land resources, labour resources, fixed assets, movable assets, financial resources, information resources, costs. The combination of these indicators makes it possible to forecast income more accurately.

Thus, the study showed the efficiency of different groups of enterprises in terms of the size of the harvested area of wheat. Their profitability and income were estimated. Since the best results are obtained from enterprises with a harvested area of 200-2000 thousand hectares (Fig. 5), this is obviously the best option for an agricultural enterprise. Nevertheless, in a war time and respective post-war times, small businesses become more manoeuvrable and accessible to the population.

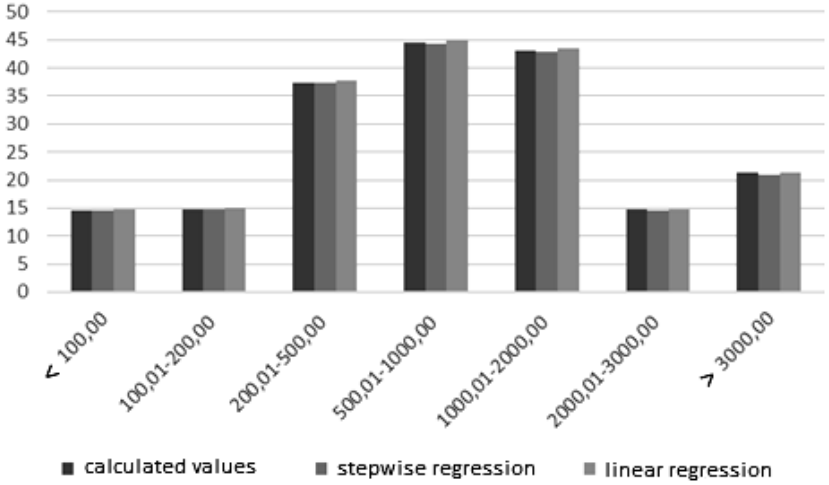


Fig. 5. Comparison of linear and power regression calculations
Source: authors' developments

General problem of conditional optimization

The general problem of conditional optimization with equality constraints is reduced to the problem of unconditional minimization using the Lagrange function, which is written in the form:

$$F(x_1,x_2,x_3,x_4,x_5,x_6,x_7,\lambda_1,\lambda_2) = f(x_1,x_2,x_3,x_4,x_5,x_6,x_7) -$$

$$-\lambda_1(u(x_1,x_2,x_3,x_4,x_5,x_6,x_7) - \lambda_2(u(x_1,x_2,x_3,x_4,x_5,x_6,x_7 - z_{\max})).$$

An indispensable condition for using expression (2) is that the number of restrictions must be less than the number of variables. If this condition is not fulfilled, then there is no optimization problem, since the number of connections of variables exceeds their number. Thus, the task is reduced to finding the minimum of the function:

$$\frac{\partial F(x_j)\lambda_k}{\partial x_j} = 0.$$

A necessary condition for the minimum of the function (2) is the equality of its gradient to zero, which leads to the system of equations. Consequently, solving this system of equations leads to finding unknown quantities.

The calculations were conducted in MathCAD, using the following formula for the maximum income:

$$(x_1, x_2, x_3, x_4, x_5, x_6, x_7, \lambda_1, \lambda_2) =$$

$$= 12800 \times 12 + 490 \times 22 + 80 \times 32 + 516 \times 42 + 566 \times 52 + 733 \times 62 + 2800 \times 72$$

variance:

$$u(x_1, x_2, x_3, x_4, x_5, x_6, x_7) =$$

$$= 7290 \times 1 + 7421 \times 2 + 7435 \times 3 + 7436 \times 5 + 7440 \times 6 + 7440 \times 6 + 7465 \times 7 .$$

Solving the equation, we obtain optimal production volumes. We observe that there has been a shift towards enterprises with an area of 200.01-500.00 thousand hectares. Taking into account inflation, in order to achieve the received income, it is necessary to obtain the production volumes indicated in Table 5.

T a b l e 5 . Auxiliary variables for calculating power regression

N	Enterprises by area, thousand hectares	Volume of production, thousand centners	Price per ton incl. costs, UAH	Revenue, UAH billion	Standard price deviation, UAH σ	σ^2	Optimum production volumes	Optimum production volumes considering inflation
1	up to 100.00	1986.0	7299.63	14.498	–113.3	12838.5	920	1220
2	100.01–200.00	2010.3	7390.78	14.858	–22.2	490.9	2550	560
3	200.01–500.00	5044.5	7421.92	37.44	8.9	80.7	16500	18000
4	500.01–1000.00	6000.8	7435.62	44.62	22.7	514.5	2530	2800
5	1000.01–2000.00	5806.4	7436.73	43.181	23.8	566.1	1940	1940
6	2000.01–3000.00	1987.1	7440.02	14.785	27.1	733.5	1200	630
7	over 3000.00	2852.1	7465.86	21.294	52.9	2800.8	0.18	1

Source: authors' developments

Solving the system of partial differential equations (1)–(4), we obtain the optimal volumes of production, with the inflation taken into account. The results are shown in Table 5. We observe that there has been a shift towards enterprises with an area of 200.01–500.00 thousand hectares. That is, enterprises with an area of 200.000–500.000 hectares are the most efficient and provide the optimal volumes of wheat production. It becomes clear, considering this study, that enterprises with an area of 200–500, 500–1000 and 1000–2000 thousand hectares are efficient and competitive, and it is most expedient to develop precisely them.

The problem was solved using the Lagrange method and the probabilistic nature of prices was taken into account. The developed technique makes it possible to calculate the necessary volumes of production during the period of inflation. Calculations show that agricultural holdings are not efficient, and preference should be given to enterprises with average cultivated areas.

CONCLUSIONS

In a current paper we study the grouping of enterprises by the size of the harvested area of wheat. Wheat is one of the most important sources of income for

the part of the population of Ukraine. Over the years, Ukraine has had agricultural land (71%), 78% of which is arable land; 97.2% of agricultural land is systematically used for economic purposes.

Here we examine the income of various groups of enterprises, using The Goldfeld–Quandt parametric test in order to determine heteroskedasticity.

Our analysis of yield, volume of production, income of each group of enterprises showed, that the best results are obtained by enterprises with an area of 200–2000 thousand hectares. Meaning, that large agricultural holdings proved to be not efficient, and preference should be given to enterprises with medium areas.

Since the phenomenon of heteroscedasticity was not detected, a linear regression formula was constructed using the method of least squares. For comparison, a power-law regression equation was found and analysed.

In order to predict the income for each group of enterprises by the size of the harvested area of wheat, we have developed equations of linear and nonlinear regression.

The scientific novelty of the work is that the equations of linear and nonlinear regression were developed to predict the income of each group of enterprises by the size of the harvested area of wheat. Application of the Lagrange method multipliers when solving the problem of optimization of agricultural enterprises makes it possible to increase profitability.

Developed models can be used to analyse the income and profitability of agricultural producers.

REFERENCES

1. M. Shahbandeh, “Global leading wheat producing countries 2022/2023,” *Statista*. 2023, March 10. Available: https://www.statista.com/statistics/237912/global-top-wheat-producing-countries/
2. *Ministry of Agrarian Policy and Food of Ukraine*. [Official website]. 2022. Available: https://minagro.gov.ua/en
3. *FAOSTAT (2022). Statistical database of food and agriculture organization of the United Nations*. [Online]. Accessed on: March 15, 2022. Available: http://faostat.fao.org/faostat/
4. *Statistical Yearbook of Ukraine, 2021. State Statistics Service of Ukraine*. [Online]. Accessed on: February 2, 2022. Available: http://www.ukrstat.gov.ua
5. *Socio-economic characteristics of households in Ukraine 2021. State Statistics Service of Ukraine*. [Online]. Accessed on: February 2, 2022. Available: http://www.ukrstat.gov.ua
6. C. Lee, G. Schluter, and B. O’Roark, “How Much Would Increasing the Minimum Wage Affect Food Prices?” *Agriculture Information Bulletin*, vol. 3, p. 747, 2000. doi: 10.22004/ag.econ.33598.
7. J. Sánchez-Fung, “Institutions for macroeconomic stability: a review of Monetary policy in low financial development countries,” *Macroeconomics and Finance in Emerging Market Economies*, vol. 15(1), 2022. Available: https:// doi.org/10.1080/17520843.2022.2096913
8. I.S. Danilova and O.M. Hetmanets, “Method of determining the degree of freshness of snail meat by the photometric method (Patent of Ukraine No. 128984),” Ministry of Economic Development and Trade of Ukraine. 2018. Retrieved from https://sis.ukrpatent.org/uk/search/detail/237538/
9. V.V. Kasianchuk and N.M. Bohatko, “Method for determining degree of freshness of beef and pork meat (Patent of Ukraine No. 59032),” Ministry of Education and Science of Ukraine. 2003. Retrieved from https://uapatents. com/3-59032-sposib-viznachennya-stupenya-svizhosti-yalovichini-ta-svinini.html

10. H. Taha, *Operations Research: An Introduction Pearson*; 10th edition, 2016, 848 p.
11. H. Rasool and Md. Tarique, "Testing the link between Rural Wages and Food Inflation in India A Co-integration Approach," *IOSR 2 (IOSR-JBM)*, vol. 20(1), pp. 87–98, 2018. doi: 10.9790/487X-2001018798.
12. A. Dorward, "The short- and medium-term impacts of rises in staple food prices," *Food Security*, vol. 4, pp. 633–645, 2012. Available: https://doi.org/10.1007/s12571-012-0210-3
13. P. Sitikantha, M. Silu, and R. Soumyajit, "Inflation expectations of households: do they influence wage-price dynamics in India?" *Macroeconomics and Finance in Emerging Market Economies,* vol. 13(3), pp. 244–263, 2020. Available: https://doi.org/10.1080/17520843.2020.1720264
14. D.I. Babmindra, *Obtaining positive financial benefits as a way to get the agro-industrial complex out of the crisis*. Available: http://nbuv.gov.ua/UJRN /Zik_2013_2_6
15. P. Biber, T. Duckett, "Dynamic maps for long-term operation of mobile service robots," *Robotics: Science and Systems I*, June 8–11, pp. 17–24, 2005. Available: http://dx.doi.org/10.15607/RSS.2005.I.003
16. R. Bhattacharya, A.Sen Gupt, "Drivers and impact of food inflation in India," *Macroeconomics and Finance in Emerging Market Economies*, vol. 11(2), pp. 146–168, 2018. Available: https://doi.org/10.1080/17520843.2017.1351461
17. E. Nwaolisa, "Effects of Exchange Rate Fluctuations on the Balance of Payment in the Nigerian Economy. Approach," *IOSR Journal of Business and Management (IOSR-JBM)*, vol. 5(12), p. 7576–7583, 2017. Available: https://doi.org/10.18535/ijsrm/v5i12.01
18. *United Nations Organization*. Available: http: //www.un.org/en/index.html
19. I. Wilson, "Strategic planning for the millennium: Resolving the dilemma," *Long rouge planning*, vol. 31, no. 4, pp. 507–551, Oxford etc., 2017. doi: 10.4236/ajibm.2017.74035.
20. A. Silva, A. Carrara, and N. Castro, "Inflation persistence for product groups in Brazil using the ARFIMA-GARCH model," *Macroeconomics and Finance in Emerging Market Economies*, vol. 15, issue 3, 2022. Available: https://doi.org/10.1080/17520843.2022.2080345
21. *Finbalance*. Available: https://finbalance.com.ua/news/nbu-serednorichna-inflyatsiya-v-2023-rotsi-bude-203
22. M. Usman, M.U. Rezekina, A. Baihaqi, and Srihandayani, "Analysis of export competitiveness of natural rubber from Indonesia and Thailand in the international market," *Scientific Journal of Accountancy, Management and Finance*, 1(4), pp. 220–230. doi: 10.33258/economit.v1i4.588.
23. I. Yuliadi, "A survey of agglomeration determinants in Indonesia," *Academic Journal of Interdisciplinary Studies*, 10(1), pp. 304–312, 2021. doi: 10.36941/ajis-2021-0026.
24. M. Dyadkova and G. Momchilov, *Constant market shares analysis beyond the intensive margin of external trade*. Bulgaria: Bulgarian National Bank, 2014.
25. E. Efendi et al., "Implementation of greedy algorithm for profit and cost analysis of swallow's nest processing dirty to finished products," *INFOKUM*, 10(02), pp. 849–858, 2022.

Received 04.07.2023

INFORMATION ON THE ARTICLE

Olga V. Tsesliv, ORCID: 0000-0002-8190-2502, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: ceslivolga@gmail.com

Tamara A. Dunaieva, ORCID: 0000-0001-8104-7836, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: dunaeva.toma @gmail.com

Julia O. Yereshko, ORCID: 0000-0002-9161-8820, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine and Technical University of Munich, Germany, e-mail: julia.jereshko@gmail.com

Oleksandr S. Tsesliv, ORCID: 0000-0002-8602-4673, Taras Shevchenko National University of Kyiv, Ukraine, e-mail: atsesliv@gmail.com

ДОСЛІДЖЕННЯ РЕНТАБЕЛЬНОСТІ СІЛЬСЬКОГОСПОДАРСЬКИХ ПІДПРИЄМСТВ В УКРАЇНІ ПІД ЧАС ВІЙСЬКОВОГО ВТОРГНЕННЯ РОСІЇ В УКРАЇНУ / О.В. Цеслів, Т.А. Дунаєва, Ю.О. Єрешко, О.С. Цеслів

Анотація. Досліджено ефективність групування сільськогосподарських підприємств за розміром збиральної площі пшениці та дано оцінку їх прибутковості. Розроблено лінійні та нелінійні рівняння регресії для прогнозування доходу для зазначених груп підприємств. Методику розроблено для випадків, коли майбутні ринкові ціни мають імовірнісний характер. За допомогою розробленої методики можна розрахувати необхідні обсяги виробництва в умовах коливання цін. Використано параметричний тест Голдфельда–Квандта для перевірки моделі на гетероскедастичність. Розрахунки показують, що агрохолдинги насправді неефективні, і перевагу слід віддавати підприємствам із середніми посівними площами. Застосування методу множників Лагранжа для вирішення завдання оптимізації сільськогосподарських підприємств дає змогу підвищити рентабельність. Розглянуто випадок цінового ризику, коли майбутні ринкові ціни не є детермінованими. Тому під час прийняття управлінських рішень необхідно керуватися двома цілями-критеріями: максимізувати очікуваний сукупний чистий дохід і мінімізувати дисперсію сукупного чистого доходу.

Ключові слова: економіко-математичні моделі, гетероскедастичність, моделі регресійного аналізу, прибутковість, дохід, лінійна регресія, нелінійна модель, повномасштабне вторгнення росії в Україну.

ТЕОРЕТИЧНІ ТА ПРИКЛАДНІ ПРОБЛЕМИ ІНТЕЛЕКТУАЛЬНИХ СИСТЕМ ПІДТРИМАННЯ ПРИЙНЯТТЯ РІШЕНЬ

UDC 519.925.51
DOI: 10.20535/SRIT.2308-8893.2024.1.06

HYBRID SYSTEM OF COMPUTATIONAL INTELLIGENCE BASED ON BAGGING AND GROUP METHOD OF DATA HANDLING

Ye. BODYANSKIY, O. KUZMENKO, He. ZAICHENKO, Yu. ZAYCHENKO

Abstract. The paper considers the problem of short- and middle-term forecasting in the financial sphere. To solve this problem, a hybrid system of computational intelligence based on the group method of data handling (GMDH) and bagging, as well as an algorithm for its training, is proposed. The odd stacks of the hybrid system are formed by ensembles of parallel connected subsystems. ARIMA and the GMDH-neo-fuzzy hybrid network were chosen as such subsystems. The proposed system does not require a large training data set, automatically determines the number of stacks during training, and provides online operation. The experimental investigations were conducted using the proposed hybrid system, as well as separately using ARIMA and GMDH-neo-fuzzy. The accuracy of the predictions obtained is compared, based on which the feasibility of using the proposed hybrid system is substantiated.

Keywords: hybrid system, bagging, hybrid GMDH-neo-fuzzy network, ARIMA, short- and middle-term forecasting.

INTRODUCTION

Today deep neural networks (DNN) are widely used for solving a large class of Data mining problems and, above all, classification (pattern recognition, image processing of various nature) and extrapolation (time series forecasting, natural language text processing, fault detection), due to their universal approximating properties, based mainly on the G. Cybenko [2] theorem. At the same time, DNNs are rather cumbersome constructs that contain too many configurable synaptic weights, in turn, requiring too much training data, which is not always available for solving specific real-world problems. In addition, the learning process of DNN requires quite a lot of time so that when solving Data Stream mining problems, especially when processing nonstationary data streams, the use of these systems encounters some difficulties.

It is possible to overcome these difficulties using ensemble methods [2–7], while the ensemble can contain a variety of computational intelligence systems the simplest type of elementary perceptron of F. Rosenblatt and N – Adaline [8] to the most complex DNNs.

The main problem when using the ensemble approach is to combine the results of individual members of this ensemble in order to obtain the optimal initial result in the sensing of the accepted learning criterion. For this purpose, Bagging procedures [9] can be used, minimizing the RMS error on the training sample, modified to work online.

The resulting ensemble approach and bagging system may be too complex from a computational point of view. To simplify its implementation, it should be decomposed into a number of simpler subsystems, while such decomposition can be implemented quite simply using the Group Method of Data Handling (GMDH) [10; 11].

It is interesting to note that J. Schmidhuber [12] believes that just based on GMDH the first deep learning systems were built. Subsequently, based on GMDH, neural networks and neuro-fuzzy systems were proposed, which demonstrated their accuracy and speed in solving a number of problems [13–17].

In our opinion, it is appropriate to introduce into consideration a hybrid system of computational intelligence (HSCI), built based on an ensemble approach and bagging, which would increase its architecture in the learning process based on GMDH ideas, would not require significant amounts of training selections and would be quite simple from a computational point of view.

ARCHITECTURE OF HSCI ON THE BASE OF BAGGING AND GMDH

In the Fig. 1 the architecture of the proposed system is presented.

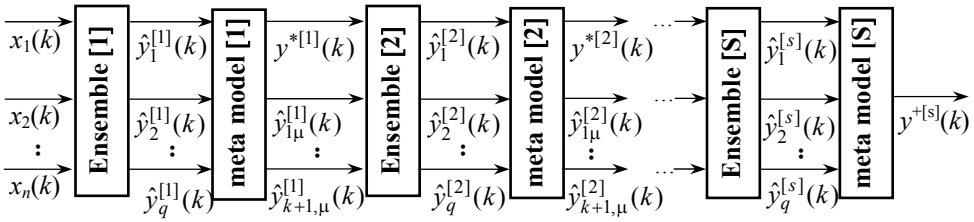


Fig. 1. Hybrid system of computational intelligence based on bagging and GMDH

The architecture of the system contains 2S sequentially-connected stacks, while odd stacks are formed by ensembles of parallel-connected subsystems that solve the same problem (recognition, prediction, etc.) and even ones are essentially learning metamodels that generalize the output signals of ensembles and form optimal results in the sense of the accepted criterion. The output signal of the first metamodel is the generalized optimal signal $y^{*1}(k)$ and $(n-1)$ output signals $\hat{y}_{i\mu}^{[1]}(k)$, $i=1,2,\ldots,n-1$ (k) "best members of the ensemble". At their core, metamodels function as selection units in traditional GMDH systems, but not only select the best results from the previous stack, but also form the optimal solution based on these results.

Further, the output signals of the first metamodel are fed to the inputs of the second ensemble, which is completely similar to the first. The outputs of the second ensemble $\hat{y}_1^{[2]}(k), \hat{y}_2^{[2]}(k),\ldots,\hat{y}_q^{[2]}(k)$ come to the second metamodel, which calculates the optimal signal $y^{*[2]}(k)$ and $(n-1)\hat{y}_{i\mu}^{[2]}(k)$ "closest" to it. The last *S*-th ensemble is similar to the first two, and the output of the last *S*-th metamodel

is $y^{*[s]}(k)$, which exactly corresponds to a priori established requirements for the quality of solving the problem under consideration.

Each of the ensembles contains q different computational intelligence systems that solve the same problem. There may still be simple neural networks such as a single-layer perceptron, radial-basis neural network (RBFN), counterpropagating neural network, etc., which do not use error backpropagation procedure for training, neuro-fuzzy systems such as ANFIS, Wang-Mendel or Takagi-Sugeno-Kang type, wavelet-neuro systems, neo-fuzzy neurons and others, the output signal of which linearly depends on the adapted parameters, which allows to use optimal speed learning algorithms.

LEARNING HSCI BASED ON BAGGING AND GMDH

The input information, on the basis of which the system is configured, is a training selection of input signals $x(1),x(2),...,x(k),...,x(N)$; $x(k)=(x_1(k),\ldots,x_i(k),\ldots,x_n(k))^T \in R^n$ and its corresponding scalar reference signals $y(1),...,y(k),...,y(N)$. On the basis of these observations, the elements of the first ensemble are tuned independently of each other, at the outputs of which q scalar signals $y_p^{[2]}(k)$, $p=1,2,...,q$, are formed, which are conveniently represented in the form of a vector $\hat{y}^{[1]}(k)=(\hat{y}_1^{[1]}(k),\ldots,\hat{y}_p^{[1]}(k),\ldots,\hat{y}_q^{[1]}(k))^T$. These signals are sent to the inputs of the first metamodel, at the outputs of which n sequences $\hat{y}^{*[1]}(k),\hat{y}_{1\mu}^{[1]}(k),\ldots,\hat{y}_{i\mu}^{[1]}(k),\ldots,\hat{y}_{n-1,\mu}^{[1]}(k)$ the main of which is $\hat{y}^{*[1]}(k)$ while others are auxiliary. The main signal of the metamodel $y^{*[1]}(k)$ is the union of the outputs of all members of the ensemble in the form of

$$y^{*[1]}(k)=\sum_{p=1}^{q} w_p^{*[1]}\,\hat{y}_p^{[1]}(k)=\hat{y}^{[1]T}(k)w^{*[1]},$$

where $w^{*[1]}=(w^{*[1]},...,w_p^{*[1]}...w_q^{*[1]})^T$ — is a vector of adapted parameters-synaptic weights on which additionally restrictions are set on unbiasedness

$$\sum_{p=1}^{q} w_p^{*[1]}=I_q^T\,w^{*[1]}=1, \qquad (1)$$

where I_q — $(q\times 1)$ is the vector of unities.

The problem of teaching the first metamodel is reduced to minimizing the standard quadratic criterion in the presence of additional constraints (1).

Thus, the problem of training the first metamodel can be solved using the standard method of penalty functions, which in this case reduces to minimizing the expression

$$J(w^{*[1]},\delta)=(Y(N)-\overline{Y}^{[1]}(N)w^{*[1]})^T(Y(N)-\overline{Y}^{[1]}(N)w^{*[1]})+\delta^{-2}(I_q^T\,w^{*[1]}-1), \qquad (2)$$

where $Y(N)=(y(1),\ldots,y(k),\ldots,y(N))^T-(N\times 1)$ is a vector, $\overline{Y}^{1}(N)=$ $=(\hat{y}^{[1]}(1),\ldots,\hat{y}^{[1]}(k),\ldots,\hat{y}^{[1]}(N))^T-(N\times q)$ is a matrix, δ is the penalty coefficient.

Minimization (2) by $w^{*[1]}$ leads to the result

$$w^{*[1]} = \lim_{\delta \to 0} w^1(\delta) = w^{LS[1]} + P^{[1]}(N)\frac{1 - I_q^T w^{LS[1]}}{I_q^T P^{[1]}(N) I_q} I_q, \qquad (3)$$

where $w^{LS[1]}$ is a standard LSM estimate:

$$w^{LS[1]} = (\overline{Y}^{[1]T}(N)\overline{Y}^{[1]}(N))^{+}\overline{Y}^{[1]T}(N)Y(N) = P^{[1]}(N)\overline{Y}^{[1]T}(N)Y(N).$$

The same result may be obtained using the indefinite Lagrange multipliers.

Introducing into consideration meta-model error

$$e^{[1]}(k) = y(k) - y^{*[1]}(k) = y(k) - \hat{y}^{[1]T}(k)w^{*[1]} = w^{*[1]T} I_q y(k) - w^{*[1]T}\hat{y}^{[1]T}(k) =$$

$$= w^{*[1]T}(I_q y(k) - \hat{y}^{[1]}(k)) = w^{*[1]T} E^{[1]}(k),$$

write Lagrange function in the form

$$L(w^{*[1]}, \lambda) =$$

$$= \sum_{k=1}^{N} w^{*[1]T} E^{[1]}(k) E^{[1]T}(k) w^{*[1]} + (I_q^T w^{*[1]} - 1) = w^{*[1]T} R(N) w^{*[1]} + \lambda(I_q^T w^{*[1]} - 1),$$

(where λ is indefinite Lagrange multiplier) and solving Kuhn–Tucker equations system

$$\begin{cases} \overline{V}_{w^{*[1]}} L(w^{*[1]}, \lambda) = 2R((N)w^{*[1]} + \lambda I_q = \vec{0}, \\ \partial L(w^{*[1]}, \lambda)/\partial\lambda = I_q^T w^{*[1]} - 1 = 0, \end{cases}$$

obtain the final result

$$\begin{cases} w^{*[1]} = R^{-1}(N) I_q (I_q^T R^{-1}(N) I_q)^{-1}, \\ \lambda = -2 I_q^T R^{-1}(N) I_q. \end{cases}$$

It was proved [3; 4] that the use of score (3) leads to results that are not inferior in accuracy to the best of the members of the first ensemble.

If observations from the training sample are processed sequentially online, it is advisable to use the least squares recurrent method in the form

$$\begin{cases} P^{[1]}(k+1) = P^{[1]}(k) - \dfrac{P^{[1]}(k)\hat{y}^{[1]}(k+1)\hat{y}^{[1]T}(k+1)P^{[1]}(k)}{1 + \hat{y}^{[1]T}(k+1)P^{[1]}(k)\hat{y}^{[1]}(k+1)}, \\ w^{LS[1]}(k+1) = w^{LS[1]}(k) + P^{[1]}(k+1)(y(k+1) - \hat{y}^{[1]T}(k+1)w^{LS[1]}(k)\hat{y}^{[1]}(k+1), \\ w^{*[1]}(k+1) = w^{LS[1]}k + 1 + P^{[1]}(k+1)(I_q^T P^{[1]}(k+1) I_q)^{-1}(1 - I_q^T w^{LS[1]}(k+1) I_q, \\ w_p^*(0) = q^{-1} \quad p = 1, 2, ..., q \end{cases}$$

or if a training sample is non-stationary we may use exponentially weighted recurrent LSM method

$$\begin{cases} P^{[1]}(k+1) = \dfrac{1}{\alpha}\left(P^{[1]}(k) - \dfrac{P^{[1]}(k)\hat{y}^{[1]}(k+1)\hat{y}^{[1]T}(k+1)P^{[1]}(k)}{\alpha + \hat{y}^T(k+1)P^{[1]}(k)\hat{y}^{[1]}(k+1)} \right), \\ w^{LS[1]}(k+1) = w^{LS[1]}(k) + \dfrac{P^{[1]}(k)(y(k+1) - \hat{y}^{[1]T}(k+1)w^{LS[1]}(k)}{\alpha + \hat{y}^{[']T}(k+1)P^{[']}(k)\hat{y}^{[']}(k+1)}\hat{y}^{[1]}(k+1), \\ w^{*[1]}(k+1) = w^{LS[1]}(k+1) + P^{[1]}(k+1)(I_q^T P^{[1]} v I_q)^{-1}(1 - I_q^T w^{LS[1]}(k+1) I_q, \\ w_p^*(0) = q^{-1} \quad p = 1, 2, ..., q, \end{cases}$$

where $0 < \alpha \le 1$ — forgetting factor.

To the parameters of the metamodel can be given meaning the levels of fuzzy membership to the optimal output signal by introducing additional restrictions on the non-negative values of these parameters, that is, in addition to the configurable parameters $w^{*[1]}$ we can also calculate the levels of this membership $\mu_p^{[1]} \ge 0,\ \ p = 1,2,...,q$.

To do this, we introduce into consideration the extended Lagrange function

$$L(\mu^{[1]}, \lambda, \rho) = (Y(N) - \overline{Y}^{[1]}(N)\mu^{[1]})^T (Y(N) - \overline{Y}^{[1]}(N)\mu^{[1]}) + \lambda(I_q^T \mu^{[1]} - 1) - \rho^T \mu^{[1]},$$

where ρ — $(q \times 1)$ is vector of non-negative indefinite Lagrange multipliers.

Using the equations system by Kuhn–Tucker

$$\begin{cases} \overline{V}_{\mu^{[1]}} L(\mu^{[1]}, \lambda, \rho) = \vec{0}, \\ \partial L(\mu^{[1]}, \lambda, \rho) / \rho\lambda = 0, \end{cases}$$

it's not difficult to get the solution in the form

$$\begin{cases} \mu^{[1]} = p^{[1]}(N)(\hat{y}^T(N)Y(N) - 0{,}5\lambda I_q + 0{,}5\rho), \\ \lambda = \dfrac{I_q^T p^{[1]}(N)\hat{y}^T(N)Y(N) - 1 + 0{,}5 I_q^T p^{[1]}(N)\rho}{0{,}5 I_q^T p^{[1]}(N) I_q} \end{cases}$$

For finding vector of non-negative Lagrange multipliers, it's reasonable to apply Arrow–Hurwitz–Uzawa procedure

$$\begin{cases} \mu^{[1]}(k+1) = w^{LS[1]}(k+1) - P^{[1]}(k+1)\dfrac{I_q^T w^{LS[1]}(k+1) - 1 + 0{,}5 I_q^T P^{[1]}(k+1)\rho(k)}{I_q^T P^{[1]}(k+1) I_q} + \\ \qquad\qquad + 0{,}5 P^{[1]}(k+1)\rho(k), \\ \rho(k+1) = P_{r+}(\rho(k) - \eta_\rho(k)\mu^{[1]}(k+1)), \end{cases} \tag{4}$$

where $P_{r+}(.)$ is a projector to positive ortant, $\eta_\rho(k)$ — learning rate, parameter.

First expression (4) after non-complex transformations may be presented in a more compact form

$$\mu^{[1]}(k+1) = w^{*[1]}(k+1) - P^{[1]}(k+1)\frac{0{,}5 I_q^T P^{[1]}(k+1)\rho(k)}{I_q^T P^{[']}(k+1) I_q} + 0{,}5 P^{[1]}(k+1)\rho(k) =$$

$$= w^{*[1]}(k+1) + 0{,}5\left(I_{qq} - \frac{P^{[1]}(k+1) I_q I_q^T}{I_q^T P^{[1]}(k+1) I_q} \right) P^{[1]}(k+1)\rho(k)$$

where instead of least squares estimates $w^{LS[1]}(k+1)$, the parameters of the metamodel $w^{*[1]}(k+1)$ are used, which simplifies the process of configuring it.

As a result of learning the first metamodel, the optimal signal $y^{*[1]}(k)$ is formed at its output, as well as q signals $\hat{y}_{p,\mu}^{[1]}(k)$ from which we choose $n-1$ $(if\ q \ge n)$ with the highest levels of fuzzy membership $\mu_p^{[1]}$, which subse-

quently in the form of $(n \times 1)$ — vector are fed to the input of the second ensemble, the outputs of which go to the inputs of the second metamodel, and so on. The process of increasing the number of ensembles and metamodels continues until the required accuracy of the last metamodel with the output $y^{*[s]}(k)$ is achieved, or the value of the criterion minimized for the bagging model begins to increase, i.e. $\bar{\varepsilon}^2(y^{*[s+1]}(k)) \geq \bar{\varepsilon}^2(y^{*[s]}(k))$.

EXPERIMENTAL INVESTIGATIONS

The experimental investigations of bagging based on GMDH were performed at the problems of short-term and middle-term forecasting of Dow–Jones Industrial average index. The dynamics of DJIA is presented in the Fig. 2.

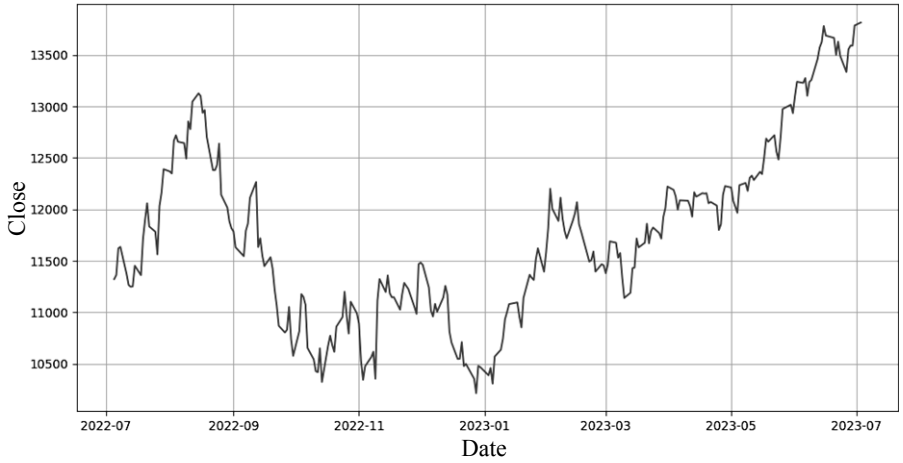


Fig. 2. Dynamics of Dow–Jones Average

The data of DJ index was taken since 05.07.22 till 03.07.23.

The correlation function of process Dow–Jones index was calculated and the correlogram was built presented in the Fig. 3.

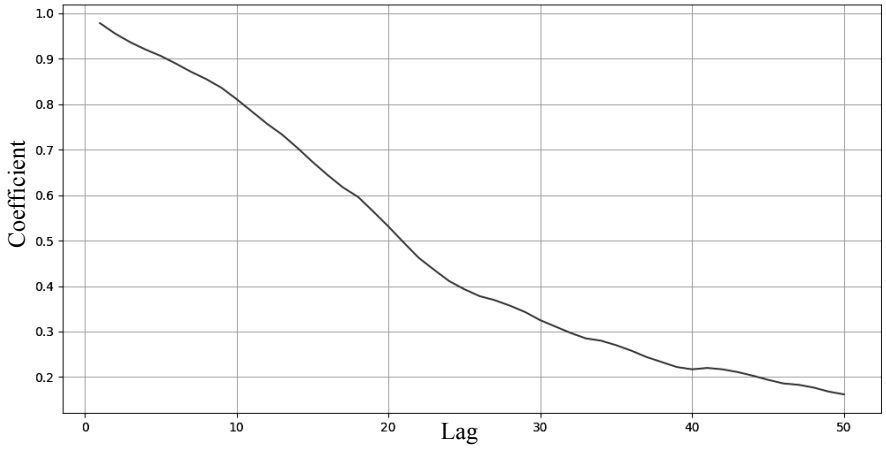


Fig. 3. Correlogram

Analyzing this correlogram we can see strong correlation between values of DJIA.

As a first model used in bagging ensemble is ARIMA. ACF function plot is presented in the Fig. 4.

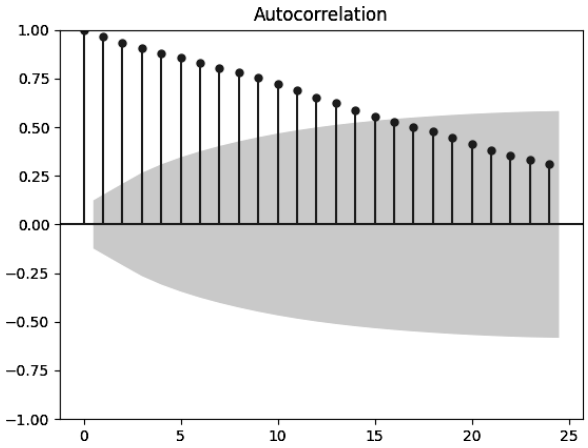


Fig. 4. ACF plot

Using differencing the process DJIA was transformed to stationary one:

Dickey–Fuller test was performed: *P*-value:1.3051439086544856e-28 < 0.05.

The experimental investigations were performed at different forecasting intervals: 1, 3, 5 (short-term) and 10, 20 days (middle-term) forecasting.

Flow-chart for 1 day forecast is presented in the Fig. 5.

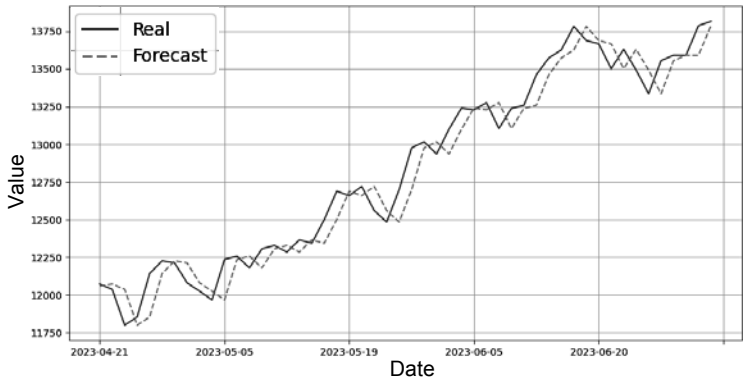


Fig. 5. Flow chart of forecast at 1 day interval for ARIMA

Flow chart of forecast by ARIMA for 5 days is shown in the Fig. 6.

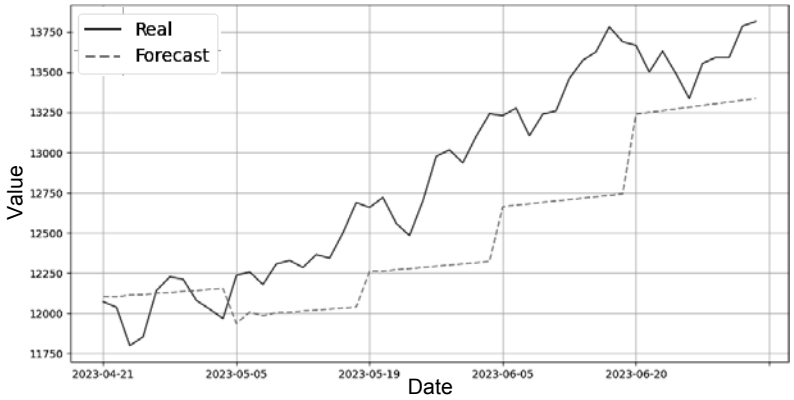


Fig. 6. Flow chart of forecast at 5 days interval for ARIMA

Next model used in ensemble is Hybrid GMDH-neo-fuzzy network. It was optimized by parameters. After that the experiments on forecasting with different forecasting intervals 1, 3, 5, 10, 20 days were performed. Some of results are presented below. Flowchart of forecast for 3 days is shown in the Fig. 7 and for 10 days in the Fig. 8.

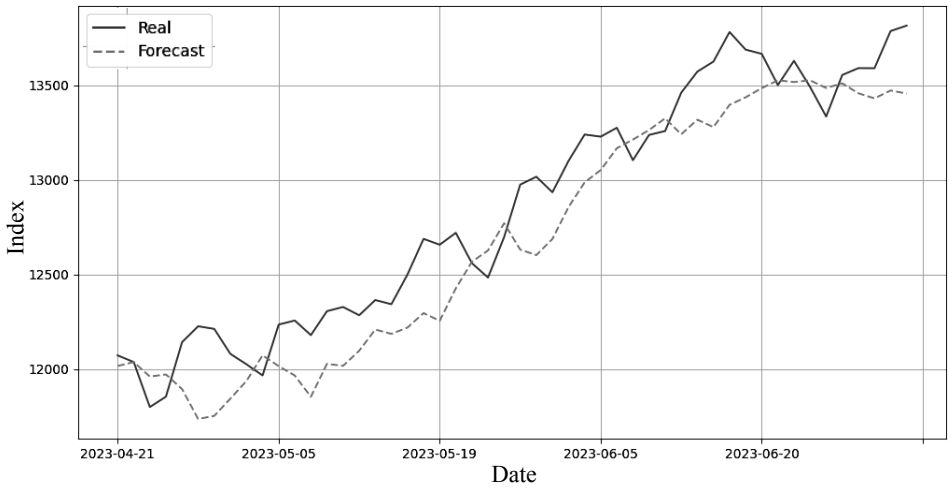


Fig. 7. Flow chart of forecast for 3 days interval by GMDH-neo-fuzzy network

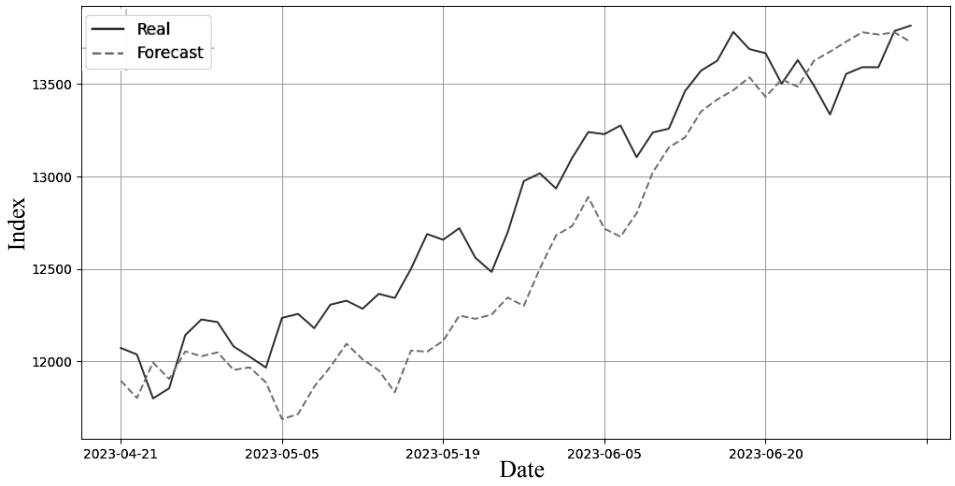


Fig. 8. Flow chart of forecast by GMDH-neo-fuzzy network for 10 days

After that the algorithm of bagging based on GMDH was implemented and experiments at short-term and middle-term forecasting were performed. The tables with criteria MSE and MAPE are presented in the Tables 1 and 2.

Table 1. Average MSE values for different intervals

Interval	ARIMA	GMDH-neo-fuzzy	HSCI-GMDH-bagging
1	17422.752	27420.394	17382.425
3	43434.022	59202.99	49332.635
5	66993.011	78330.284	58153.202
10	235427.989	108358.616	104324.0
20	696291.974	253693.345	241508.146

T a b l e 2. Average MAPE values for different intervals

Interval	ARIMA	GMDH-neo-fuzzy	HSCI-GMDH-bagging
1	0.83	1.049	0.828
3	1.416	1.63	1.398
5	1.616	1.903	1.577
10	3.134	2.208	2.108
20	5.579	3.433	3.308

Average MSE values for different intervals are presented in the Fig. 9 and MAPE values — in the Fig. 10.

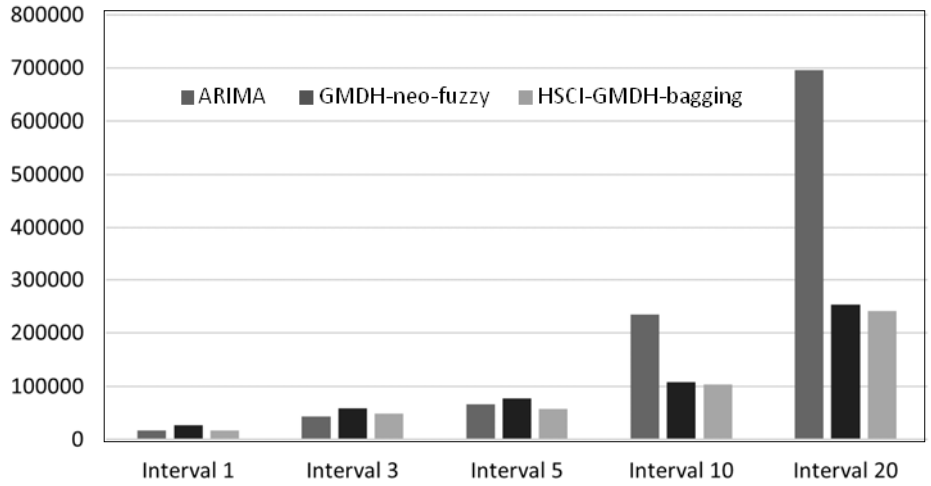


Fig. 9. Average MSE values for different forecasting intervals

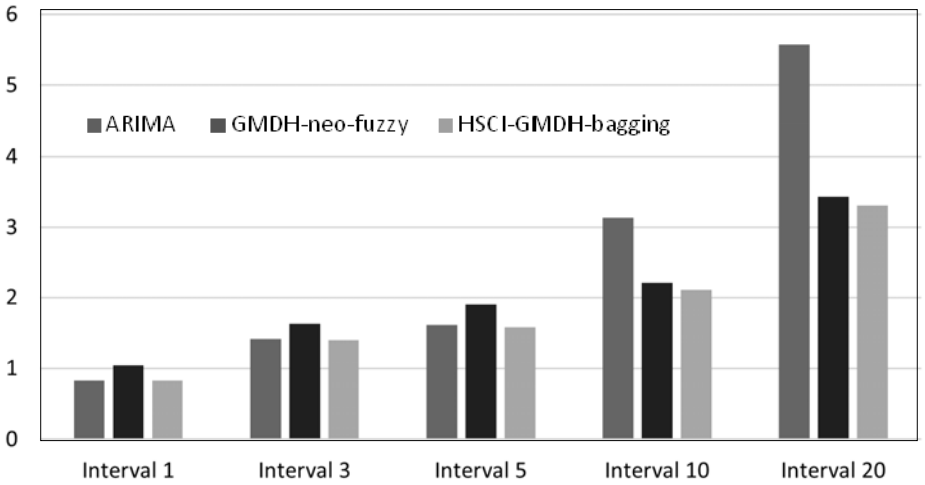


Fig. 10. Average MAPE values for different forecasting intervals

Analyzing the presented results, one may conclude that bagging procedure based on GMDH has the best results as compared with separate models in ensemble, the second place takes GMDH-neo-fuzzy network for all the intervals at the exception 1, 3 and 5 days. And for short-term forecasting 1, 3 and 5 days ARIMA appears to be better than hybrid GMDH-neo-fuzzy network.

In a whole the obtained results well comply with theoretical statements as for properties of bagging procedure and the proposed bagging based on GMDH in HSCI system appeared to be very efficient procedure which demand minimum calculations due to application of GMDH.

CONCLUSION

The architecture and learning algorithms of hybrid computational intelligence system, which is built based on Group Method of Data Handling (GMDH) method and bagging approach, are proposed.

The system consists of a sequence of stacks, while odd stacks are essentially ensembles formed by parallel connected different subsystems that solve the same problem, while even ones are metamodels that implement the bagging procedure and calculate the levels of fuzzy membership of each member of the ensemble to the optimal result. The process of increasing the number of stacks is based on GMDH principles until the desired accuracy of the final results is achieved. The proposed system does not require large volumes of training samples, provides online work and automatically determines the number of its layers — stacks in the learning process.

Experimental investigations confirm that Hybrid system of computational intelligence based on bagging and GMDH is effective for short- and middle-term forecasting in the financial sphere and has better metrics than ARIMA and GMDH neo fuzzy.

REFERENCES

1. G. Cybenko, “Approximation by superpositions of a sigmoidal function,” *Mathematics of Control, Signals and Systems*, vol. 2, pp. 303–314, 1989.
2. L.K. Hansen, P. Salomon, “Neural network ensembles,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 993–1000,1990.
3. Ye. Bodyanskiy, I. Pliss, “Adaptive generalized forecasting of multivariate stochastic signals,” *Proc. Latvian Sign. Proc. Int. Conf. Riga*, vol. 2, pp. 80–83, 1990.
4. Ye.V. Bodyanskiy, I.A. Rudneva, “On one adaptive algorithm for detecting discords in random sequences,” *Automation and Remote Control*, 56, no. 10, pp. 1439–1443, 1995.
5. A.J.C. Sharkey, “On combining artificial neural nets,” *Connect. Sci.*, vol. 8, pp. 299–313, 1996.
6. S. Hashem, “Optimal linear combination of neural networks,” *Neural Networks*, vol. 10, no. 4, pp. 599–614, 1997.
7. U. Naftaly, N. Intrator, and D. Horn, “Optimal ensemble averaging of neural networks,” *Network: Comput. Neural Syst.*, vol. 8, pp. 283–296, 1997.
8. D.T. Pham, X. Liu, “Neural Networks for Identification,” *Prediction and Control*. London, Springer Verlag, 1995, 238 p.
9. L. Breiman, *Bagging Predictors. Techn. Report No. 421*. Dept. of Statistics, Univ. of California, Berkeley, CA 94720, 1994, 19 p.
10. A.G. Ivakhnenko, V.G. Lapa, *Cybernetic forecasting devices*. K.: “Naukova Dumka”, 1965, 216 p.
11. A.G. Ivakhnenko, G.A. Ivakhnenko, and J.A. Mueller, “Self-organization of the neural networks with active neurons,” *Pattern Recognition and Image Analysis*, 4(2), pp. 177–188, 1994.
12. J. Schmidhuber, “Deep learning in neural networks: An overview,” *Neural Networks*, vol. 61, pp. 85–117, 2015.

13. Yuriy Zaychenko, Yevgeniy Bodyanskiy, Oleksii Tyshchenko, Olena Boiko, and Galib Hamidov, "Hybrid GMDH-neuro-fuzzy system and its training scheme," *Int. Journal Information Theories and Applications*, vol. 24, no. 2, pp. 156–172, 2018.
14. Yuriy Zaychenko, Yevgeniy Bodyanskiy, Olena Boiko, and Galib Hamidov, "Evolving Hybrid GMDH-NeuroFuzzy Network and Its Application," *International Conference IEEE-SAIC 2018, Kyiv, IASA, 8–11 October, 2018*.
15. Yevgeniy Bodyanskiy, Nonna Kulishova, Yuriy Zaychenko, and Galib Hamidov, "Spline-Orthogonal Extended Neo-Fuzzy Neuron," *International conference CISP-BMEI 2019*.
16. Yevgeniy Bodyanskiy, Yuriy Zaychenko, Olena Boiko, Galib Hamidov, and Anna Zelikman, "The Hybrid GMDH-Neo-fuzzy Neural Network in Forecasting Problems in Financial Sphere," *Intern. Conference IEEE SAIC 2020 in book "Advances in Intelligent Computing", Springer, 2020*, vol. 1075, pp. 221–225.
17. Ye. Bodyanskiy, O. Vynokurova, and I. Pliss, "Hybrid GMDH-neural network of computational intelligence," *Proc. 3rd Int. Workshop on Inductive Modeling, Krynica, Poland, 2009*, pp. 100–107.

Received 15.08.2023

INFORMATION ON THE ARTICLE

Yevgeniy V. Bodyanskiy, ORCID: 0000-0001-5418-2143, Kharkiv National University of Radio Electronics, Ukraine, e-mail: yevgeniy.bodyanskiy@nure.ua

Oleksii V. Kuzmenko, ORCID: 0000-0003-1581-6224, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: oleksii.kuzmenko@ukr.net

Helen Yu. Zaichenko, ORCID: 0000-0002-4630-5155, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: syncmaster@bigmir.net

Yuriy P. Zaychenko, ORCID: 0000-0001-9662-3269, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: zaychenkoyuri@ukr.net

ГІБРИДНА СИСТЕМА ОБЧИСЛЮВАЛЬНОГО ІНТЕЛЕКТУ НА ОСНОВІ БЕГГІНГУ ТА МЕТОДУ ГРУПОВОГО УРАХУВАННЯ АРГУМЕНТІВ / Є.В.Бодянський, О.В. Кузьменко, Ю.П. Зайченко, О.Ю. Зайченко

Анотація. Розглянуто проблему короткострокового та середньострокового прогнозування у фінансовій сфері. Для її вирішення запропоновано гібридну систему обчислювального інтелекту на основі методу групового урахування аргументів (МГУА) та беггінгу, а також алгоритм її навчання. Непарні стеки гібридної системи сформовані ансамблями паралельно з'єднаних підсистем. Як такі підсистеми обрано ARIMA та гібридну мережу МГУА-нео-фаззі. Запропонована система не потребує великого обсягу навчальної вибірки, автоматично визначає кількість стеків у процесі навчання та забезпечує роботу у режимі online. Проведено експериментальні дослідження з використанням запропонованої гібридної системи, а також окремо ARIMA та МГУА-нео-фаззі. Порівняно точність прогнозів, отриманих експериментальним шляхом, на основі чого обґрунтовано доцільність застосування запропонованої гібридної системи.

Ключові слова: гібридна система, беггінг, гібридна мережа МГУА-нео-фаззі, ARIMA, короткострокове та середньострокове прогнозування.

UDC 004.2
DOI: 10.20535/SRIT.2308-8893.2024.1.07

STATISTICAL METHODS OF FEATURE ENGINEERING FOR THE PROBLEM OF FOREST STATE CLASSIFICATION USING SATELLITE DATA

Y.V. SALII, A.M. LAVRENIUK, N.M. KUSSUL

Abstract. Timely detection of forest diseases is an important task for their prevention and spread limitation. The usage of satellite imagery provides capabilities for large-scale forest monitoring. Machine learning models allow to automate the analysis of these data for anomaly detection indicating diseases. However, selecting informative features is key to building an effective model. In this work, the application of Bhattacharyya distance and Spearman's rank correlation coefficient for feature selection from satellite images was investigated. A greedy algorithm was applied to form a subset of weakly correlated features. The experiment showed that selected features allow for improving the classification quality compared to using all spectral bands. The proposed approach demonstrates effectiveness for informative and weakly correlated feature selection and can be utilized in other remote sensing tasks.

Keywords: Sentinel-2, vegetation indices, Bhattacharyya distance, feature engineering, greedy algorithms, Spearman's rank correlation coefficient.

INTRODUCTION

Monitoring the condition of forest areas is a task for successfully identifying tree diseases and preventing their further spread. The use of high-resolution satellite images makes it possible to regularly obtain up-to-date information on large forest areas [1]. The process of processing and analyzing these data can be automated using machine learning methods that can detect signs of abnormal vegetation changes that may indicate the presence of diseases [2; 3].

One of the key steps in building an effective machine learning model for classification of the forest condition is the careful selection of the most informative features of the input data for the machine learning model. This allows to simplify the model and reduce the training time, without losing the quality of the classification. There are a large number of approaches to evaluating the informativeness of features. Most approaches are statistical, based on evaluating the similarity of data distributions. Among the most well-known methods for assessing the similarity of distributions, it is worth noting the Kullback–Leibler divergence, which calculates the relative entropy between two probability distributions. The higher the divergence value, the more distinct the distributions are [4]. However, this characteristic is not symmetrical, which limits its use. More universal are methods for determining the distance between data distributions, which include the Euclidean metric, that calculates the Euclidean distance between the means of two distributions, the Wasserstein distance, which measures the minimum "work" required to transform one distribution into another, or the chi-square distance, that compares the frequencies of samples from two distributions. Another symmetric

metric that allows you to measure the difference between two distributions is the Bhattacharyya distance [5].

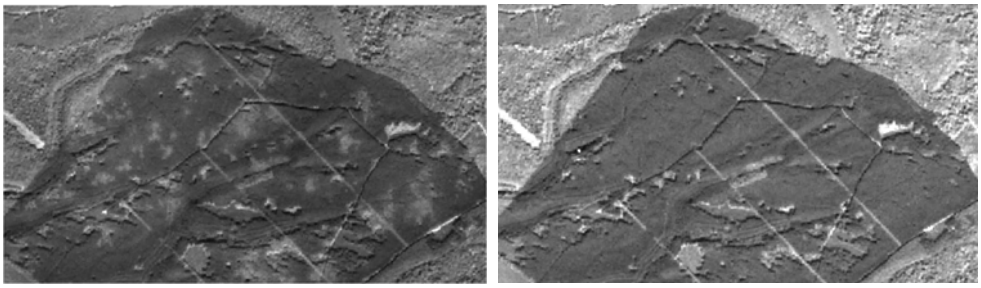
This paper investigates the possibility of using the Bhattacharyya distance and the Spearman correlation coefficient to select the most informative and at the same time weakly correlated features of multispectral satellite images.

FORMULATION OF THE PROBLEM

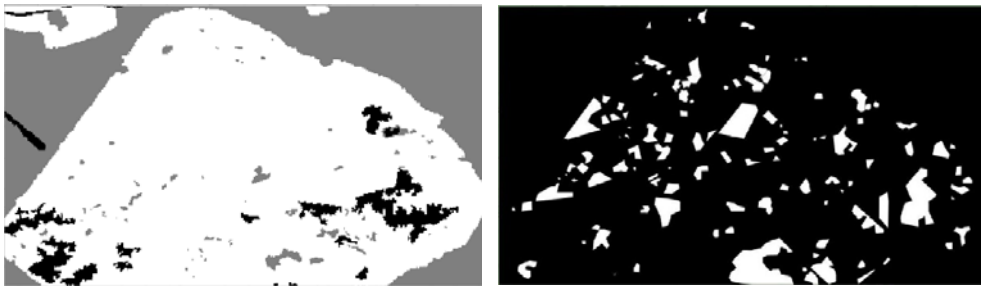
Let us consider the task of detecting disease in forest areas based on the analysis of Sentinel-2 satellite images [6]. The goal of our research is to develop an effective model that will be able to automatically determine whether a certain area of the forest is diseased on the basis of multispectral satellite data at different times.

To achieve this goal, two multispectral images will be used: current (Fig. 1, *a*) and past (Fig. 1, *b*). Since coniferous forests were studied, past images are not limited to a specific date, but it is important that the forest is healthy. Additionally, a vector mask of forest type (Fig. 1, *c*) from the Forest Type 2018 geospatial dataset [6] is used, which allows us to determine the areas where the forest is located and exclude non-forest areas from the analysis.

In the terminology of machine learning, the task is to build a binary classifier of each pixel of a multispectral image into the classes “healthy” and “stressed” (Fig. 1, *d*) by building and training a machine learning model. For training and testing of the model, the experts provided a ground truth (disease) mask (Fig. 1, *d*), which will be used for training the model and evaluation of its effectiveness. The experimental study was conducted on the data of the eastern part of the Grand Est region, France. The area for which training data is available is shown in Fig. 2.



(*a*) Current image (*b*) Past image



(*c*) Forest Type 2018 (*d*) Ground truth mask

Fig. 1. Example of input data: (*a*, *b*) RGB (B4, B3, B2) composite of Sentinel-2 images; (c) white — coniferous, gray — deciduous, black — non-forest; (*d*) white — sick, black — healthy

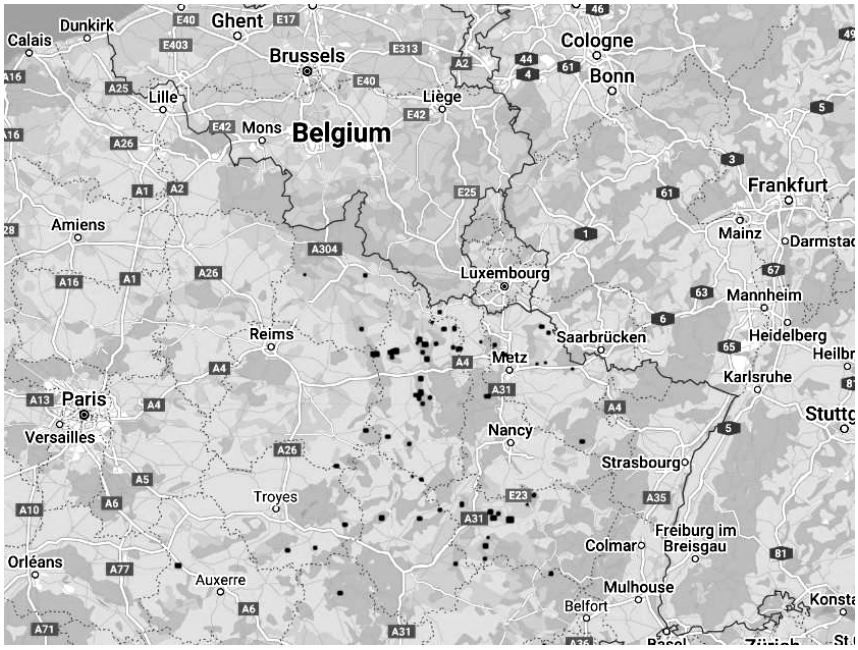


Fig. 2. Sections for which train areas are available. The locations of the areas are marked with black dots

SATELLITE DATA

The work uses multispectral images of the Sentinel-2 satellite obtained in the eastern part of France. The images contain data in 13 spectral channels (bands) with a spatial resolution of 10 to 60 meters. Images of areas of coniferous forests were selected for analysis. All channels with a resolution of 10 m (Fig. 3, *a*) and 20 m (Fig. 3, *b*) were used, as well as channel B9 (water vapor) with a resolution of 60 m (Fig. 3, *c*). This choice is due to the fact that these ranges are sensitive to the content of chlorophyll, moisture and other indicators of vegetation, the change of which may indicate the presence of diseases.

The choice of bands for research is determined by the following considerations. The B4 band (red range) is sensitive to the chlorophyll content of vegetation because chlorophyll strongly absorbs red light for photosynthesis. A decrease in the content of chlorophyll during plant stress or disease leads to an increase in reflection in the red range. B5 band (red-edge) is in the region of rapid change in reflectance from low (red light) to high (infrared). Changes in this band can carry information about the content of chlorophyll, which often changes during the development of the disease. B9 band (water vapor) is used mainly for atmospheric correction, but can help in cases of changes in the water content of vegetation under stress. Bands B11 and B12 (short-wave infrared range) are sensitive to the water content of plants because water absorbs strongly in this range. A decrease in water content under stress leads to an increase in reflectance, which may indicate the development of the disease.

In addition to the values of the spectral bands themselves, their combinations (so-called vegetation indices) will also be used as input data. Depending on the mathematical form of the index, they can highlight information about the state of

the vegetation cover; eliminate or minimize the influence of negative factors (for example, brightness).

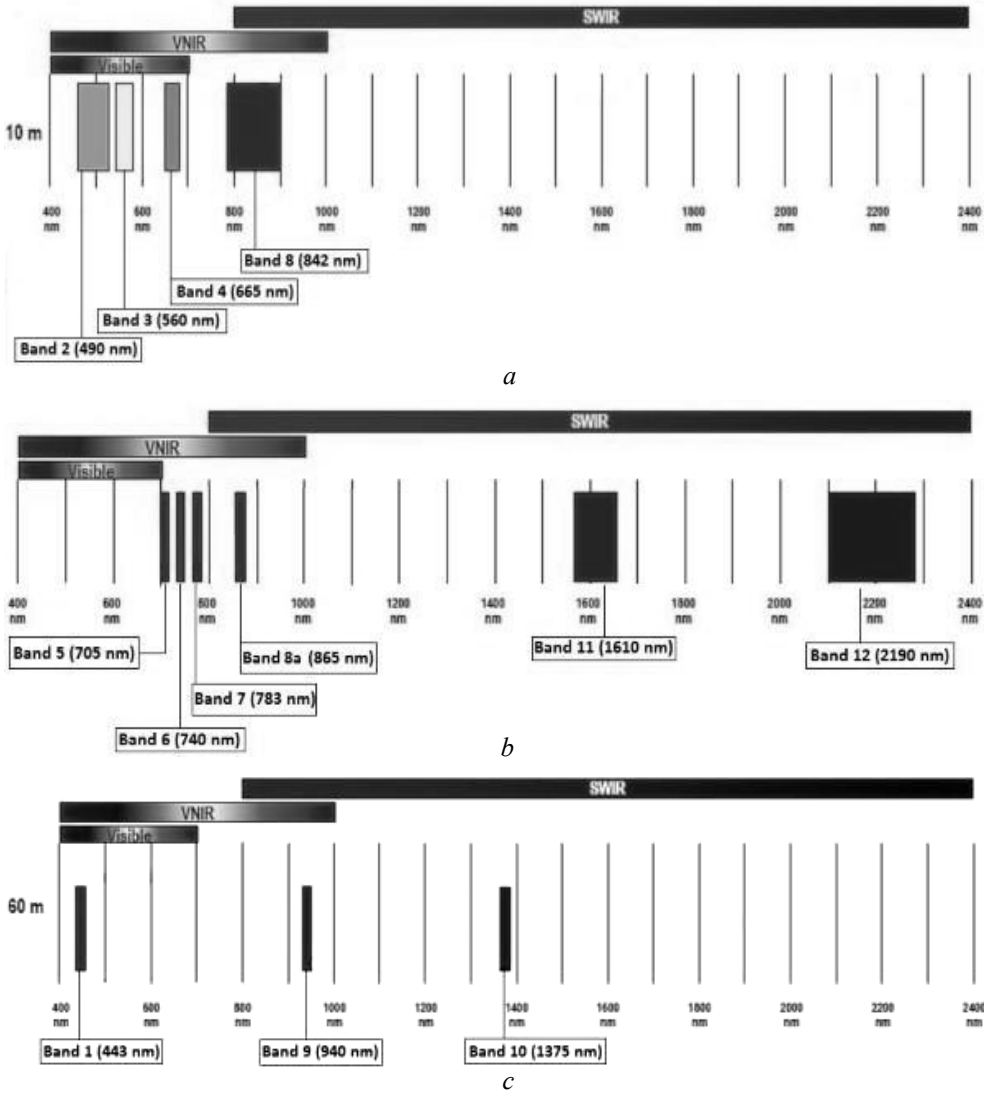


Fig. 3. Bands of Sentinel-2 satellite images [7]. Bands with a spatial resolution: 10m (*a*), 20m (*b*), 60m (*c*)

Among the well-known vegetation indices, the following can be noted:

- Green Leaf Index (GLI) — used to assess the health and development of green leaves of the plant cover, physiological state, detection of stress, drying or damage of plants, as well as monitoring of their growth and phenological changes.
- Normalized Difference Vegetation Index (NDVI) — measures the health and density of vegetation on the Earth's surface. This index is used to assess the state of ecosystems, monitor the impact of climate change and control land use.
- Disease Stress Water Index (DSWI) — used to identify plant diseases, especially coniferous forests. A decrease in DSWI values indicates a deterioration of the physiological state of the plant cover, which can be caused by diseases, for example, infectious diseases or stressful conditions.

- Chlorophyll Vegetation Index (CVI) — used to estimate the concentration of chlorophyll in the vegetation cover.

Since vegetation indices are mathematical functions, they can be generalized by classes of functions. For example, the following classes can be distinguished for the vegetation indices used in the article [2, Table 2]:

$$B(A) = A \text{ : B2, B3, B4, … , B9, B11, B12;}$$

$$NORMP(A,B) = \frac{A-B}{A+B} \text{ : NDWI, NGRDI, NDRE2, NDVI, GNDVI, NDRE3;}$$

$$FRAC(A,B) = \frac{A}{B} \text{ : RDI, PBI, CIG;}$$

$$GLIbased(A,B,C) = \frac{(A-B)+(A-C)}{(A+B)+(A+C)} \text{ : GLI;}$$

$$NORPP(A,B,C,D) = \frac{A+B}{C+D} \text{ : DSWI;}$$

$$CVIbased(A,B,C,D) = \frac{A \cdot B}{C \cdot D} \text{ : CVI;}$$

$$DIST(A,B) = \sqrt{A^2 + B^2} \text{ : DRS.}$$

Set of values of various spectral bands and possible vegetation indices will be used as input data for building and training models for classifying the state of the forest into healthy and stressed.

METHODOLOGY

Bhattacharyya distance calculation

to evaluate the informativeness of the features, the bhattacharyya distance will be used between the distributions of the feature values for the healthy and damaged forest classes. The larger the value of this distance, the better the feature separates these classes.

Bhattacharyya distance is calculated by the formula:

$$D_B(H,S) = -\ln(BC(H,S)),$$

where $BC(H,S) = \sum_{i=1}^{n} \sqrt{H_i S_i}$ — Bhattacharyya coefficient, where H_i, S_i — probabilities of the *i*-th value (the height of the *i*-th columns of the histograms H and S).

The value of the Bhattacharyya coefficient is sensitive to the number of histogram columns. If their number is too low, the coefficient will be underestimated, if it is too large, it will be overestimated. Because of this, it is reasonable to take the number of histogram columns equal to the square root of the number of observations of the stressed class.

Greedy feature selection algorithm

In the previous section, classes of functions were presented, on the basis of which a large set of vegetation indices can be created. In this section, the most relevant features from this variety of indices will be defined. Our task is to select a subset

of features that jointly satisfy two key criteria. First, they should be distinguished by high values of the Bhattacharyya distance that characterizes their significance. Secondly, these features should be as independent as possible, i.e. carry as much new information as possible.

This approach is aimed at building a compact and at the same time informative subset of features that helps to understand key relationships and features in the data. This selection of features will simplify the machine learning model and increase its effectiveness.

A greedy algorithm will be used to form an optimal set of informative and weakly correlated features. At each step of the work, greedy algorithms choose a locally optimal solution, which makes it possible to obtain a satisfactory global approximation in an acceptable time [8]. Heuristics based on the Bhattacharyya distance and the Spearman correlation coefficient will be used as a criterion for local optimality of the greedy algorithm.

Spearman correlation coefficient

Spearman correlation coefficient [9] is used to determine the statistical dependence of two variables and shows to what extent the dependence between variables can be described using a monotonic function. The Spearman correlation coefficient is defined as the Pearson correlation coefficient for variable ranks, that is, it does not operate with the values of quantities, but with their serial numbers. Its values lie within [–1, 1].

This value is defined as $r_s(X,Y) = \dfrac{cov(R(x),R(Y))}{\sigma(R(x))\sigma(R(y))}$, where X, Y — values of variables; $R(X)$, $R(Y)$ — transformation of variable values into their ranks; $cov(R(X),R(Y))$ — covariance of rank values; $\sigma(R(x))$, $\sigma(R(y))$ — dispersion of the corresponding rank values.

Independence coefficient

Since the goal is to find the most independent features, a value of 1 should correspond to independent features, and 0 should correspond to fully dependent features. Thus, the coefficient of independence will take the form: $c_i(X,Y) = 1 - |r_s(X,Y)|$, where $r_s(X,Y)$ is the Spearman correlation coefficient.

Proposed algorithm

The algorithm works as follows:

1. The weight of each feature is set equal to the value of the Bhattacharyya distance of the given feature.
2. The feature with the largest current weight is selected.
3. The weight of each feature is multiplied by the independence coefficient between the current feature and the feature selected in the previous step.
4. Return to step 2 until the required number of features have been selected

The pseudocode of the algorithm can be seen in Fig. 4. It is important to note that it uses a modified coefficient of independence, namely, a constant value C is added to it. Such change allows you to adjust what the algorithm should pay more

attention to: the independence of features ($C \to 0$) or the informativeness of features ($C > 0$).

```
procedure SELECTINDEPENDENTSUBSET(indexes, k)
    indexes ← list of all indices and their Bhattacharyya distance
    k ← required number of features
    if len(indexes) ⩽ k then return indexes
    end if
    output ← new list[k]
    i ← 0
    while i < k do
        selected ← indexes[argmax(indexes, indexes.distance)]
        output[i] ← selected
        indexes.distance ← indexes.distance ∗ (1 + C − |spearman(indexes, selected)|)
        i ← i + 1
    end while
    return output
end procedure
```

Fig. 4. Pseudocode of the proposed algorithm

This approach allows to consistently add to the set the most informative at the moment and weakly correlated with previous signs. As a result, a set is formed that contains a maximum of information for a given number of features.

Machine learning models for classification

To compare the effectiveness of different sets of features, machine learning models will be used for binary classification of the forest state.

Multilayer perceptron [10] with a different number of hidden layers will be used as a basic classifier. Optimal architecture and hyperparameters of the model will be tuned using a genetic algorithm.

The models will be trained on a training sample of satellite images with ground truth masks. To assess the quality of the classification, such metrics as [11] will be used: overall accuracy (accuracy), Jacquard coefficient (IoU), cross-entropy (log-loss), area under the ROC curve (ROC AUC) on the validation sample.

Note that the overall accuracy metric makes sense only when assessing the classification accuracy with a balanced distribution of classes, which occurred on the validation sample.

Five-fold cross-validation was used to analyze metrics.

Comparing the results of models built on different sets of features allows us to assess the contribution of the proposed feature selection method to improving the quality of classification.

EXPERIMENT RESULTS

Description and results of the experiment on evaluating the informativeness of features

In order to evaluate the informativeness of various features of the image, the Bhattacharyya distance between the classes of “stressed” and “healthy” coniferous forest was calculated for various spectral channels and vegetation indices.

Sentinel-2 images containing fragments of healthy and stressed coniferous forest in eastern France were used as input data. Each image contains 12 bands.

For each pixel of each image, 43.644 vegetation indices were calculated based on 12 spectral channels according to the given index classes.

The territories were divided into 3 parts, for each of which Bhattacharyya distances were calculated separately between the obtained histograms of class distributions. Histograms were built within the values of the "stressed" class with the number of columns equal to the square root of the number of pixels of this class.

As a result, average estimates of the Bhagattacharya distance ($Avg(D_B(H,S))$) for each of the features were obtained.

Among the original spectral channels, bands B4, B11 and B12 showed the greatest informativeness (Table 1). Among the known vegetation indices, 7 (RDI, NDWI, NGRDI, DSWI, NDRE2, NDVI, GLI) demonstrated higher separation rates (Table 1) compared to spectral channels. A comparison of Bhattacharyya distance values in vegetation indices from table 1 and the corresponding class of indices (Table 2) shows that there are instances of the class with a larger distance value.

T a b l e 1 . Bhattacharyya distance

Sentinel-2 bands		**Well-known vegetation indices**		
Band	$Avg(D_B(H,S))$	Index	Formula	$Avg(D_B(H,S))$
$B12$	0.332	RDI	$\frac{B12}{B8A}$	0.581
$B4$	0.306	NDWI	$\frac{B8A-B11}{B8A+B11}$	0.577
$B11$	0.228	NGRDI	$\frac{B3-B4}{B3+B4}$	0.562
$B1$	0.116	DSWI	$\frac{B8+B3}{B4+B11}$	0.513
$B9$	0.115	NDRE2	$\frac{B7-B5}{B7+B5}$	0.470
$B5$	0.107	NDVI	$\frac{B8A-B4}{B8A+B4}$	0.448
$B7$	0.092	GLI	$\frac{(B3-B4)+(B3-B2)}{(B3+B4)+(B3+B2)}$	0.389
$B2$	0.073	PDI	$\frac{B8}{B3}$	0.219
$B8$A	0.068	CIG	$\frac{B8A}{B3}-1$	0.183
$B6$	0.068	GNDVI	$\frac{B8A-B3}{B8A+B3}$	0.182
$B8$	0.062	NDRE3	$\frac{B8A-B7}{B8A+B7}$	0.126
$B3$	0.041	CVI	$\frac{B8A\cdot B5}{B3^2}$	0.051

Table 2. The largest Bhattacharyya distance for classes of vegetation indices

Class	Instance	$Avg(D_B(H,S))$
$CVIbased(A,B,C,D)$	$\frac{B3 \cdot B6}{B11 \cdot B4}$	0.686
$NORPP(A,B,C,D)$	$\frac{B2 + B6}{B12 + B4}$	0.646
$GLIbased(A,B,C)$	$\frac{(B6 - B4) + (B6 - B11)}{(B6 + B4) + (B6 + B11)}$	0.630
$NORMP(A,B)$	$\frac{B6 - B12}{B6 + B12}$	0.609
$FRAC(A,B)$	$\frac{B6}{B12}$	0.603
$DIST(A,B)$	$\sqrt{B12^2 + B4^2}$	0.354
$B(A)$	$B12$	0.331

So, the experiment confirmed that the Bhattacharyya distance can be used to evaluate the informativeness of features, confirmed the advantage of using vegetation indices in comparison with the original image data, and showed that for each class of indices it is possible to find instances with better informativeness (within the scope of the task) than in well known indices.

This makes it possible to form an effective set of features for further training of machine learning models without the need for their previous training.

Determination of the optimal set of informative features

The analysis of the results of feature selection showed that among the initial set of 43.644 calculated vegetation indices, 3.128 had a Bhattacharyya distance above 0.4, which indicates their high informativeness.

Using the proposed greedy algorithm, sets of 12 and 24 features were obtained. The algorithm used a modified coefficient of independence: $c_i(X,Y) + C$, where $C = 0.6$. At this value of the C parameter, the model showed the best result.

As it was found during experiments, the use of classes $GLIbased(A,B,C)$, $NORPP(A,B,C,D)$, $CVIbased(A,B,C,D)$ during selection leads to a decrease in metric values, so their use was abandoned. Rejecting them allows you to reduce the execution time of the algorithm by an order of magnitude.

As a result of this problem, it seems reasonable to search for an effective combination within each class separately, and then somehow combine them. However, the identification of the causes of this problem and methods of solvingit require a separate study.

As can be seen in Fig. 5, the obtained set of features mostly contains features with relatively little informativeness. This indicates that although a large number of signs are informative, they are also highly correlated. This is also confirmed by the fact that the selected features with high informativeness are more correlated with each other (have a darker color) than the features with low.

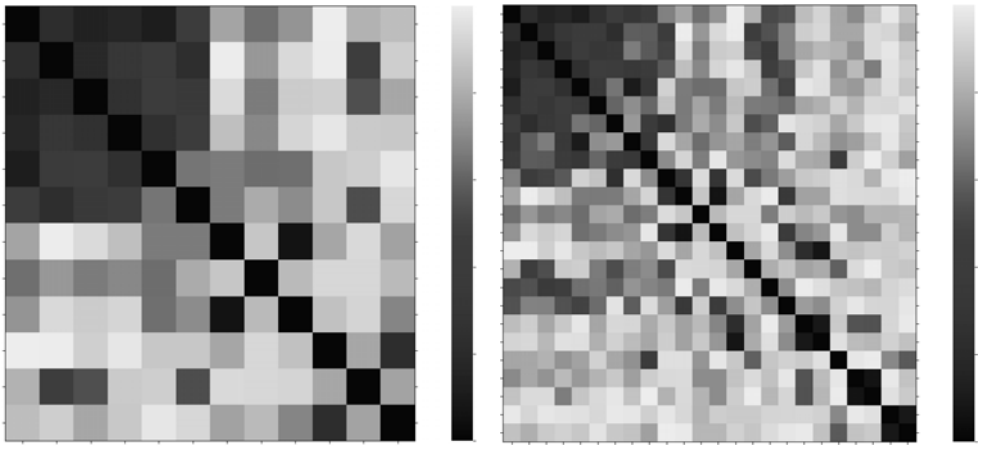


Fig. 5. Independence matrix for (*a*) — 12, (*b*) — 24 features selected by the proposed algorithm among the classes *NORMP*(*A*, *B*), *FRAC*(*A*, *B*), DIST(*A*, *B*)

Fig. 5 shows the example of pairwise independence of selected features, where features are placed from top-left to bottom-right in descending order of their individual informativeness. Brightness of each cell corresponds to their independence, where brighter means higher.

Overall, the obtained set of features mostly contains features with relatively little informativeness. As can be seen in fig. 5, the selected features with high informativeness are more correlated with each other (have a darker color) than the features with low.

Analysis of machine learning results

To evaluate how successful the sets turned out to be, let's compare the accuracy of models built using them compared to models using only spectral channels, known vegetation indices, and their combination.

Fig. 6 immediately shows that the model built on the features proposed by the algorithm shows much better knowledge of metrics and their dynamics. Thus, already after about 7 epochs, the model based on 12 proposed features shows metrics close to the metrics of the model based on known vegetation indices. This comparison confirms that the proposed algorithm is effective.

It should also be noted that the models from the 12 proposed features show close (but still slightly worse) values of metrics to the model based on spectral classes. At the same time, the algorithm worked an order of magnitude longer than model training. Based on this, it can be concluded that the spectral channels are a good enough basis, and the multilayer perceptron model is able to build a vegetative index with high informativeness on their basis.

However, when the number of features is doubled, the proposed algorithm was able to find such a set of features that improves the accuracy of the model and the rate of its learning. This suggests that the feature sets previously failed to maximize the amount of information, which one would think, as the model based on a combination of known vegetation indices and spectral channels showed almost the same accuracy and learning rate as the model based on spectral channels alone. And also confirms the usefulness and necessity of feature engineering.

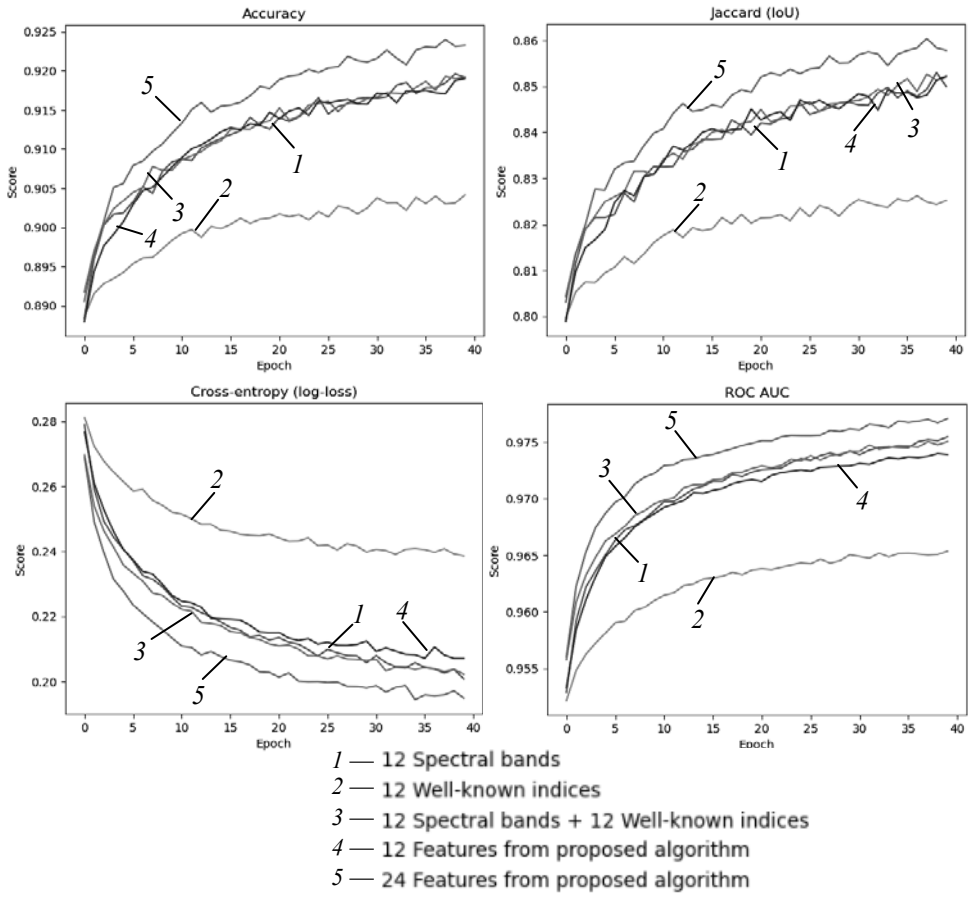


Fig. 6. Metrics of built models on the validation sample

In the future, the proposed algorithm can be applied to optimize forest state classification models based on other types of data and for other tasks of remote sensing of the Earth.

CONCLUSIONS

In this work, the possibility of using Bhattacharyya distance to assess the relative importance of features in the task of forest state classification based on satellite images was investigated.

The analysis of real data showed that it makes sense to consider not specific (known) vegetation indices, but their classes. It was confirmed that within each class, with a high probability, a vegetative index can be found that is more informative compared to the known and original Sentinel-2 spectral channels.

The proposed greedy feature selection algorithm based on the Bhattacharyya distance and the Spearman correlation coefficient made it possible to form a set of 12 features with similar accuracy indicators, and a set of 24 features with significantly better ones, compared to the model based only on the spectral channels of the image.

Therefore, the proposed approach is effective for selecting informative and weakly correlated features based on satellite images. It can be applied to find an

effective set of features for building machine learning models in forest condition monitoring tasks and other fields of Earth remote sensing data analysis without the need to pre-train the models.

REFERENCES

1. N. Kussul, G. Lemoine, J. Gallego, S. Skakun, and M. Lavreniuk, "Parcel based classification for agricultural mapping and monitoring using multi-temporal satellite image sequences," *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), IEEE, 2015*. doi: 10.1109/igarss.2015.7325725.
2. J. Zhang, S. Cong, G. Zhang, Y. Ma, Y. Zhang, and J. Huang, "Detecting Pest-Infested Forest Damage through Multispectral Satellite Imagery and Improved UNet++," *Sensors*, vol. 22, issue 19, 2022. doi: 10.3390/s22197440.
3. N.N. Kussul, N.S. Lavreniuk, A.Y. Shelestov, B.Y. Yailymov, and I.N. Butko, "Land Cover Changes Analysis Based on Deep Machine Learning Technique," *Journal of Automation and Information Sciences*, vol. 48, no. 5, pp. 42–54, 2016. doi: 10.1615/jautomatinfscien.v48.i5.40.
4. T. van Erven, P. Harrëmos, "Rényi divergence and kullback-leibler divergence," *IEEE Transactions on Information Theory*, 60(7), 2014. Available: https://doi.org/10.1109/TIT.2014.2320500
5. A. Ilnitskiy, O. Burba, "Statistical criteria for assessing the informativity of the sources of radio emission of telecommunication networks and systems in their recognition," *Cybersecurity: Education, Science, Technique*, 1(5), pp. 83–94, 2019. doi: 10.28925/2663-4023.2019.5.8394.
6. *Forest type 2018*. Accessed on: April 07, 2023. [Online]. Available: https://land.copernicus. eu/pan-european/highresolution-layers/forests/forest-type-1/status-maps/forest-type-2018.
7. "Spatial Resolutions," *Sentinel Online*. Accessed on: August 13, 2023. [Online]. Available: https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/resolutions/spatial
8. "What is a Greedy Approach? - Algorithms for Coding Interviews in Java," *educative.io*. Accessed on: May 08, 2023. [Online]. Available: https://www.educative.io/courses/algorithms-coding-interviews-java/3j1R50KnNjQ
9. C. Croux, C. Dehon, "Influence functions of the Spearman and Kendall correlation measures," *Statistical Methods and Applications*, vol. 19, pp. 497–515, 2010. doi: 10.1007/s10260-010-0142-z.
10. P.M. Atkinson, A.R. Tatnall, "Introduction neural networks in remote sensing," *International Journal of Remote Sensing*, vol. 18(4), 1997. doi: 10.1080/014311697218700.
11. "Metrics for semantic segmentation," *ilmonteux.github.io*. Accessed on: May 27, 2023. [Online]. Available: https://ilmonteux.github.io/2019/05/10/segmentation-metrics.html

Received 06.09.2023

INFORMATION ON THE ARTICLE

Yevhenii V. Salii, ORCID: 0009-0006-0395-8099, Educational and Research Institute of Physics and Technology of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: yevhenii.salii@gmail.com

Alla M. Lavreniuk, ORCID: 0000-0002-5791-0377, Educational and Research Institute of Physics and Technology of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: alla.lavrenyuk@gmail.com

Nataliia M. Kussul, ORCID: 0000-0002-9704-9702, Educational and Research Institute of Physics and Technology of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: nataliia.kussul@gmail.com

СТАТИСТИЧНІ МЕТОДИ ІНЖЕНЕРІЇ ОЗНАК ДЛЯ ЗАДАЧІ КЛАСИФІКАЦІЇ СТАНУ ЛІСІВ ЗА СУПУТНИКОВИМИ ДАНИМИ / Є.В. Салій, А.М. Лавренюк, Н.М. Куссуль

Анотація. Своєчасне виявлення хвороб лісу є важливим завданням для запобігання їх поширенню та обмеження наслідків. Використання супутникових зображень надає можливості для великомасштабного моніторингу лісів. Моделі машинного навчання дають змогу автоматизувати аналіз цих даних для виявлення аномалій, що можуть свідчити про наявність хвороб. Відбір інформативних ознак є ключовим етапом побудови ефективної моделі. Досліджено можливість застосування відстані Бгаттачар'я та коефіцієнта кореляції Спірмена для відбору ознак із супутникових зображень. Застосовано жадібний алгоритм для формування підмножини слабко корельованих ознак. Експеримент показав, що обрані ознаки дозволяють покращити якість класифікації порівняно з використанням усіх спектральних каналів. Запропонований підхід продемонстрував ефективність для відбору інформативних і слабко корельованих ознак та може застосовуватися в інших задачах дистанційного зондування Землі.

Ключові слова: Sentinel-2, вегетаційні індекси, відстань Бгаттачар'я, інженерія ознак, жадібні алгоритми, коефіцієнт кореляції Спірмена.

UDC 004.942:537.525
DOI: 10.20535/SRIT.2308-8893.2024.1.08

EVALUATION OF THE THERMAL REGIME OF THE CATHODE OPERATION OF A HIGH-VOLTAGE GLOW DISCHARGE ELECTRON GUN, WHICH FORMS A RIBBON ELECTRON BEAM

I.V. MELNYK, S.B. TUHAI, D.V. KOVALCHUK, M.S. SURZHIKOV, I.S. SHVED, M.Yu. SKRYPKA, O.M. KOVALENKO

Abstract. The article discusses the various methods of estimating the surface cathode temperature of the high-voltage glow discharge electron gun, which forms a ribbon electron beam with a linear focus. Numerical estimations have been made to design the cathode assembly of an industrial gun. It is shown that the most effective way to make approximate estimates of the temperature of the cathode surface in high-voltage glow discharge electron guns for various technological purposes is to use arithmetic-logical ratios for modeling the geometry of the cathode assembly and locus functions for estimating the temperature distribution. The accuracy of such estimates, made using the heat balance equation, was 5–10%, sufficient at the initial stage of designing an electron gun. It is shown that using the SolidWorks CAD software complex for designing high-voltage glow discharge electron guns is effective only for solving the complex engineering design tasks and preparing the corresponding technical documentation. The results of the theoretical research published in the article are of interest to a wide range of specialists engaged in developing electron beam equipment and its implementation in industrial production.

Keywords: electron gun, high-voltage glow discharge, CAD-systems, thermodynamic equation, heat balance equation, arithmetic-logical relation, locus.

INTRODUCTION

Today, high-voltage glow discharge (HVGD) electron guns are widely used in various branches of industry [1–3]. The main fields of application of such electron guns are electron-beam welding of thin-walled parts and their assemblies, deposition of ceramic coatings in an environment of active or inert gases, depending on the requirements of the technological process, as well as electron-beam cleaning of refractory metals and ceramic materials for the further use of high-refining materials in modern production technologies. The main advantages of HVGD electron guns over traditional electron sources with incandescent cathodes are the following [1–3].

1. Stability and reliability of operation in conditions of low vacuum in the environment of various gases, in particular inert and active ones.

2. Simplicity of the design of HVGD electron guns with the possibility of replacing spent units, in particular the cold cathode of the HVGD.

3. Simplicity of vacuum technological equipment, without the need to ensure the operation of the electron gun in conditions of high vacuum.

4. Simple control of the power of the electron beam. In general, there are two ways to provide such control in order to improve the quality of industrial products that are subject to heat treatment with an electron beam. There are the following: changing the pressure in the discharge gap [4] or changing the concentration of ions in the anode plasma due to the lighting of the additional auxiliary discharge [5].

In connection with the physical features of the operation of HVGD electron guns listed above and the existing advantages of these types of guns over traditional electron sources with incandescent cathodes as well as over other beam technologies, including laser ones [2; 3], the main industrial areas of application of HVGD electron guns in modern electron beam technologies are the following.

1. High-speed electron beam welding of capsules and contacts of modern electronic devices, in particular cryogenic ones [6; 7].

2. Production of ceramic films for high-quality and durable capacitors with the aim of creating a new component base of the modern electronic industry [8–11].

3. Production of ceramic films for receiving and transmitting antennas of modern microwave communication electronic means [8–11].

4. Deposition of heat-resistant and heat-protective coatings in the modern automotive, aviation, and space industries. For example, advanced modern technologies are the application of heat-protective coatings on the blades and other parts of automobile and aircraft engines [12–15].

5. Electron beam refining of refractory metals and ceramic materials in order to increase their purity and quality [16–20]. For example, in works [17–20], the features of applying electron-beam remelting technologies to obtain pure polycrystalline silicon were considered. The aim of this advanced electron-beam technology is to improve the manufacturing technology of modern semiconductor electronics devices, in particular, electronic microcircuits.

6. The modern technologies of three-dimensional printing on the metal's substrates [21].

STATEMENT OF THE PROBLEM

One of the features of the HVGD electron gun is a rather low current density from the metallic surface of the cold cathode, the maximum value of which is about $10^3 \frac{A}{m^2}$. In connection with this, an advanced direction for the introduction of HVGD electron guns in modern electron-beam technologies is the formation of profiled electron beams with a linear and circular focus from a large cathode surface. The main advantage of using such types of HVGD electron guns in modern electron-beam welding and surface treatment technologies is the extremely high productivity of the technological process, since the use of profiled electron beams allows for a short period of time to process the products of a long duration with complex geometry without the need to provide the additional scanning of the electron beam on the surface of detail, which is treated [2; 3].

The main advantages of HVGD electron guns, which form profile electron beams, are the following [2; 3].

1. A wide range of power, from 10 kW to 600 kW, which allows designing optimal equipment for a specific technological application.

2. Stability of operation in a wide vacuum range, the partial gas pressure can be from 10–4 to 10–1 Pa, including in conditions of dynamic pressure changes in the process chamber.

3. The use of different active and noble gases, as well as mixtures of them, is also possible.

4. The HVGD electron guns are compact and light in weight.

5. The HVGD electron guns are simple to exploit and repair, and they are also reliable in operation.

6. The cold cathodes of HVGD electron guns have a significant time of operation without necessary cleaning, up to 100 amps for hours or more.

The construction diagram of the electrode system of the HVGD electron gun, which forms an electron beam with a linear focus, is shown in Fig. 1. The main features of the electrode systems of HVGD electron guns, which form electron beams with a ring and linear focus, as well as methods of evaluating their geometric design parameters, were considered in papers [22–24].

In connection with this, the most important tasks related to the mathematical model of HVGD electron sources, from a practical point of view, are estimates of the thermal modes of operation of the cold cathode, which form electron beams with a linear focus. In general, from the point of view of HVGD physics, the thermal mode of operation of the cold cathode significantly affects its emission properties [1].

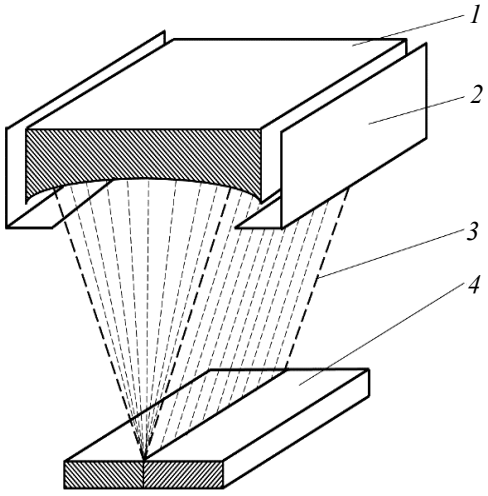


Fig. 1. Structural diagram of the HVGD electrode system, which forms an electron beam with a linear focus: *1* — cathode; *2* — anode; *3* — electron beam; *4* — product being processed

Methods of analysis of the self-aligned electron-ion optics of HVGD, taking into account the assessment of the position of the plasma boundary relative to the cathode, have been studied quite thoroughly and were described in the papers [1, 25–27]. Generally, anode plasma in HVGD electrode systems is considered a moving electrode with a given potential that is transparent to electrons. The peculiarities of estimating the position of the plasma boundary in HVGD electrode systems, which form electron beams with a linear and circular focus, were considered in a paper [20]. The general theoretical approach to estimating the operating temperature of the cold cathodes of the HVGD electron guns was considered in [28].

Another approach to simulation and experimental study of the properties of powerful and intensive electron beams in vacuum and plasma devices, including microwave ones, is presented in the papers [29–31]. In the paper [29], the radiation of plasma diodes in the microwave range has been studied experimentally. In the paper [30] the penetration of an intensive electron beam into an irradiated ob-

ject in dependence on the incidence angle of the electron beam has been studied. In the paper [31] model of an electron beam for an advanced method of computational dosimetry for radiation processing has been elaborated and tested.

Therefore, the main goal of this work is to evaluate the thermal modes of operation of the cold cathode in HVGD electrode systems that form electron beams with a linear focus.

GENERAL THEORETICAL APPROACHES AND THE BASIC EQUATIONS FOR CALCULATING THE TEMPERATURE DISTRIBUTION ON THE COLD CATHODE SURFACE OF HIGH-VOLTAGE GLOW DISCHARGE ELECTRON SOURCES

The main conclusion, which is a consequence of the general theory of HVGD, is that the accelerated electrons in HVGD do not significantly affect the temperature of the working gas [1]. Accordingly, limitations on the power of HVGD electron guns can only be associated with ensuring the dissipation of energy on the electrodes. In addition, the maximum power of HVGD electron guns is limited by the emissive properties of the cold cathode, as well as by contraction of the discharge and step ionization under conditions of high pressure [1]. The maximum value of the current density of the secondary ion-electron emission is $10^3 A/m^2$. In any case, in the general theory of the HVGD, it is justified that, according to the relations for the current-voltage characteristic of the HVGD, the HVGD electron guns, within the physical conditions of the existence of the discharge under a defined range of the pressure and the accelerating voltage, are operated stably, without unwanted surges of current up to extremely high values [1].

The second conclusion of the general theory of HVGD is that to reduce the temperature of the electrodes and, accordingly, the effect of their heating on the discharge parameters, it is desirable to use light gases, for example, hydrogen or helium, in the powerful electron guns. In addition, it is known that when light gases are used, the sputtering coefficient of the cathode surface S decreases, and, accordingly, the operation time of the HVGD electron gun without replacing the cathode increases [1]. However, the choice of operation gas or gas mixture is always combined with the task of ensuring a high value of the coefficient of secondary ion-electron emission γ_g, which allows, under the condition of stable operation of the cathode, to increase the power of the electron beam that is formed [1–3]. Another way to reduce the operating temperature of the electrodes' surfaces is to increase the gas flow rate. However, the application of this method must be compatible with the gun current stabilization system, since the gun current depends on the pressure of the operation gas, and, according to the main equation of vacuum technology, the gas pressure in the HVGD gun directly depends on the gas flow rate [4]. Usually, in modern electron-beam technological equipment, the control of the current of the HVGD gun is carried out by changing the gas pressure by automatically adjusting the gas flow into the electron gun [4]. Therefore, this method of cooling is generally not used in industrial HVGD electron guns.

That is why ensuring the effectiveness of water cooling of the cold cathode due to the thermal power released on it as a result of bombardment with residual gas ions is the main technical factor that ensures the stable operation of HVGD electron guns [28].

Generally, the following four approaches are used to solve applied problems of thermodynamics [32–35].

1. A simplified approach based on the transition to a one-dimensional model without taking edge effects into account and solving the heat balance equation [32; 33]:

$$P_t + c_m(T) m_m \frac{dT_s}{dt} = P_b; \quad P_t = \int_0^{l_m} \frac{S_s(T_s - T_{cl})}{\lambda_m(T)} dl, \qquad (1)$$

where c_m is the specific heat capacity of the heated material; λ_m is the thermal conductivity of the heated material; S_s is the surface area irradiated by the flow of energetic particles; l_m is the thickness of the body being heated; T_s is the temperature on the surface of the substance, m_m is the mass of the heated material; T_{cl} is the temperature coolant; P_b is the power absorbed by the substance; and P_t is the lost power due to thermal conductivity.

Estimates of the surface temperature, which are based on relations (1), are most accurate for metals that have high thermal conductivity, and, due to this, the temperature of the surface of the metal, which is heated by a high-energy stream of charged particles, can be assumed to be the same with great accuracy [28]. Therefore, for the preliminary analysis of the temperature of the emission surface of the cathode of the HVGD electron guns, the ratio (1) is used. For electrode systems with direct cooling of the cathode by a coolant, the heat balance equation (1) is written as follows [28]:

$$T_c = \frac{W_c \dfrac{\gamma_g}{1+\gamma_g} \cdot \dfrac{R_c + l_c - \sqrt{R_c^2 - R^2}}{\lambda_c} + \dfrac{l_к}{\lambda_к}}{\alpha_c R_c^2 \left(\arcsin \dfrac{R}{2R_c} - \dfrac{R}{2R_c} \right)} + T_c, \quad \alpha_c = k_1 + k_2 \sqrt{v_c}, \qquad (2)$$

where α_c is the heat transfer coefficient of the liquid through the base of the cathode; λ_c is the thermal conductivity of the cathode material; R_c — the radius of the cathode emission surface; R — the transversal size of the cathode; v_c is the rate of the cooling liquid, m^3/s, k_1 and k_2 are the empiric coefficients. For example, for water, at room temperature and atmospheric pressure, $k_1 \approx 350$ and $k_2 \approx 21000$ [34; 35].

In the case of a complex geometry of the cathode, a mathematical approach for describing the geometric form of complex three-dimensional objects, based on the application of arithmetic-logical relations [36] and Rvachov functions [37–39], or the locus, can be effectively used to find the temperature of its surface using relations (2). The appropriate software means can be effectively implemented using the programming tools of the MatLab system of scientific and technical calculations [40].

2. Analytical solution of the heat conduction equation [32; 33]:

$$\frac{\partial}{\partial t}(\rho c_v T) = \frac{\partial}{\partial x}\left(\lambda \frac{\partial T}{\partial x}\right) + \frac{\partial}{\partial y}\left(\lambda \frac{\partial T}{\partial y}\right) + \frac{\partial}{\partial z}\left(\lambda \frac{\partial T}{\partial z}\right) + F(x, y, z), \qquad (3)$$

where x, y, and z are the space coordinates; T is the temperature of medium; λ is the thermal conductivity of the substance; ρ is its density by mass; $F(x, y, z)$ — the density of the thermal sources on the surface boundary.

Solving the generalized equation of thermal conductivity with partial derivatives (3) is carried out analytically by expanding its solution under given boundary conditions into a functional series. To carry out this operation, consider the kernel of equation (3), which, in its general form, is written as follows [28; 32; 33]:

$$\Phi(x,t) = \frac{1}{(2a_t\sqrt{\pi t})^n}\exp\left(-\frac{x^2}{4a_t^2 t}\right), \quad a_t^2 = \frac{\lambda}{\rho c_v}, \tag{4}$$

where x is the coordinate, that is under consideration and by which the equation (3) is solved; t is the time; a_t is the coefficient of thermal diffusion; c_v is the gas heat capacity by the volume, and n is the number of the basic function. In the theory of thermal conductivity, it has been proven that the series based on the use of functions (3) are orthogonal and coincide [41–44]. Generally, the coefficients of thermodynamic tasks in the functional row (4) must corresponded to partial differential equation (3), given above [28; 32; 33]. The general disadvantage of this approach is the difficulty of automating the formation of functional series using computer programs [32; 33]. That is, the application of such an approach requires a large amount of routine work from professional mathematicians and engineers. On the other hand, the obtained functional series can be analyzed, in particular, take the derivatives of analytical functions and search for their extrema [41; 42]. This sometimes, to some extent, simplifies the search for optimal engineering solutions regarding the geometry of the cathode assembly details [32; 33].

3. Numerical solution of the heat conduction equation (3). Usually, the right-hand difference method or the implicit Crank–Nicholson method is used to solve the complex non-stationary heat-conduction equation (3) [30; 31; 38]. However, if a stationary thermodynamic problem is considered, the hyperbolic heat conduction equation (3) is generally reduced to the elliptic Poisson equation [32; 33; 40; 43; 44]. The advantage of numerical methods for solving engineering problems in thermodynamics is that, if they are correctly set, it is possible to obtain high accuracy in modeling the temperature distribution, taking into account all edge effects. The general disadvantage of this approach is that, in order to obtain optimal designs of thermodynamic systems, it is necessary to analyze a large number of variants, which is usually associated with huge time costs. Today, even with the use of advanced computers and network technologies in cloud computing, solving one variant of an engineering task with the complex spatial geometry of the simulated object can take several hours [36; 39]. Another disadvantage of using this approach is that the research engineer has only the final results of the provided calculations, and after those, it is usually extremely difficult to find the optimal design from the point of view of the laws of thermodynamics. It is much easier to do this through the analysis of the simple analytical solution of a thermodynamic problem or the corresponding functional series.

4. Today, with the development of computer systems for engineering design and modeling, computer modeling methods using built-in computer calculation systems and engineering systems for automated design (CAD) are often used to estimate the temperature regimes of the HVGD cold cathode. Such methods of modeling thermodynamic systems are based on the numerical solution of the heat conduction equation (3) using the implicit Crank–Nicholson scheme, and calcula-

tions are performed using the finite element method [40; 43; 44]. When using this method, the accuracy of calculations for axially symmetric cathode cooling systems is proportional to the minimum dimensions of the calculation elements of the finite-difference cells and is estimated by the following ratio [43; 44]:

$$\varepsilon_2 = \sqrt{\left(\varepsilon_1 + \frac{\partial T(r,z)}{\partial r} h_r\right)^2 + \left(\varepsilon_1 + \frac{\partial T(r,z)}{\partial z} h_z\right)^2}, \qquad (5)$$

where ε_1 is the temperature calculation error at nodal points; ε_2 is the error of approximation of the temperature dependence between nodes; h_r is the discretization step along the r coordinate, and h_z is the discretization step along the z coordinate.

It is clear from relation (5) that to increase the accuracy of simulation of HVGD cathode cooling systems, the density of the finite-difference grid should be increased in regions with a high temperature gradient. Usually, such areas are the boundaries between elements of the cooling system, which have different values of thermal conductivity.

The work [28] presented the results of modeling the cooling system of the cathode unit through the copper base in the case of considering a gap between the cathode and the base and without such a gap. The simulation was carried out in the **pdetool** program of the MatLab system for scientific and technical calculations [28; 40].

COMPUTER SIMULATION TOOLS OF THE SOLIDWORKS CAD SOFTWARE

The work [28] considered the means of modeling the cooling system of the cathode unit of the powerful HVGD electron gun, implemented in the modern **SolidWorks** software complex, in the **Flow Simulation** program [45; 46]. The advantage of using this software tool in engineering activities is that it is designed specifically for drawing and has high-quality graphic means for visualizing complex three-dimensional parts, assemblies and structures. In the **SolidWorks** software complex, drawings of individual parts of the cathode assembly are first created, and then, through the connection of these parts, an assembly drawing is performed with the possibility of three-dimensional visualization of the entire structure [45; 46]. After that, the final version of the drawing of the cathode assembly is uploaded to the **Flow Simulation** program, where the power of the heat flow that is applied to the surface of the cathode is set. Manual and automatic settings of the rate of the coolant are also possible. The **Flow Simulation** software has its own database on the necessary values of thermodynamic parameters of metals, ceramics, organic materials, as well as liquids and gases [45; 46], so it is enough to specify most of the material from which the corresponding part of the cathode assembly is made, the coolant and, if necessary, the type of operation gas and its pressure [45; 46]. Basic mechanical, thermodynamic, aerodynamic, and electrophysical properties of structural materials, given in reference literature [34, 35], already entered into the database of the **SolidWorks** CAD software [45; 46].

A generalized review of the possibilities of using all the approaches described above regarding the analysis and optimization of the temperature regime of the cathode cooling systems of HVGD electron guns was generally carried out in [28].

PRELIMINARY ESTIMATES OF THE TEMPERATURE OF THE COLD CATHODE SURFACE OF A HIGH-VOLTAGE GLOW DISCHARGE ELECTRON GUN, INTENDED FOR THE FORMATION OF A RIBBON ELECTRON BEAM, USING THE HEAT BALANCE EQUATION, ARITHMETIC-LOGICAL RELATIONS, AND RVACHOV FUNCTIONS

The structural diagram of the electrode system of the HVGD device, which forms a ribbon electron beam, is shown in Fig. 2. The device intended for the formation of a ribbon beam contains a cathode 1, which is fixed on a high-voltage insulator 4, and is located in a capsule 2. The cooling of the cathode surface with water is carried out directly, cold water enters through tubes that are passed through the insulator. The length of the rubber hoses used to supply water should be such that they provide reliable electrical insulation, taking into account the fact that a voltage of 15 kV is applied to the cathode.

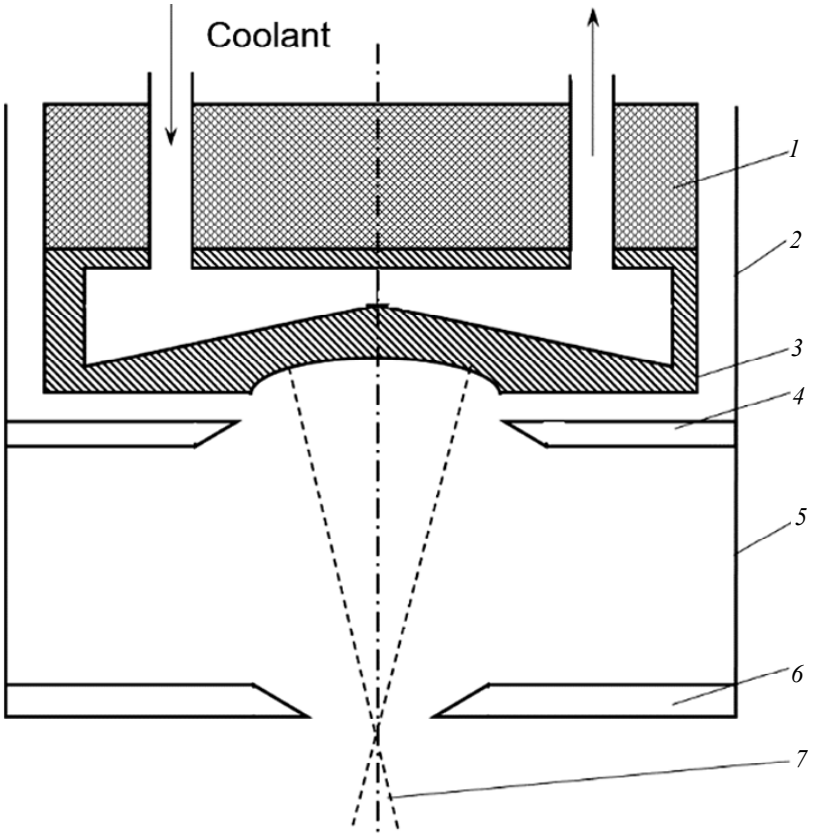


Fig. 2. Structural diagram of the HVGD electrode system designed to form a ribbon electron beam: *1* — insulator; *2* — capsule; *3* — cathode; *4* — near-cathode diaphragm; *5* — anode; *6* — base flange for mounting the gun on the technological chamber with a hole for the output of the electron beam; *7* — ribbon electron beam with a linear focus

The capsule of the device, taking into account the near-cathode diaphragm 4 and the base flange 6 with a hole for outputting the electron beam, forms a discharge chamber in which a high-voltage discharge burns [1–3]. Diaphragm 4, located near the cathode, ensures the distribution of the electric field necessary for the formation and focusing of the ribbon electron beam [2; 3].

The calculation of water cooling of the cathode surface for the HVGD electrode system, whose design scheme is shown in Fig. 2, was carried out taking into account the geometric features of the cathode assembly. Fig. 3 shows the construction of the cathode assembly used and the corresponding geometric parameters of the cathode. Therefore, the boundary conditions for solving thermodynamic task are presented in Fig. 4.

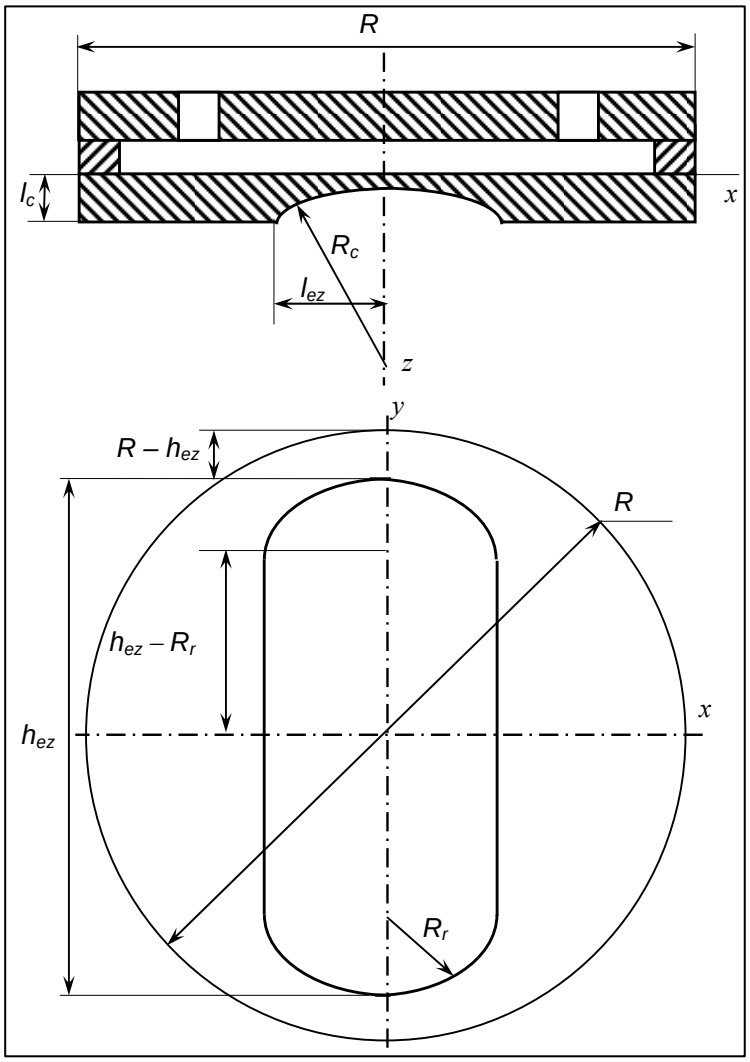


Fig. 3. Geometrical parameters of the cathode of the HVGD electrode system, the structural scheme of which is shown in *Fig. 2*: R_c is the radius of the cylindrical emission surface from which the flow of electrons is formed; R is the cathode radius as a structural element; l_c is the thickness of cathode; l_{ez} is the wideness of the cathode emission zone; h_{ez} is the longitude length of the cathode emission zone; and R_r is the radius of cathode rounding

Analytical relation (2) has been used to estimate the temperature of the surface of the HVGD cold cathode, and the geometry of the cathode assembly was described using arithmetic and logical functions [36]. Under such conditions, the arithmetic-logical relationship that was used to calculate the temperature of the cold cathode of the HVGD device, which forms a ribbon electron beam with a linear focus, has the following form:

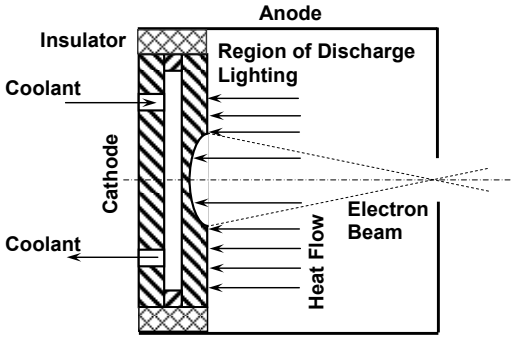


Fig. 4. The boundary conditions for calculation the heat regime of electrodes system of high voltage glow discharge electron gun

$$C_1 = ((|x| < l_{ez}) \,\&\, (|y| < h_{ez})) \cdot \left(\frac{W_c \gamma_g \dfrac{R_c + 2l_c - \sqrt{R_c^2 - x^2}}{\lambda_c}}{\alpha_c R_c^2 (1+\gamma_g) \arcsin\left(\dfrac{l_{ez} - R}{2R_c}\right)} + T_{cl} \right),$$

$$C_2 = ((|x| \ge l_{ez}) \,\&\, (|y| < h_{ez}) \,\&\, (|x| \le R)) \left(\frac{W_c \gamma_g (R - l_{ez})}{\alpha_c R_c^2 \lambda_c (1+\gamma_g) \arcsin\left(\dfrac{R - R_r}{2R_c}\right)} + T_{cl} \right),$$

$$C_3 = \left(\left(|x| < \sqrt{R_r^2 - (|y| < h_{ez})^2} \right) \,\&\, (|y| \ge h_{ez}) \,\&\, (|y| \le (h_{ez} + R)) \right) \times$$

$$\times \left(\frac{W_c \gamma_g \cdot \left(R_r + R_c - \sqrt{R_c^2 - x^2} \right)}{\alpha_c R_k^2 \lambda_k (1+\gamma_g) \arcsin\left(\dfrac{R - R_r}{2R_c}\right)} + T_{cl} \right), \qquad (6)$$

$$C_4 = \left(\left(|x| \ge \sqrt{R_r^2 - (|y| < h_{ez})^2} \right) \,\&\, (|y| \ge h_{ez}) \,\&\, (|y| \le (h_{ez} + R)) \right) \times$$

$$\times \left(\frac{W_c \gamma_g l_c}{\alpha_c \lambda_c (1+\gamma_g) \left(R - \sqrt{R_r^2 - (|y| - h_{ez})^2} \right)^2 \arcsin\left(\dfrac{R - R_r}{2R_c}\right)} + T_{cl} \right),$$

$$C_5 = (|y| > h_{ez}) \left(\frac{W_c \gamma_g l_c}{\alpha_c R_c^2 (1+\gamma_g)} + T_{cl} \right), \quad T_c(x,y) = C_1 + C_2 + C_3 + C_4 + C_5,$$

where $T_c(x,y)$ is temperature distribution function on the cathode surface by x and y coordinates; α_c is the heat transfer coefficient of the coolant liquid through the base of the cathode according to the second ratio of equations set (2).

Calculations using relations (6) were carried out using logical and matrix programming tools of the MatLab -system of scientific and technical calculations for the following parameters of the cathode assembly: $T_{cl} = 10\,^{\circ}\mathrm{C}$, $W_c - 1 - 5\,\mathrm{kW}$, $\gamma_g = 3$, $R_c = 0.05\,\mathrm{m}$, $R = 0.03\,\mathrm{m}$, $R_r = 0.012\,\mathrm{m}$, $l_c = 0.005\,\mathrm{m}$, $l_{ez} = 0.012\,\mathrm{m}$, $h_{ez} = 0.026\,\mathrm{m}$, $v_c = 0.03\,\mathrm{m}^3/\mathrm{s}$, $k_1 = 350\,k$ and $k_2 = 21000$.

The simulation results obtained using arithmeticlogical relations (6) are shown in Fig. 5.

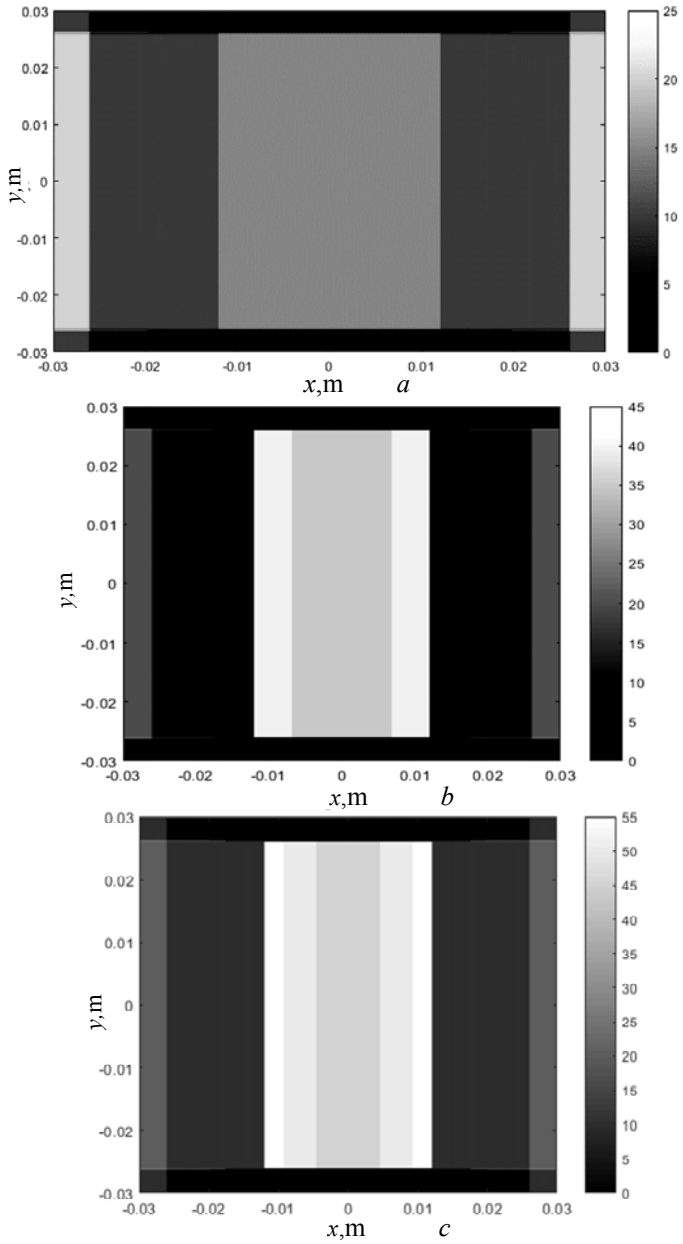


Fig. 5. Contour graphs of the temperature distribution on the cold cathode surface $T_c(x, y)$, obtained using arithmetic-logic relation (6): *a* — $W_c = 1\,\mathrm{kW}$; *b* — $W_c = 5\,\mathrm{kW}$; *c* — $W_c = 7\,\mathrm{kW}$

From the graphic dependences obtained using arithmetic-logic relations (6) and shown in Fig. 5, it can be seen that the preliminary, generalized estimates based on the use of the heat balance equation indicate that in the range of power released at the cathode, within the range of $W_c = 1-7$ kW, the surface temperature of the water-cooled aluminum cathode does not exceed 60°C. Such a temperature mode of operation is normal for the cold cathode of technological electron sources, based on HVGD [1–3].

Another, more accurate approach to describing the geometry of complex three-dimensional objects is the use of Rvachev functions, or locus [37–39]. The essence of this approach is that instead of the logical ratios used in the analytical expression (6), atomic functions or locus are used. Locus is smooth functions with a high value of the derivative. Therefore, with their help, a jump-like coordinate difference is approximated for the surface being modulated. Locus functions are formed on the basis of arithmetic-logical relations. This approach makes it possible to effectively use locus for solving problems of electrophysics and thermodynamics in real devices with complex electrode geometry [37–39].

Analytical expressions for locus functions are given in Table 1 [35–37].

Table 1. Rvachov R-functions R_α, or locus

The accompanying logical function of the Boolean algebra	Set of functions R_α
$F_1(x_1, x_2) = x_1 \wedge x_2$	$y_1(x_1, x_2) = x_1 \wedge_\alpha x_2 = \frac{1}{1+\alpha} \times \left(x_1 + x_2 - \sqrt{x_1^2 + x_2^2 - 2\alpha x_1 x_2} \right)$
$F_2(x_1, x_2) = x_1 \vee x_2$	$y_2(x_1, x_2) = x_1 \vee_\alpha x_2 = \frac{1}{1+\alpha} \times \left(x_1 + x_2 + \sqrt{x_1^2 + x_2^2 - 2\alpha x_1 x_2} \right)$
$F_3(x_1) = \bar{x}_1$	$y_3(x_1) = \bar{x}_1 = -x_1$

The coefficient α in the ratios given in Table 1 is selected in the range [0; 1], depending on the formulation of the modeling task.

Let's rewrite the logical expressions of the arithmetic-logical functions (6) in terms of the Rvachev R_α functions defined in Table 1.

$$L_{11} = (l_{ez} - |x|); \quad L_{12} = (h_{ez} - |y|);$$

$$R_1(L_{11}, L_{12}) = L_{11} \wedge_\alpha L_{12} = \frac{1}{1+\alpha} \times \left((l_{ez} - |x|) + (h_{ez} - |y|) - \right.$$

$$\left. - \sqrt{(l_{ez} - |x|)^2 + (h_{ez} - |y|)^2 - 2\alpha (l_{ez} - |x|)(h_{ez} - |y|)} \right);$$

$$L_{21} = (|x| - l_{ez}), \quad L_{22} = (h_{ez} - |y|), \quad L_{23} = (R - |x|);$$

$$R_{21}(L_{21}, L_{22}) = L_{21} \wedge_\alpha L_{22} = \frac{1}{1+\alpha} \times \left((|x| - l_{e3}) + (h_{e3} - |y|) - \right.$$

$$\left. - \sqrt{(|x| - l_{ez})^2 + (h_{ez} - |y|)^2 - 2\alpha (|x| - l_{ez})(h_{ez} - |y|)} \right);$$

$$R_2(R_{21}, L_{23}) = R_{21} \wedge_\alpha L_{23} = \frac{1}{1+\alpha} \times \left(R_{21} + (R - |x|) - \right. \quad (7)$$

$$\left. - \sqrt{R_{21}^2 + (R - |x|)^2 - 2\alpha R_{21} (R - |x|)} \right);$$

$$L_{31} = |x| - \sqrt{R_r^2 - (|y| < h_{e3})^2}\,, L_{32} = |y| - h_{ez}\,, L_{33} = (h_{ez} + R) - |y|;$$

$$R_{31}(L_{31}, L_{32}) = L_{31} \wedge_{\alpha} L_{32} = \frac{1}{1+\alpha} \times \left(\left(|x| - \sqrt{R_r^2 - (|y| - h_{ez})^2} \right) + (|y| - h_{ez}) - \right.$$

$$\left. - \sqrt{\left(|x| - \sqrt{R_r^2 - (|y| - h_{ez})^2} \right)^2 + (|y| - h_{ez})^2 - 2\alpha \left(|x| - \sqrt{R_r^2 - (|y| < h_{ez})^2} \right)(|y| - h_{ez})} \right);$$

$$R_3(R_{31}, L_{33}) = R_{31} \wedge_{\alpha} L_{33} = \frac{1}{1+\alpha} \times \left(R_{31} + ((h_{ez} + R) - |y|) \right. -$$

$$\left. - \sqrt{R_{31}^2 + ((h_{ez} + R) - |y|)^2 - 2\alpha R_{31}((h_{ez} + R) - |y|)} \right);$$

$$L_{41} = \sqrt{R_r^2 - (|y| < h_{ez})^2} - |x|, L_{42} = |y| - h_{ez}\,, L_{43} = (h_{ez} + R) - |y|;$$

$$R_{41}(L_{41}, L_{42}) = L_{41} \wedge_{\alpha} L_{42} = \frac{1}{1+\alpha} \times \left(\left(\sqrt{R_r^2 - (|y| - h_{ez})^2} - |x| \right) + (|y| - h_{ez}) - \right.$$

$$\left. - \sqrt{\left(\sqrt{R_r^2 - (|y| - h_{ez})^2} - |x| \right)^2 + (|y| - h_{ez})^2 - 2\alpha \left(\sqrt{R_r^2 - (|y| < h_{ez})^2} - |x| \right)(|y| - h_{ez})} \right);$$

$$R_4(R_{41}, L_{43}) = R_{41} \wedge_{\alpha} L_{43} = \frac{1}{1+\alpha} \times \left(R_{41} + ((h_{ez} + R) - |y|) \right. -$$

$$\left. - \sqrt{R_{41}^2 + ((h_{ez} + R) - |y|)^2 - 2\alpha R_{41}((h_{ez} + R) - |y|)} \right);$$

$$L_5 = |y| - h_{ez};$$

$$R_5(L_5, 1) = L_5 \wedge_{\alpha} 1 = \frac{1}{1+\alpha} \times \left((|y| - h_{ez}) + 1 - \sqrt{((|y| - h_{ez}))^2 - 2\alpha(|y| - h_{ez}) + 1} \right).$$

Then the sophisticated arithmetic-logical relation (6), taking into account (7), will be rewritten as follows:

$$T_c(x, y) = R_1 \left(\frac{W_c \gamma_g \dfrac{R_c + 2l_c - \sqrt{R_c^2 - x^2}}{\lambda_c}}{\alpha_c R_c^2 (1 + \gamma_g) \arcsin\left(\dfrac{l_{ez} - R}{2R_c} \right)} + T_{cl} \right) +$$

$$+ R_2 \left(\frac{W_c \gamma_g (R - l_{ez})}{\alpha_c R_c^2 \lambda_c (1 + \gamma_g) \arcsin\left(\dfrac{R - R_r}{2R_c} \right)} + T_{cl} \right) +$$

$$+R_3\left(\frac{W_c\gamma_g\left(R_r+R_c-\sqrt{R_c^2-x^2}\right)}{\alpha_c R_c^2\lambda_c(1+\gamma_g)\arcsin\left(\frac{R-R_r}{2R_c}\right)}+T_{cl}\right)+ \tag{8}$$

$$+R_4\cdot\left(\frac{W_c\gamma_g l_c}{\alpha_c\lambda_c(1+\gamma_g)\left(R-\sqrt{R_r^2-(|y|-h_{ez})^2}\right)^2\arcsin\left(\frac{R-R_r}{2R_c}\right)}+T_{cl}\right)+$$

$$+R_5\left(\frac{W_c\gamma_g l_c}{\alpha_c R_c^2(1+\gamma_g)}+T_{cl}\right).$$

The results of calculating the temperature distribution on the surface of the cathode using content (8), which have been obtained under the condition α = 0.95, are shown in Fig. 6.

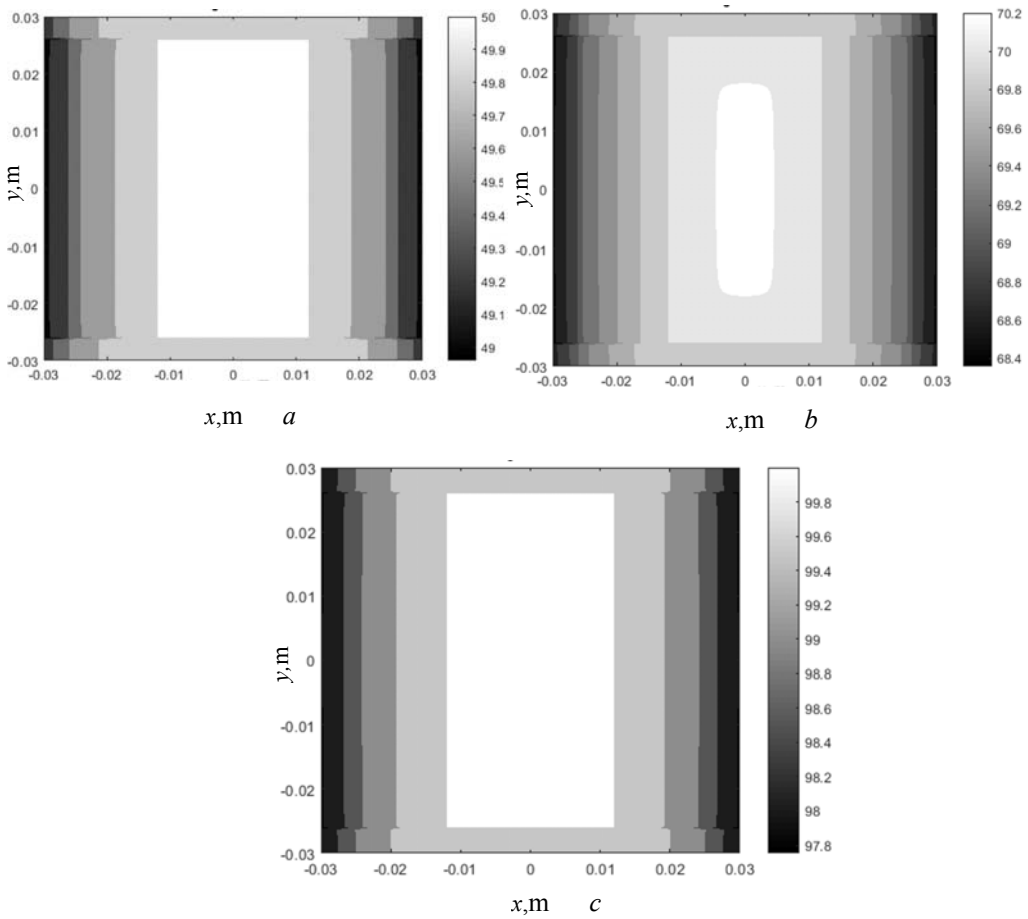


Fig. 6. Contour graphs of temperature distribution on the surface of the cold cathode $T_c(x, y)$, obtained using relation (8): *a* — $W_c = 1$ kW; *b* — $W_c = 3$ kW; *c* — $W_c = 5$ kW

ESTIMATION OF THE SURFACE TEMPERATURE OF THE COLD CATHODE USING MODELING TOOLS OF THE SOLIDWORKS CAD SOFTWARE

The CAD software system **SolidWorks** is a well-known and widely used package that allows one to create 3D models of parts based on drawings or sketches. That is why this engineering CAD is often used by designers and industrialists. The system also has the powerful algorithms for calculating loads of various types. The calculations are based on the physical parameters of materials implemented in the program, which include almost all widely used metals, alloys, nonmetals, and liquids. Mechanical loads and deformations, as well as thermal loads can be simulated. It is also possible in this CAD system to study the aerodynamics or hydrodynamics of a simulated object in a liquid or gas flow [45; 46].

Today, the creation of an engineering design of any device cannot be done without the prior use of its simulation tools. High-voltage devices, in case of failure during testing, pose a significant danger, therefore, it is much more efficient and safer to make their preliminary modeling and corresponding simple calculations. Modern systems of automatic design and their corresponding tools allow, with sufficient accuracy, to simulate the conditions and loads to which a real model of the device will be subjected. The **SolidWorks** CAD system has several internal applications for modeling loads of various types [45; 46].

In the course of the provided engineering researches described in this section of paper, the application **Flow Simulation** has been used [45; 46]. This program allows you to simulate flows of liquids or gases within certain geometric limits, taking into account the heat exchange between interacting solid-states materials, liquids, and gases. Since this software is primarily focused on the development of three-dimensional models, it calculates the physical parameters of materials with a certain value of error, which does not allow obtaining an absolutely accurate result. However, usually this accuracy of calculations is sufficient to draw general conclusions regarding the correctness of testing and applying the system being researched and designed [45; 46].

To simulate the design of the device in the **SolidWorks** CAD system, one need to make an assembly drawing or a necessary detail in the form of a three-dimensional model. In the provided and described engineering research, an assembly drawing of an electronic device intended for the formation of an electron beam with a ribbon focus has been made. First of all, one needs to open the assembly drawing file and click on the option **Flow Simulation**.

In the next step, one needs to create a new project with specification of physical quantities, materials, and initial conditions for modeling.

Then, by selecting the **Calculation Area** option, one can select the space where the object of research is located. Before this, it is necessary to make an axisymmetric section of the assembled drawing for better visualization [28; 45; 46].

As noted in Section 4 of the paper, the **SolidWorks** CAD system has its own powerful database of mechanical, electrical, and electrophysical properties of various structural materials. Therefore, the next step in simulation is the selection of materials from which separate parts of the structure are made. If the necessary material is not available, it is possible to create a new record in the database and enter the necessary parameters of the material there [45; 46].

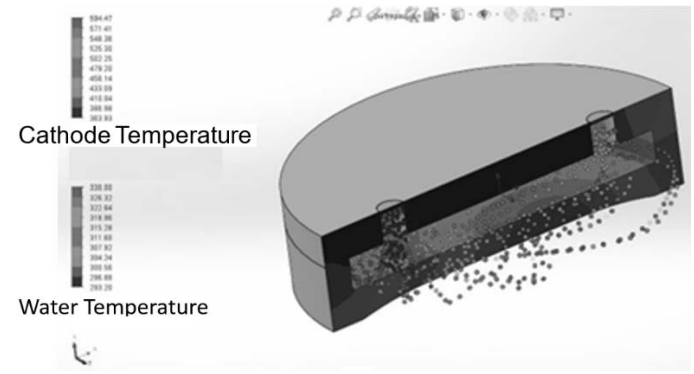
Having chosen the materials for the parts and assemblies of the device, one can move on to defining the boundary conditions for the operation of the device being modeled [45; 46]. The parameters of the model for studying the temperature mode of operation of the cathode unit of the HVGD electron gun, which forms a ribbon beam with a linear focus, are given in Table 2.

Table 2. The parameters of the cathode assembly model of the HVGD electron gun, which forms an electron beam with a linear focus, created in the **SolidWorks** CAD system

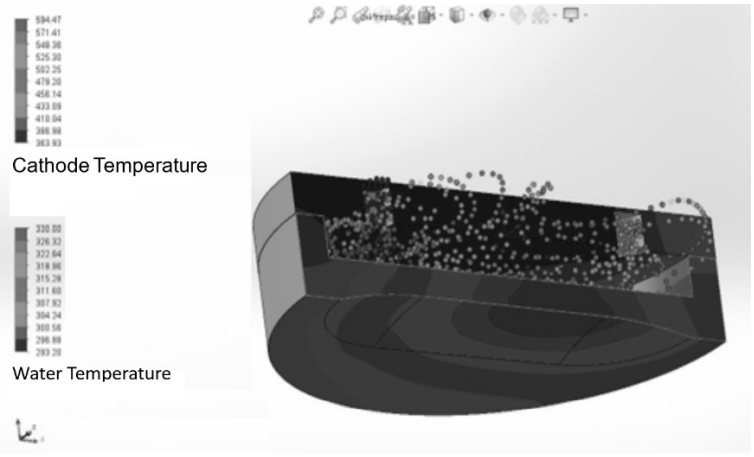
Model parameter	Value	
A model of the heat conduction in a solid body	Enabled	
Only heat conduction in a solid	Disabled	
Radiation heat exchange	Disabled	
Heat exchange in gas by radiation	Disabled	
Non-stationarity	Disabled	
Gravitational effects	Disabled	
Rotation	Local Area (Averaging)	
The power released at the cathode	2 kW	
Type of liquid flow	Laminar and turbulent	
High-value Mach number flow model	Disabled	
Cavitation	Enabled	
Concentration of dissolved gas by mass, kg/m^3	10^{-4}	
Free surface	Disabled	
Surface roughness, μm	0.1	
Water flow rate in the normal direction to the cathode surface, m/s	10	
Water consumption, m^3/s	10^{-5}	
Thermal conditions on the outer walls	Adiabatic wall	
Number of nodes in the finite-elements mesh	$5 \cdot 10^3$	
Thermodynamic parameters	Static pressure, Pa	Temperature, °K
	101325	293.2
Intensity and scale of turbulence	Intensity, %	Scale (length, m)
	2	$1.5 \cdot 10^{-4}$

ANALYSIS OF OBTAINED SIMULATION RESULTS AND PRACTICAL RECOMMENDATIONS

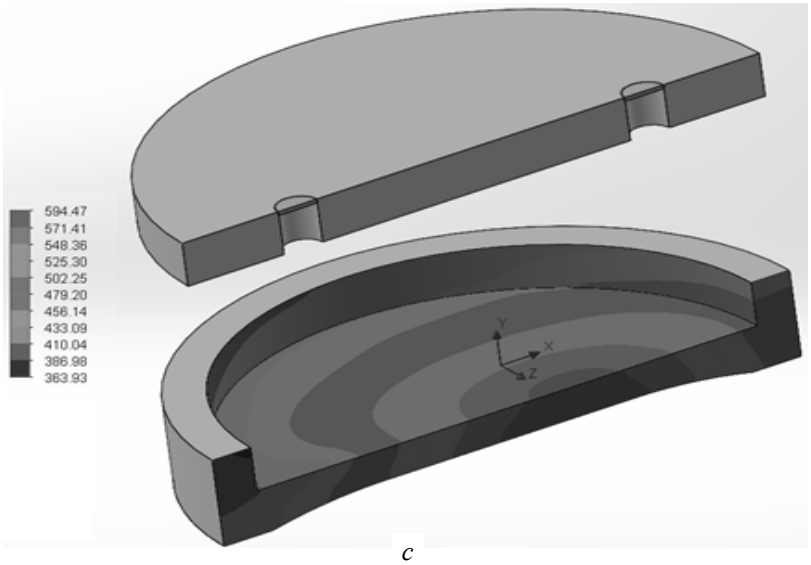
Fig. 7 shows the results of modeling the thermal modes of operation of the cathode node of an electronic device that forms a ribbon beam with a linear focus, obtained using the **Flow Simulation** program in the **SolidWorks** engineering computer CAD system. The obtained results largely coincide with the results obtained by numerically solving the heat balance equation and modeling the geometry of the cathode node by arithmetic-logical relations or locus functions, which are shown in Fig. 6.



a



b



c

Fig. 7. The temperature distribution of the cathode assembly T_c (x, y, z), obtained using modeling tools and the design toolkit of the SolidWorks CAD software

The results of modeling the temperature regime of the cathode assembly obtained in the provided research generally indicate that a simplified approach to solving the heat conduction equation (3), connected with the transition to the analytical solution of the heat balance equation (2), taking into account the real geometry of the cooling constructive elements of the electronic gun through the use of arithmetic-logical relations, is quite effective, and the accuracy obtained from modeling is usually sufficient for preliminary generalized estimates of the thermal mode of operation of structural units at the initial stage of gun design. The method of describing the complex geometry of construction details by locus or Rvachev functions [28; 37–39] is also effective. To calculate the thermal regime of the cathode assembly, we used relation (2), which is quite universal and was used in work [28] to calculate the surface temperature of the cathode with direct cooling in electron guns that form beams with a point focus. The difference of the proposed model lies in the fact that the real geometry of the cathode assembly for the HVGD electron guns, which forms a ribbon electron beam, was described using the appropriate arithmetic-logical ratios. It should also be noted that the accuracy of thermal calculations in the SolidWorks software complex is also low, about 20%. However, an important advantage of this system, from a practical point of view, is the possibility of numerical calculations for real structures. In this regard, if the appropriate mathematical apparatus is available, the proposed model of the cathode node, based on the use of locus, or Rvachev functions [28; 37–39], can be recommended to designers of electron beam technological equipment for further practical use. To carry out more accurate estimates of the temperature of the surface of the cathode, taking into account the boundary effects, in work [28] it was recommended to use the **pdetool** toolkit of the MatLab system of scientific and technical calculations, however, working with this program requires certain qualifications and knowledge of the appropriate mathematical apparatus from the designers.

The testing experiments in SolidWorks software were realized in the such range of geometry parameters of cathode item: $R_c = 0.05 \pm 0.01$ m, $R = 0.03 \pm$ ± 0.01 m, $R_r = 0.012 \pm 0.01$ m, $l_c = 0.005 \pm 0.001$ m, $l_{ez} = 0.012 \pm 0.001$ m. By the power of electron gun regimes, range of 1–150 kW are considered. Pointed out above error of simulation, corresponding to the obtained experimental results, have been in the same range. The dependences of focal beam parameters on geometry of electrodes' system and technological tolerances have been analyzed in the papers [22–24].

The scientific results, given in this article, have been obtained in the Scientific Laboratory of Electron Beam Technological Devices of National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnical Institute" customed by "Chervona Khvyla" Open Joint Stock Company.

These results are of interest to a wide range of specialists who are engaged in the development and introduction into production of modern electron-beam technological equipment.

CONCLUSION

The paper considers various theoretical approaches to estimating the operating temperature of the surface of the cold cathode of the HVGD electron gun, which

forms a ribbon-like converging electron beam with a linear focus. Based on the test calculations that have been performed, it is shown that the accuracy of estimates using ratio (2), obtained as a solution of the system of heat balance equations (1), is quite high. For the considered models of HVGD electron guns, the calculated surface temperature of the cathode, in the case of using Rvachev arithmetic-logical relations and functions (8) to describe the geometry of the cathode assembly, is about 5–10% and is close to the estimated results obtained using the **SolidWorks** software complex. To carry out more accurate estimates, taking into account edge effects and the influence of the temperature of the peripheral region of the cathode on the emission of electrons from its surface, you can use either estimates through functional series using relation (4) or the **pdetool** toolkit of the MatLab scientific and technical calculations system. The method of using ratio (4) to estimate the surface temperature of the cathode of the HVGD electron gun, which forms a beam with a point focus, as well as the corresponding mathematical approaches for the use of special functions, were considered and analyzed in work [28]. Such evaluations can be interesting and useful for developers of new types of HVGD electron guns with improved emission characteristics. To carry out such calculations, a sufficiently high qualification of development engineers with appropriate knowledge of mathematical functions and corresponded software is required. In the case of preliminary evaluations, without taking into account the influence of the temperature of the peripheral region of the cathode surface on its emission characteristics, calculations using arithmetic-logical relations and Rvachev functions similar to relations (8) are sufficient. To carry out such assessments, it is also necessary to develop suitable mathematical approaches and appropriate modifications of the software, since each design of the HVGD electron guns for technological purposes has its own specific features [2; 3].

Main recommendations to engineers, who designed HVGD electron guns, have been obtained as a results of scientific research, are as follows.

1. For electron guns with the power range of 1–5 kW using of cathode item without cooling for simplifying the gun construction is possible.

2. For electron guns with the power range of 5–30 kW using of cooling cathode through cooper base for simplifying the gun construction is possible.

3. For electron guns with the power, grater, then 30 kW, the direct cooling of the cathode surface with coolant is necessary. The scheme of corresponded gun construction is given in Fig. 2, and the geometrical parameters are noted at Fig. 3. The temperature of cathode emission surface for the cathode from aluminum has to be in range 150–200 $^{\circ}$C [2; 28].

The theoretical assumptions presented in the article are of great practical importance regarding the creation of new types of structures for HVGD electron guns and their industrial application. The results presented in the article may be interesting and useful for a wide range of specialists who are engaged in the development of modern electron-beam equipment and its application in industry.

REFERENCE

1. I. Melnyk, S. Tuhai, and A. Pochynok, “Universal Complex Model for Estimation the Beam Current Density of High Voltage Glow Discharge Electron Guns,” *Lecture Notes in Networks and Systems*; Eds: M. Ilchenko, L. Uryvsky, and L. Globa, 152, pp. 319–341, 2021. Available: https://www.springer.com/gp/book/ 9783030583583

2. S. Denbnovetsky et al., "Principles of operation of high voltage glow discharge electron guns and particularities of its technological application," *Proceedings of SPIE, The International Society of Optical Engineering,* pp. 10445–10455, 2017. Available: https://spie.org/Publications/Proceedings/Paper/10.1117/12.2280736
3. S.V. Denbnovetsky, V.G. Melnyk, and I.V. Melnyk, "High voltage glow discharge electron sources and possibilities of its application in industry for realizing of different technological operations," *IEEE Transactions on plasma science*, vol. 31, no. 5, pp. 987–993, October, 2003. Available: https://ieeexplore.ieee.org/document/1240048
4. S.V. Denbnovetsky, V.I. Melnyk, I.V. Melnyk, and B.A. Tugay, "Model of control of glow discharge electron gun current for microelectronics production applications," *Proceedings of SPIE. Sixth International Conference on "Material Science and Material Properties for Infrared Optoelectronics"*, vol. 5065, pp. 64–76, 2003. doi: 10.1117/12.502174.
5. I.V. Melnyk and S.B. Tugay, "Analytical calculations of anode plasma position in high-voltage discharge range in case of auxiliary discharge firing," *Radioelectronic and Communication Systems*, vol. 55, no. 11, pp. 50–59, 2012. Available: http://radio.kpi.ua/article/view/S0021347012110064
6. A.A. Druzhinin, I.P. Ostrovskii, Y.N. Khoverko, N.S. Liakh-Kaguy, and A.M. Vuytsyk, "Low temperature characteristics of germanium whiskers," *Functional materials*, 21 (2), pp. 130–136, 2014. Available: https://nanoscalereslett.springeropen.com/ articles/10.1186/s11671-017-1923-1
7. A.A. Druzhinin, I.A. Bolshakova, I.P. Ostrovskii, Y.N. Khoverko, and N.S. Liakh-Kaguy, "Low temperature magnetoresistance of InSb whiskers," *Materials Science in Semiconductor Processing*, 40, pp. 550–555, 2015. Available: https://academic-accelerator.com/search?Journal=Druzhinin
8. A. Zakharov, S. Rozenko, S. Litvintsev, and M. Ilchenko, "Trisection Bandpass Filter with Mixed Cross-Coupling and Different Paths for Signal Propagation," *IEEE Microwave Wireless Components Letters*, vol. 30, no. 1, pp. 12–15, 2020.
9. A. Zakharov, S. Litvintsev, and M. Ilchenko, "Trisection Bandpass Filters with All Mixed Couplings," *IEEE Microwave Wireless Components Letters*, vol. 29, no. 9, pp. 592–594, 2019. Available: https://ieeexplore.ieee.org/abstract/document/8782802
10. A. Zakharov, S. Rozenko, and M. Ilchenko, "Varactor-tuned microstrip bandpass filter with loop hairpin and combline resonators," *IEEE Transactions on Circuits Systems, II, Exp. Briefs*, vol. 66, no. 6, pp. 953–957, 2019. Available: https://ieeexplore.ieee.org/document/8477112
11. A. Zakharov, S. Litvintsev, and M. Ilchenko, "Transmission Line Tunable Resonators with Intersecting Resonance Regions," *IEEE Transactions on Circuits Systems II, Exp. Briefs*, vol. 67, no. 4, pp. 660–664, 2020.
12. T.O. Prikhna et al., "Electron-Beam and Plasma Oxidation-Resistant and Thermal-Barrier Coatings Deposited on Turbine Blades Using Cast and Powder Ni(Co)CrALY(Si) Alloys I. Fundamentals of the Production Technology, Structure, and Phase Composition of Cast NiCrAlY Alloys," *Powder Metallurgy and Metal Ceramics*, 61(1-2), pp. 70–76, *2022*. Available: https://www.springer.com/journal/11106
13. T.O. Prikhna et al., "Electron-Beam and Plasma Oxidation-Resistant and Thermal-Barrier Coatings Deposited on Turbine Blades Using Cast and Powder Ni(Co)CrAlY(Si) Alloys Produced by Electron-Beam Melting II. Structure and Chemical and Phase Composition of Cast CoCrAlY Alloys," *Powder Metallurgy and Metal Ceramicsthis*, 61(3-4), pp. 230–237, 2022.
14. I.M. Grechanyuk et al., "Electron-Beam and Plasma Oxidation-Resistant and Thermal-Barrier Coatings Deposited on Turbine Blades Using Cast and Powder Ni(Co)CrAlY(Si) Alloys Produced by Electron Beam Melting IV. Chemical and Phase Composition and Structure of Cocralysi Powder Alloys and Their Use," *Powder Metallurgy and Metal Ceramics*, 61(7-8), pp. 459–464, 2022.
15. M.I. Grechanyuk et al., "Electron-Beam and Plasma Oxidation-Resistant and Thermal-Barrier Coatings Deposited on Turbine Blades Using Cast and Powder Ni

(Co)CrAlY (Si) Alloys Produced by Electron Beam Melting III. Formation, Structure, and Chemical and Phase Composition of Thermal-Barrier Ni(Co)CrAlY/ZrO_2–Y_2O_3 Coatings Produced by Physical Vapor Deposition in One Process Cycle," *Powder Metallurgy and Metal Ceramics*, 61(5-6), pp. 328–336, 2022.

16. V. Vassilieva, K. Vutova, and V. Donchev, "Recycling of alloy steel by electron beam melting," *Electrotechnics and Electronics (E+E)*, vol. 47, no. 5–6, pp. 142–145, 2012.
17. H. Sasaki, Y. Kobashi, T. Nagai, and M. Maeda, "Application of electron beam melting to the removal of phosphorous form silicon: toward production of solar-grade silicon by metallurgical process," *Advanced in material science and engineering*, vol. 2013, article ID 857196. Available: https://cyberleninka.org/article/n/299910/viewer
18. T. Kemmotsu, T. Nagai, and M. Maeda, "Removal Rate of Phosphorous form Melting Silicon," *High Temperature Materials and Processes*, vol. 30, no. 1–2, pp. 17–22, 2011. Available: https://www.degruyter.com/journal/key/htmp/30/1-2/html
19. J.C.S. Pires, A.F.B. Barga, and P.R. May, "The purification of metallurgically grade silicon by electron beam melting," *Journal of Materials Processing Technology*, vol. 169, no. 1, pp. 347–355, 2005. Available: https://www.academia.edu/9442020/The_purification_of_metallurgical_grade_silicon_by_electron_beam_melting
20. D. Luo, N. Liu, Y. Lu, G. Zhang, and T. Li, "Removal of impurities from metallurgically grade silicon by electron beam melting," *Journal of Semiconductors*, vol. 32, no. 3, Article ID 033003, 2011. Available: http://www.jos.ac.cn/en/article/doi/10.1088/1674-4926/32/3/033003
21. A.E. Davis et all., "Tailoring equiaxed β-grain structures in Ti-6Al-4V coaxial electron beam wire additive manufacturing," *Materialia*, vol. 20, December 2021, 101202. Available: https://www.sciencedirect.com/science/article/abs/pii/S2589152921002052
22. I.V. Melnik, "Simulation of geometry of high voltage glow discharge electrodes' systems, formed profile electron beams," *Proceedings of SPIE (2006), Volume 6278, Seventh Seminar on Problems of Theoretical and applied Electron and Ion Optics*, pp. 627809-1–627809-13. Available: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/6278/627809/Simulation-of-geometry-of-high-voltage-glow-discharge-electrodes-systems/10.1117/12.693202.short
23. I.V. Melnyk and A.V. Pochynok, "Modeling of electron sources for high voltage glow discharge forming profiled electron beams," *Radioelectronic and Communication Systems*, vol. 62, no. 6 (684), pp. 311–323, 2019. Available: http://radioelektronika.org/issue/view/2019-06
24. I. Melnyk, V. Melnyk, B. Tugai, S. Tuhai, N. Mieshkova, and A. Pochynok, "Simplified Universal Analytical Model for Defining of Plasma Boundary Position in the Glow Discharge Electron Guns for Forming Conic Hollow Electron Beam," *2019 IEEE 39th International Conference on Electronics and Nanotechnology (ELNANO). Conference proceedings. April 16-18, 2019, Kyiv, Ukraine*, pp. 548–552. Available: https://ieeexplore.ieee.org/xpl/conhome/8767103/proceeding?isnumber=8783212&refinementName=Author&refinements=Author:I.V.%20Melnyk
25. J.I. Etcheverry, N. Mingolo, J.J. Rocca, and O.E. Martınez, "A Simple Model of a Glow Discharge Electron Beam for Materials Processing," *IEEE Transactions on Plasma Science*, vol. 25, no. 3, pp. 427–432, 1997.
26. I.V. Melnyk, "Numerical simulation of distribution of electric field and particle trajectories in electron sources based on high-voltage glow discharge," *Radioelectronic and Communication Systems*, vol. 48, no. 6, pp. 61–71, 2005. Available: http://radioelektronika.org/article/view/S0735272705060087
27. S.V. Denbnovetsky, J. Felba, V.I. Melnik, and I.V. Melnik, "Model of Beam Formation in A Glow Discharge Electron Gun With a Cold Cathode," *Applied Surface Science*, 111, pp. 288–294, 1997. Available: https://www.sciencedirect.com/science/ article/pii/S0169433296007611?via%3Dihub
28. I. Melnyk, S. Tuhai, M. Surzhykov, I. Shved, V. Melnyk, and D. Kovalchuk, "Different Approaches for Analytic and Numerical Estimation of Operation Temperature

of Cooled Cathode Surface in High Voltage Glow Discharge Electron Guns," in *Ilchenko M., Uryvsky L., Globa L.: editors. Progress in Advanced Information and Communication Technology and Systems. MCiT 2021. Lecture Notes in orks and Systems*, vol. 548. Springer, 2023, Cham, pp. 575–595. Available: https://link.springer.com/book/10.1007/978-3-031-16368-5;

29. A.F. Tseluyko, V.T. Lazurik, D.L. Ryabchikov, V.I. Maslov, and I.N. Sereda, "Experimental study of radiation in the wavelength range 12.2-15.8 nm from a pulsed high-current plasma diode," *Plasma Physics Reports*, 34(11), pp. 963–968, 2008. Available: https://link.springer.com/article/10.1134/S1063780X0811010X
30. V.G. Rudychev, V.T. Lazurik, and Y.V. Rudychev, "Influence of the electron beams incidence angles on the depth-dose distribution of the irradiated object," *Radiation Physics and Chemistry*, 186, 109527, 2021. Available: https://www.sciencedirect.com/science/article/abs/pii/S0969806X21001778
31. V.M. Lazurik, V.T. Lazurik, G.Popov, and Z. Zimek, "Two-parametric model of electron beam in computational dosimetry for radiation processing," *Radiation Physics and Chemistry*, 124, pp. 230–234, 2016. Available: https://www.sciencedirect.com/science/article/abs/pii/S0969806X1530133X
32. Theodore L. Bergman, Adrienne S. Lavine, Frank P. Incropera, and David P. DeWitt, *Fundamentals of Heat and Mass Transfer*; 8th Edition. Wiley, 2020, 992 p. Available: https://www.amazon.com/Fundamentals-Heat-Transfer-Theodore-Bergman/dp/1119722489/ref=zg_bs_14587_sccl_6/144-8112897-9282521?psc=1
33. P. Atkins, *The Laws of Thermodynamics: A Very Short Introduction*. Oxford University Press, 2010, 103 p.
34. H. Kuchling, *Taschenbuch der Physik* (in German); 21 Edition. Hanser Verlag, 2014.
35. W. Espe, *Materials of High Vacuum Technology*. Pergamon Press, 1966.
36. I. Melnyk and A. Luntovskyy, "Estimation of Energy Efficiency and Quality of Service in Cloud Realizations of Parallel Computing Algorithms for IBN," *Future Intent-Based Networking. On the QoS Robust and Energy Efficient Heterogeneous Software Defined Networks. Lecture Notes in Electrical Engineering*, vol. 831, pp. 339–379. Springer, 2022. Available: https://link.springer.com/chapter/10.1007/978-3-030-92435-5_20
37. E. Wolf, *Practical Guide to Continuous Delivery*. Addison-Wesley Professional, 2017, 288 p.
38. J.F. Epperson, *An Introduction to Numerical Methods and Analysis*; Revised Edition. Wiley-Interscience, 2007, 590 p.
39. I. Melnyk and A. Luntovskyy, "Networked Simulation with Compact Visualization of Complex Graphics and Interpolation Results," in *Klymash M., Luntovskyy A., Beshley M., Melnyk I., Schill A. (Editors) Emerging Networking in the Digital Transformation Age. TCSET 2022. Lecture Notes in Electrical Engineering*, vol. 965, pp. 175–196. Springer, Cham. Available: https://doi.org/10.1007/978-3-031-24963-1_10
40. J.H. Mathews and K.D. Fink, *Numerical Methods. Using MATLAB*. Amazon, 1998. Available: https://www.amazon.com/Numerical-Methods-Engineers-Steven-Chapra/dp/007339792X
41. *Handbook on Mathematical Functions with Formulas, Graphs and Mathematical Tables*. Edited by Milton Abramovich and Irene Stegun, National Bureau of Standards, Applied Mathmatic Series, 55, Washington, 1964, 1046 p.
42. I.N. Bronshtein, K.A. Semendyayev, G. Musiol, and H. Mühlig, *Handbook of Mathematics*; 5th Edition. Springer, 2007, 1164 p.
43. M.K. Jain, S.R.K. Iengar, and R.K. Jain, *Numerical Methods for Scientific & Engineering Computation*. New Age International Pvt. Ltd., 2010, 733 p. Available: https://www.google.com.ua/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwippcuT7rX8AhUhlYsKHRfBCG0QFnoECEsQAQ&url=https%3A%2F%2Fwww.researchgate.net%2Fprofile%2FAbiodun_Opanuga%2Fpost%2Fhow_can_solve_a_non_linear_PDE_using_numerical_method

%2Fattachment%2F59d61f7279197b807797de30%2FAS%253A284742038638596%2540144899200343%2Fdownload%2FNumerical%2BMethods.pdf&usg=AOvVaw0MjNl3K877lVWUWw-FPwmV

44. S.C. Chapra and R.P. Canale, *Numerical Methods for Engineers*; 7th Edition. McGraw Hill, 2014, 992 p.
45. K.H. Chang, *Machining Simulation Using Solidworks CAM 2021.*
46. P.J. Schilling and R.H. Shih, *Parametric Modeling with Solidworks 2021.*

Received 01.06.2023

INFORMATION ON THE ARTICLE

Igor V. Melnyk, ORCID: 0000-0003-0220-0615, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: imelnik@phbme.kpi.ua

Serhii B. Tuhai, ORCID: 0000-0001-7646-1979, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: sbtuhai@gmail.com

Dmytro V. Kovalchuk, ORCID: 0000-0001-9016-097X, the enterprise "Chervona Khvyla" Open Joint Stock Company, e-mail: dv_kovalchuk@yahoo.com

Mykola S. Surzhykov, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: nikolajsurzikov@gmail.com

Iryna S. Shved, ORCID: 0009-0008-6603-586X, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: shvettd@gmail.com

Mykhailo Yu. Skrypka, ORCID: 0009-0006-7142-5569, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: scientetik@gmail.com

Oleksandr M. Kovalenko, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: sashakovalenko51640@gmail.com

ОЦІНЮВАННЯ ТЕПЛОВОГО РЕЖИМУ РОБОТИ КАТОДА ЕЛЕКТРОННИХ ГАРМАТ ВИСОКОВОЛЬТНОГО ТЛІЮЧОГО РОЗРЯДУ, ЯКІ ФОРМУЮТЬ СТРІЧКОВИЙ ЕЛЕКТРОННИЙ ПУЧОК / І.В. Мельник, С.Б. Тугай, Д.В. Ковальчук, М.С. Суржиков, І.С. Швед, М.Ю. Скрипка, О.М. Коваленко

Анотація. Розглянуто різні способи оцінювання температури поверхні катода електронної гармати високовольтного тліючого розряду (ВТР), яка формує стрічковий електронний пучок з лінійним фокусом. Оцінювання виконано для конструкції катодного вузла гармати промислового призначення. Показано, що найбільш ефективним для проведення наближених оцінок температури поверхні катода в гарматах ВТР різного технологічного призначення є використання арифметико-логічних співвідношень для моделювання геометрії катодного вузла та локусів для оцінювання розподілу темпеsратури. Точність оцінювання, проведеного з використанням рівняння теплового балансу, складала 5–10%, що є достатнім на початковому етапі проектування електронної гармати. Показано, що використання для проектування електронних гармат ВТР програмного комплексу CAD SolidWorks є ефективним лише для вирішення завдань комплексного інженерного проектування гармат ВТР та оформлення відповідної технічної документації. Результати теоретичних досліджень є цікавими для широкого кола фахівців, які займаються розробленням електронно-променевого обладнання та його впровадженням у промислове виробництво.

Ключові слова: електронна гармата, високовольтний тліючий розряд, автоматизоване проектування, рівняння теплопровідності, рівняння теплового балансу, арифметико-логічне співвідношення, функції Рвачова.

УДК 519.711.7
DOI: 10.20535/SRIT.2308-8893.2024.1.09

КОНСТРУКЦІЇ МЕРЕЖ ПЕТРІ ІЗ СИЛЬНОЮ АНТИСИПАЦІЄЮ ЗА ПОЗИЦІЄЮ ТА ЗА ПЕРЕХОДОМ У ВИПАДКУ ДІЙСНИХ ФУНКЦІЙ

В.М. СТАТКЕВИЧ

Анотація. Запропоновано розширити класичні мережі Петрі та врахувати сильну антисипацію в сенсі Д. Дюбуа двома способами. Пропонується ввести в правило запуску переходу новий доданок, який містить дійснозначну функцію від нової кількості фішок у даній позиції (сильна антисипація за позицією) та від нової кількості фішок у вхідній позиції для даного переходу (приклад сильної антисипації за переходом). На відміну від класичних мереж Петрі умови цілочисловості вагової функції та цілочисловості маркування не накладаємо аналогічно неперервним мережам Петрі. Розглянуто виконання таких мереж, указано важливі властивості, для декількох прикладів побудовано графи досяжності та сформульовано відмінності порівняно з класичними мережами Петрі. Також досліджено умови виконання рівності маркувань для послідовностей запусків переходів $t_j t_k$ і $t_k t_j$.

Ключові слова: мережа Петрі, сильна антисипація, правило запуску переходу, граф досяжності, цілочислова функція, функція наступного стану, послідовність запусків переходів, гранична досяжність.

ВСТУП

Мережі Петрі, запропоновані К.А. Петрі в 1962 р., є зручним та потужним інструментом для проектування, аналізу та моделювання різних процесів, мереж та систем [1–3]. Нині відомо багато різних модифікацій класичних мереж Петрі, зокрема, інгібіторні, стохастичні, кольорові, неперервні, часові та інші мережі [2–5], у яких правило запуску переходу може відрізнятись від класичного правила запуску переходу.

У неперервних мереж Петрі, запропонованих у праці [4] (див. детальніше монографію [3]), конструктивна відмінність від класичних мереж полягає в тому, що кількість фішок у позиції може бути будь-яким дійсним невід'ємним числом, а модифіковане правило запуску переходу дозволяє запускати перехід нецілу кількість разів, тобто вводиться поняття ступеня запуску переходу.

Відомі також часові неперервні мережі та інші типи мереж Петрі, у яких кількість фішок у позиції також може не бути цілим числом [3]. Зазначимо, що для стохастичних мереж Петрі середня кількість фішок у позиції природним чином може не бути цілим числом [2].

Поняття антисипації, коли новий стан об'єкта залежить не тільки від попередніх станів, а також від оцінок майбутніх станів, отриманих за допомогою внутрішньої прогнозної моделі, розглядалось багатьма авторами [6].

Д. Дюбуа у 1992 р. запропонував поняття сильної антисипації, коли замість оцінок майбутніх станів використовуються саме майбутні стани [7; 8]. У працях [7–10; 12] досліджувались системи різної природи із сильною антисипацією.

Модифікація класичних мереж Петрі із сильною антисипацією за позицією була запропонована в [11]: у правило запуску переходу вводився новий доданок, який містив цілочислову функцію $f : \mathbf{N} \cup \{0\} \to \mathbf{N} \cup \{0\}$ від нової кількості фішок у позиції. Для такої функції кількість фішок у позиції залишається цілим невід'ємним числом, як і у випадку класичних мереж Петрі.

У роботі ідея врахування сильної антисипації в класичній мережі Петрі реалізується у двох напрямках. Пропонується ввести в правило запуску переходу новий доданок, який містить дійсну функцію $f : [0;+\infty) \to \mathbf{R}$ від нової кількості фішок

– у позиції (антисипація за позицією),

– у вхідній позиції для даного переходу (приклад антисипації за переходом).

Для такої функції вважається корисним відмовитись від умов цілочисловості вагової функції та цілочисловості маркування аналогічно неперервним мережам Петрі, але дозволити запускати перехід лише цілу кількість разів, як і у класичних мережах Петрі. Наскільки автору відомо, такі мережі Петрі з антисипацією не розглядались.

ПОПЕРЕДНІ ВІДОМОСТІ

Мережа Петрі — це набір $\langle P, T, W, \mu_0 \rangle$, де $P = \{p_1, \ldots, p_m\}$ — скінченна множина позицій; $T = \{t_1, \ldots, t_n\}$ — скінченна множина переходів; $W = (T \times P) \cup (P \times T) \to \mathbf{N} \cup \{0\}$ — вагова функція; $\mu_0 : P \to \mathbf{N} \cup \{0\}$ — початкове маркування. Мережу Петрі зображають у вигляді дводольного орієнтованого мультиграфу [1–3]. Перехід t називають дозволеним, якщо для кожної вхідної позиції p виконується нерівність $\mu(p) \geq W(p,t)$. Якщо перехід t дозволений, то він може (але не обов'язково має) бути запущеним, а кількість фішок у позиції p змінюється згідно з правилом запуску переходу

$$\mu_{k+1}(p) = \mu_k(p) - W(p,t) + W(t,p). \qquad (1)$$

Функцію наступного стану позначають δ [1].

МЕРЕЖІ ПЕТРІ З АНТИСИПАЦІЄЮ ЗА ПОЗИЦІЄЮ ТА ДІЙСНИМИ ФУНКЦІЯМИ

Розглянемо розширення класичної мережі Петрі

$$\langle P, T, W, \mu_0, \{f_{p_1}, \ldots, f_{p_m}\} \rangle. \qquad (2)$$

Тут $W = (T \times P) \cup (P \times T) \to [0;+\infty)$ є дійсною ваговою функцією, а $\mu_0 : P \to [0;+\infty)$ — дійсним початковим маркуванням, тобто на відміну від класичних мереж Петрі умови цілочисловості вагової функції та цілочисловості маркування не накладаємо. Функція $f_{p_i} : [0;+\infty) \to \mathbf{R}$ відповідає пози-

ції p_i, $1 \le i \le m$ (також будемо використовувати позначення f_i). Рівняння запуску переходу t (вважаємо перехід t дозволеним):

$$\mu_{new}(p) = \mu_k(p) - W(p,t) + W(t,p) + f_p(\mu_{new}(p)), \qquad (3)$$

де $\mu_{new}(p)$ — нова кількість фішок у позиції p, узагальнює правило запуску переходу (1) класичної мережі Петрі. Запропоноване розширення (2) назвемо мережею Петрі з антисипацією за позицією та дійсними функціями. Зазначимо, що випадок цілочислової вагової функції, цілочислових маркувань та цілочислових функцій f_i запропоновано у праці [11].

Розглянемо виконання таких мереж.

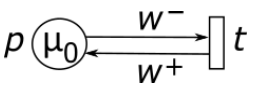


Рис. 1

Твердження 1. Нехай у мережі Петрі з антисипацією за позицією та дійсною функцією, яка зображена на рис. 1, $\mu_0 \in [0;+\infty)$, $w^- \in [0;+\infty)$, $w^+ \in [0;+\infty)$. Позиції p відповідає лінійна функція $f(x) = -ax + b$, $a > 0$, $b \in \mathbf{R}$. Тоді:

1) якщо $\mu_0 \ge w^-$ і $b + w^+ \ge (1+a)w^-$, то послідовність запусків переходів t^k є дозволеною для всіх $k \in \mathbf{N}$, відповідна послідовність маркувань $\mu_0, \mu_1, \mu_2, \ldots$ має вигляд

$$\mu_k = (1+a)^{-k}\mu_0 + (b - w^- + w^+)(1 - (1+a)^{-k})a^{-1}, \qquad (4)$$

і збігається до $\mu^* = (b - w^- + w^+)a^{-1} \ge w^-$, причому швидкість збіжності є лінійною.

2) якщо $\mu_0 < w^-$ або $b + w^+ < (1+a)w^-$, то існує таке $K \in \mathbf{N} \cup \{0\}$, що послідовність t^k є дозволеною для всіх $k \le K$, але послідовність t^{K+1} вже не є дозволеною (тобто маркування μ_K є тупиковим).

Доведення. Рівняння запуску переходу t з поточного маркування μ_k має вигляд

$$\mu_{k+1} = \mu_k - w^- + w^+ + f(\mu_{k+1}) = \mu_k - w^- + w^+ - a\mu_{k+1} + b; \qquad (5)$$

$$\mu_{k+1} = (1+a)^{-1}(\mu_k - w^- + w^+ + b) = (1+a)^{-1}\mu_k + (1+a)^{-1}(b - w^- + w^+). \qquad (6)$$

Доведемо методом математичної індукції, що $\mu_k \ge w^-$, якщо умови п. 1) виконані. База індукції: виконується нерівність $\mu_0 \ge w^-$ і перехід t є дозволеним. Припущення індукції: нехай $\mu_k \ge w^-$. Тоді перехід t є дозволеним, а з рівностей (6) випливає оцінка

$$\mu_{k+1} \ge (1+a)^{-1}(w^- - w^- + (1+a)w^-) = w^-,$$

яка доводить крок індукції. Тобто послідовність запусків переходів t^k є дозволеною для всіх $k \in \mathbf{N}$.

Доведемо методом математичної індукції рівність (4). База індукції випливає з рівностей (6), крок індукції набуває вигляду

$$\mu_{k+1} = (1+a)^{-1}\mu_k + (1+a)^{-1}(b - w^- + w^+) =$$

$$= (1+a)^{-1}[(1+a)^{-k}\mu_0 + (b - w^- + w^+)(1-(1+a)^{-k})a^{-1}] + (1+a)^{-1}(b - w^- + w^+) =$$

$$= (1+a)^{-(k+1)}\mu_0 + (b - w^- + w^+)[(1+a)^{-1} - (1+a)^{-(k+1)} + (1+a)^{-1}a]a^{-1} =$$

$$= (1+a)^{-(k+1)}\mu_0 + (b - w^- + w^+)(1-(1+a)^{-(k+1)})a^{-1}.$$

Зазначимо, що рівності (6) визначають арифметико-геометричну прогресію $u_{n+1} = qu_n + d$, тому рівність (4) також можна отримати із загальних міркувань.

Із рівності (4) граничним переходом за $k \to \infty$ отримуємо $\mu_k \to \mu^*$, якщо $k \to \infty$, а також $\mu^* \geq w^-$. Зазначимо, що збіжність $\mu_k \to \mu^*$ також можна довести за допомогою принципу стискальних відображень, не використовуючи рівність (4). Справді, оператор $T : \mathbf{R} \ni x \mapsto (1+a)^{-1}(x - w^- + w^+ + b) \in \mathbf{R}$ є стиском

$$|Tx - Ty| = \Big|(1+a)^{-1}(x - w^- + w^+ + b) -$$

$$-(1+a)^{-1}(y - w^- + w^+ + b)\Big| = (1+a)^{-1}|x - y|,$$

а тому має єдину нерухому точку x^*, яку знаходимо з рівняння (5):

$$x^* = x^* - w^- + w^+ - ax^* + b, \quad x^* = (b - w^- + w^+)a^{-1}.$$

Оцінимо швидкість збіжності послідовності $\mu_0, \mu_1, \mu_2, \ldots$ до μ^*:

$$\left|\mu_{k+1} - \mu^*\right| = \left|(1+a)^{-1}\mu_k + (1+a)^{-1}(b - w^- + w^+) - (b - w^- + w^+)a^{-1}\right| =$$

$$= \left|(1+a)^{-1}\mu_k - (b - w^- + w^+)a^{-1}(1+a)^{-1}\right| = (1+a)^{-1}\left|\mu_k - \mu^*\right| = \ldots$$

$$\ldots = (1+a)^{-(k+1)}\left|\mu_0 - \mu^*\right|$$

(згідно з принципом математичної індукції). Таким чином, швидкість збіжності є лінійною, доведення п. 1) завершено.

Для доведення п. 2) міркування такі. Якщо $\mu_0 < w^-$, то перехід t вже не є дозволеним у початковому маркуванні, тому $K = 0$. Якщо ж $\mu_0 \geq w^-$, але $b + w^+ < (1+a)w^-$ (для зручності дві наведені нерівності можна об'єднати в одну подвійну нерівність $(b - w^- + w^+)a^{-1} < w^- \leq \mu_0$), то для певного K виконується оцінка $\mu_K < w^-$. Справді, рівність (4) виконується в даному разі для всіх $k \leq K$, тому

$$\mu_K = (1+a)^{-K}\mu_0 + (b - w^- + w^+)(1-(1+a)^{-K})a^{-1} < w^-,$$

$$(1+a)^{-K}(\mu_0 - (b - w^- + w^+)a^{-1}) < w^- - (b - w^- + w^+)a^{-1},$$

$$(1+a)^{-K} < \frac{w^- - (b - w^- + w^+)a^{-1}}{\mu_0 - (b - w^- + w^+)a^{-1}},$$

$$K > \log_{1+a} \frac{\mu_0 - (b - w^- + w^+)a^{-1}}{w^- - (b - w^- + w^+)a^{-1}},$$

$$K = \left[\log_{1+a} \frac{a\mu_0 - b + w^- - w^+}{aw^- - b + w^- - w^+} \right] + 1$$

(тут $[x]$ позначає цілу частину числа x). Отже, перехід t стає недозволеним, маркування μ_K стає тупиковим, а послідовність маркувань $\mu_0, \mu_1, \mu_2, \ldots, \mu_K$ стає скінченною. Твердження 1 доведено.

Із твердження 1 випливають такі наслідки:

1) існують такі w^-, w^+, a, b та початкове маркування μ_0, що μ^* може дорівнювати довільному наперед заданому невід'ємному числу;

2) для структури $\langle P, T, W \rangle$ класичної мережі Петрі показаного на рис. 1 вигляду, тобто для довільних ваг w^- і w^+ існує така функція $f(x) = -ax + b$ і таке початкове маркування μ_0, що μ^* може дорівнювати довільному наперед заданому невід'ємному числу;

3) згідно з принципом стискальних відображень $(\mu_0 = \mu^*) \Rightarrow$ $\Rightarrow (\forall k \in \mathbf{N} : \mu_k = \mu^*)$;

4) рівність (4) свідчить про те, що точка μ_k ділить відрізок $[\mu_0; \mu^*]$ або $[\mu^*; \mu_0]$ у заданому відношенні: $\mu_k = (1+a)^{-k}\mu_0 + (1-(1+a)^{-k})\mu^*$.

Нагадаємо, що у класичних мережах Петрі послідовність кількостей фішок у певній позиції має скінченну границю тоді і тільки тоді, коли ця послідовність стабілізується, починаючи з деякого номера (також послідовність кількостей фішок у певній позиції може мати нескінченну границю $\omega = +\infty$). Але з введенням антисипації за позицією з дійсною функцією послідовність μ_k може мати довільну скінченну границю μ^* і при цьому в загальному випадку $\mu_k \neq \mu^*$. Така властивість є характерною саме для неперервних мереж Петрі і називається граничною досяжністю або lim-досяжністю [3, с. 135–136; 5].

Зазначимо, що випадок $a = 0$ відповідає класичній мережі Петрі без антисипації, якщо змінити вагу дуги $W(t, p) = b + w^+$ (за умови $b \geq -w^+$). У випадку $-1 < a < 0$ оператор T не є стиском, рівність (4) зберігається, але $\mu_k \to \infty$, якщо $k \to \infty$, і можна вводити символ $\omega = +\infty$. Випадок $a < -1$ не розглядаємо, оскільки значення μ_{k+1} згідно з рівнянням (5) може стати від'ємним. Випадок $a = -1$ розглянемо окремо у прикладі 1.

Приклад 1. Розглянемо мережу Петрі з антисипацією за позицією та дійсною функцією, зображену на рис. 1, де $\mu_0 = 4$, $w^- = 1$, $w^+ = 0$, $a = -1$, $b = -3$. Тоді рівняння запуску переходу t з початкового маркування μ_0 має вигляд $\mu_1 = 4 - 1 + \mu_1 - 3$ згідно з рівнянням (5), звідки $\mu_1 \geq 0$. Таким чином, отримуємо нове маркування $\mu_1 = [0; +\infty)$, яке взагалі не є маркуванням у класичному сенсі та яке не є зліченною множиною. Звернемо увагу, що за

антисипації за позицією та цілочислових функцій $f:\mathbf{N}\cup\{0\}\to\mathbf{N}\cup\{0\}$ множина $\delta(\mu,t)$ зліченна і незліченною бути не може [11, с. 106].

Рівняння запуску переходу t з маркування μ_1 має вигляд $\mu_2=\mu_1-1+\mu_2-3$, воно має розв'язки $\mu_2\geq 0$, якщо $\mu_1=4$, і не має розв'язків, якщо $\mu_1\in[1;+\infty)\setminus\{4\}$ (перехід є дозволеним, якщо $\mu_1\geq 1$). Тому нове маркування, яке збігається з уже отриманим раніше "маркуванням" $\mu_1=[0;+\infty)$, отримане не класичним запуском переходу t, а умовним за умови $\mu_1=4$.

Отже, будуємо граф досяжності (рис. 2). Для даної мережі граф є скінченним, не є деревом і містить єдине класичне маркування μ_0 та маркування μ_1, яке не з'являється у класичних мережах Петрі. Граф не містить тупикових маркувань. Також у графі існують класичний запуск переходу t та умовний, причому умовні переходи характерні саме для неперервних мереж Петрі [3, с. 116–119].

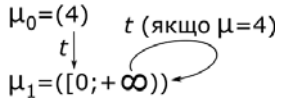


Рис. 2

Зауваження 1. У монографії [3, с. 115–116] зазначається, що у неперервних мережах Петрі кількість маркувань може бути нескінченною, а тому пропонується оригінальний механізм макромаркувань, кількість яких скінченна.

Розглянемо довільну мережу Петрі з антисипацією за позицією, у якій позиціям p_i відповідають лінійні функції $f_i(x)=-a_ix+b_i$, $1\leq i\leq m$. Нехай $R(\mu)$ — множина маркувань, досяжних з маркування μ, $\bullet t$ і $t\bullet$ — множини вхідних та вихідних позицій переходу t відповідно [2, 3].

Твердження 2. Якщо $a_i>-1$, $1\leq i\leq m$, то для кожного маркування $\mu\in R(\mu_0)$ та для кожного переходу t множина $\delta(\mu,t)$ містить не більше одного елемента.

Доведення випливає з твердження 1, оскільки рівняння вигляду (5) не може мати двох або більше розв'язків.

Теорема 1. Нехай виконуються такі умови:

1) $a_i>-1$, $1\leq i\leq m$;

2) для деяких j і k обидві послідовності запусків переходів t_jt_k і t_kt_j є дозволеними у поточному маркуванні $\mu\in R(\mu_0)$, а обидві множини $\delta(\mu,t_jt_k)=\{\mu_{jk}\}\neq\varnothing$ і $\delta(\mu,t_kt_j)=\{\mu_{kj}\}\neq\varnothing$ є непорожніми;

3) $p_i\in\bullet t_j$ або $p_i\in t_j\bullet$, $p_i\in\bullet t_k$ або $p_i\in t_k\bullet$.

Тоді виконується еквівалентність

$$(\mu_{jk}(p_i)=\mu_{kj}(p_i))\Leftrightarrow(a_i=0 \text{ або } -W(p_i,t_j)+W(t_j,p_i)=$$

$$=-W(p_i,t_k)+W(t_j,p_k)).$$

Доведення. Позначимо $W(p_i,t_j)=w_{ij}^-$, $W(t_j,p_i)=w_{ij}^+$, $W(p_i,t_k)=w_{ik}^-$, $W(t_k,p_i)=w_{ik}^+$. Двічі застосовуємо рівності (6) і отримуємо:

$$\delta(\mu, t_j t_k) = (1+a_i)^{-1}((1+a_i)^{-1}(\mu - w_{ij}^- + w_{ij}^+ + b_i) - w_{ik}^- + w_{ik}^+ + b_i);$$

$$\delta(\mu, t_k t_j) = (1+a_i)^{-1}((1+a_i)^{-1}(\mu - w_{ik}^- + w_{ik}^+ + b_i) - w_{ij}^- + w_{ij}^+ + b_i).$$

Тоді

$$(\mu_{jk}(p_i) = \mu_{kj}(p_i)) \Leftrightarrow ((1+a_i)^{-2}(-w_{ij}^- + w_{ij}^+) + (1+a_i)^{-1}(-w_{ik}^- + w_{ik}^+) =$$

$$= (1+a_i)^{-2}(-w_{ik}^- + w_{ik}^+) + (1+a_i)^{-1}(-w_{ij}^- + w_{ij}^+)) \Leftrightarrow$$

$$\Leftrightarrow (-w_{ij}^- + w_{ij}^+ + (1+a_i)(-w_{ik}^- + w_{ik}^+) = -w_{ik}^- + w_{ik}^+ + (1+a_i)(-w_{ij}^- + w_{ij}^+)) \Leftrightarrow$$

$$\Leftrightarrow (a_i(-w_{ik}^- + w_{ik}^+) = a_i(-w_{ij}^- + w_{ij}^+)).$$

Рівність $-w_{ik}^- + w_{ik}^+ = -w_{ij}^- + w_{ij}^+$ означає рівність i-х елементів k-го та j-го рядків матриці A у класичному рівнянні стану мережі Петрі $A^T x = \Delta\mu$ (нагадаємо, що згідно з [2] рядкам матриці A відповідають переходи, стовпцям — позиції, вектор x є лічильником запусків переходів). У класичній мережі Петрі без антисипації рівність $\mu_{jk}(p_i) = \mu_{kj}(p_i)$ виконується завжди, якщо обидві послідовності запусків переходів $t_j t_k$ і $t_k t_j$ є дозволеними; це співвідноситься з рівностями $a_i = b_i = 0$ для всіх $1 \le i \le m$. З уведенням антисипації за позицією множини $\delta(\mu, t_j t_k)$ і $\delta(\mu, t_k t_j)$, $\mu \in R(\mu_0)$ у загальному випадку стають різними, більш того жодна з цих множин не обов'язково у загальному випадку має бути підмножиною іншої [11]. Теорема 1 важлива тим, що надає достатні умови рівності $\delta(\mu, t_j t_k) = \delta(\mu, t_k t_j)$ у випадку лінійних функцій.

Випадок $a_i = -1$ для деякого i розглянемо окремо в прикладі 2.

Приклад 2. Розглянемо мережу Петрі з антисипацією за позицією та дійсними функціями, зображену на рис. 3, де $w_{11}^- \in [0;+\infty)$, $w_{11}^+ \in [0;+\infty)$, $w_{12}^- \in [0;+\infty)$, $w_{12}^+ \in [0;+\infty)$, $\mu_0 \in [\max(w_{11}^-, w_{12}^-);+\infty)$, $a = -1$, $b \in \mathbf{R}$. Використаємо міркування прикладу 1.

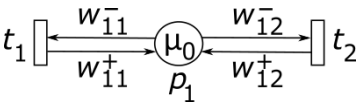


Рис. 3

Рівняння запуску переходу t_1 з початкового маркування μ_0 має вигляд $\mu_1 = \mu_0 - w_{11}^- + w_{11}^+ + \mu_1 + b$, звідки $\mu_1 \ge 0$, якщо виконується рівність $\mu_0 = w_{11}^- - w_{11}^+ - b$. Таким чином отримуємо нове маркування $\mu_1 = [0;+\infty)$. Рівняння запуску переходу t_2 з маркування μ_1 має вигляд $\mu_2 = \mu_1 - w_{12}^- + w_{12}^+ + \mu_2 + b$, воно має розв'язки $\mu_2 \ge 0$, якщо $\mu_1 = w_{12}^- - w_{12}^+ - b$, і не має розв'язків, якщо $\mu_1 \ne w_{12}^- - w_{12}^+ - b$ (перехід є дозволеним за $\mu_1 \ge w_{12}^-$). Тому нове маркування, яке збігається з уже отриманим раніше маркуванням $\mu_1 = [0;+\infty)$, отримане не класичним запуском переходу t_2, а умовним за умови $\mu_1 = w_{12}^- - w_{12}^+ - b$.

Аналогічним чином, розглядаючи послідовність запусків переходів t_2t_1, отримуємо:

1) якщо виконуються обидві рівності $\mu_0 = w_{11}^- - w_{11}^+ - b$ і $\mu_0 = w_{12}^- - w_{12}^+ - b$, то $\delta(\mu, t_1t_2) = \delta(\mu, t_2t_1) = [0;+\infty)$;

2) якщо не виконується перша (відповідно, друга) рівність, то перехід t_1 (відповідно, t_2) формально є дозволеним у класичному сенсі, але $\delta(\mu, t_1) = \varnothing$ (відповідно, $\delta(\mu, t_2) = \varnothing$).

Зауваження 2. Ситуацію, коли у мережі Петрі з антисипацією за позицією певний перехід t може бути дозволений у класичному сенсі, але рівняння його запуску не має розв'язків і $\delta(\mu, t) = \varnothing$, виявлено у праці [11].

ПРИКЛАДИ МЕРЕЖ ПЕТРІ З АНТИСИПАЦІЄЮ ЗА ПЕРЕХОДОМ ТА ДІЙСНИМИ ФУНКЦІЯМИ

Розглянемо класичну мережу Петрі $\langle P, T, W, \mu_0 \rangle$, кожен перехід якої має не більше однієї вхідної позиції $|\bullet t| \le 1$, і введемо інше розширення:

$$\langle P, T, W, \mu_0, \{f_{t_1}, \ldots, f_{t_n}\} \rangle. \tag{7}$$

Як і раніше, $W = (T \times P) \cup (P \times T) \to [0;+\infty)$ — дійсна вагова функція, $\mu_0 : P \to [0;+\infty)$ — дійсне початкове маркування, тобто умови цілочисловості також не накладаємо. Проте на відміну від попереднього розширення (2) функція $f_{t_i} : [0;+\infty) \to \mathbf{R}$ відповідає переходу t_i, $1 \le i \le n$, а не позиції p_i (також будемо використовувати позначення f_i, якщо це не буде призводити до конфлікту позначень). Якщо перехід t є дозволеним, то нова кількість фішок $\mu_{new}(p)$ у позиції $p \in \bullet t$ задовольняє рівняння запуску переходу

$$\mu_{new}(p) = \mu_k(p) - W(p,t) + W(t,p) + f_t(\mu_{new}(p)), \tag{8}$$

а кількість фішок у позиції $p \in t \bullet$ змінюється згідно з класичним правилом запуску переходу (1). Якщо ж позиція p є одночасно і вхідною, і вихідною позицією для переходу t, то також використовуємо рівняння (8). Рівняння (8) узагальнює класичне правило запуску переходу (1), але відрізняється від рівняння (3) — тому розширення (7) відрізняється від розширення (2). Структура мережі $\langle P, T, W \rangle$ є суттєвою — факт $|\bullet t| \le 1$ ураховується у рівнянні (8). Запропоноване розширення (7) ілюструє приклад антисипації за переходом та дійсними функціями.

Приклад 3. Розглянемо мережу Петрі з антисипацією за переходом та дійснозначними функціями, зображену на рис. 4. Нехай переходам t_1 і t_2 відповідають лінійні функції $f_i : [0;+\infty) \to \mathbf{R}$ $(i = 1,2)$, $f_1(x) = x$, $f_2(x) = 2x - 7$. Якщо перехід t_i є дозволеним, то нова кількість фішок $\mu_{new}(p_1)$

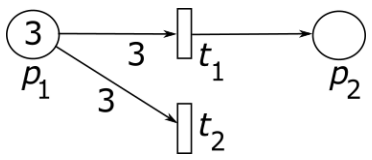


Рис. 4

у вхідній позиції p_1 задовольняє рівняння запуску переходу (8), окрім того, запуск переходу t_1 додає одну фішку у вихідну позицію p_2.

Під час запуску дозволеного переходу t_1 з початкового маркування $\mu_0 = (3,0)$ рівняння (8) набуває вигляду $\mu_{new}(p_1) = 3 - 3 + f_1(\mu_{new}(p_1))$, звідки $\mu_{new}(p_1) = \mu_{new}(p_1)$, $\mu_{new}(p_1) \geq 0$. Таким чином, отримуємо нове маркування $\mu_1 = ([0;+\infty),1)$, яке, як і у прикладі 1, не є класичним маркуванням і яке не є зліченною множиною. Із запуском дозволеного переходу t_2 з маркування μ_0 маємо $\mu_{new}(p_1) = 3 - 3 + f_2(\mu_{new}(p_1))$, звідки $\mu_{new}(p_1) = 2\mu_{new}(p_1) - 7$. Таким чином, отримуємо нове маркування $\mu_2 = (7,0)$.

Під час запуску переходу t_1 з маркування μ_1 маємо $\mu_{new}(p_1) = \mu_1(p_1) - 3 + \mu_{new}(p_1)$. Дане рівняння має розв'язки $\mu_{new}(p_1) \geq 0$ за умови $\mu_1(p_1) = 3$ і не має розв'язків, якщо $\mu_1(p_1) > 3$ (t_1 не є дозволеним у випадку $\mu_1(p_1) < 3$). Таким чином, нове маркування $\mu_3 = ([0;+\infty),2)$ отримано не класичним запуском переходу t_1, а умовним за умови $\mu_1(p_1) = 3$. За запуску переходу t_2 з маркування μ_1 маємо рівняння $\mu_{new}(p_1) = \mu_1(p_1) - 3 + 2\mu_{new}(p_1) - 7$, звідки $\mu_{new}(p_1) = 10 - \mu_1(p_1)$. Ураховуючи обмеження $\mu_{new}(p_1) \geq 0$, отримуємо нове маркування $\mu_4 = ([0;7],1)$ умовним запуском переходу t_2 за умови $3 \leq \mu_1(p_1) \leq 10$ (t_2 не є дозволеним, якщо $\mu_1(p_1) < 3$).

Під час запуску з маркування μ_2 переходу t_1, який є дозволеним, отримуємо рівняння $\mu_{new}(p_1) = 7 - 3 + \mu_{new}(p_1)$, яке не має розв'язків. Аналогічно зауваженню 2 знайдена ситуація, коли перехід формально є дозволеним, але $\delta(\mu_2, t_1) = \varnothing$. За запуску з маркування μ_2 переходу t_2, який також є дозволеним, маємо $\mu_{new}(p_1) = 7 - 3 + 2\mu_{new}(p_1) - 7$, звідки $\mu_{new}(p_1) = 3$. Таким чином, отримуємо маркування $(3,0)$, яке збігається з початковим. Зазначимо, що для класичних мереж Петрі послідовність маркувань $\mu_0 = (3,0)$, $\mu_2 = \delta(\mu_0, t_2) = (7,0)$, $\delta(\mu_2, t_2) = \mu_0 = (3,0)$ отримати не можна: у такому випадку $\delta(\mu_2, t_2)$ має дорівнювати $(11,0)$.

Нарешті, після запуску з маркування μ_4 переходу t_1 маємо рівняння $\mu_{new}(p_1) = \mu_4(p_1) - 3 + \mu_{new}(p_1)$, звідки отримуємо маркування μ_3 умовним запуском за умови $\mu_4(p_1) = 3$, а після запуску з маркування μ_4 переходу t_2 маємо рівняння $\mu_{new}(p_1) = \mu_4(p_1) - 3 + 2\mu_{new}(p_1) - 7$ і далі $\mu_{new}(p_1) = 10 - \mu_4(p_1)$, звідки отримуємо нове маркування $\mu_5 = ([3;7],1)$ умовним запуском за умови $\mu_4(p_1) \geq 3$. Із маркування μ_5 отримуємо вже існуюче маркування μ_3 умовним запуском переходу t_1 за умови $\mu_5(p_1) = 3$, отримуємо те саме маркування μ_5 класичним запуском переходу t_2.

Отже, можемо побудувати граф досяжності, зображений на рис. 5. Для спрощення запису умовних переходів на рис. 5 використаємо позначення $\mu_k^1 = \mu_k(p_1)$, $\mu_k^2 = \mu_k(p_2)$ для $k \in \mathbf{N}$. Для даної мережі граф нескінченний ($\mu_{3k} = ([0;+\infty),k+1)$, $\mu_{3k+1} = ([0;7],k)$, $\mu_{3k+2} = ([3;7],k)$ для $k \in \mathbf{N}$), не є де-

ревом і не містить жодного тупикового маркування, хоча без уведення антисипації за переходом класична мережа Петрі на рис. 4 мала б два тупикові маркування $(0,0)$ і $(0,1)$. Існують тільки два класичні маркування μ_0 і μ_2, інші маркування в класичних мережах Петрі не з'являються. Також у графі існують як класичні запуски переходів t_1, t_2, так і умовні.

Твердження 1 для мережі на рис. 1 в точності переноситься для випадку з антисипацією за переходом та дійсними функціями, оскільки t має в точності одну вхідну позицію. Наведемо аналог теореми 1 для часткового випадку мережі, зображеної на рис. 3.

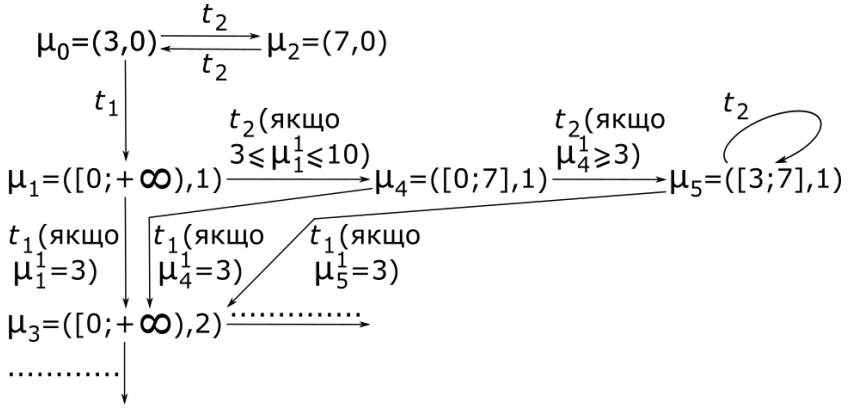


Рис. 5

Твердження 3. Нехай виконуються такі умови:

1) $\mu_0 \in [0;+\infty)$, $w_{11}^- \in [0;+\infty)$, $w_{11}^+ \in [0;+\infty)$, $w_{12}^- \in [0;+\infty)$, $w_{12}^+ \in [0;+\infty)$;

2) переходам t_i відповідають функції $f_i(x) = -a_i x + b_i$, $a_i > -1$, $b_i \in \mathbf{R}$ $(i = 1,2)$;

3) обидві послідовності запусків переходів t_1t_2 і t_2t_1 є дозволеними у поточному маркуванні $\mu \in R(\mu_0)$, а обидві множини $\delta(\mu, t_1t_2) = \{\mu_{12}\} \neq \varnothing$ і $\delta(\mu, t_2t_1) = \{\mu_{21}\} \neq \varnothing$ є непорожніми.

Тоді виконується еквівалентність

$$(\mu_{12} = \mu_{21}) \Leftrightarrow \left(\begin{vmatrix} a_1 & -w_{11}^- + w_{11}^+ + b_1 \\ a_2 & -w_{12}^- + w_{12}^+ + b_2 \end{vmatrix} = 0 \right).$$

Доведення. За запуску переходу t_1 з поточного маркування μ маємо рівняння

$$\mu_1 = \mu - w_{11}^- + w_{11}^+ + f_1(\mu_1) = \mu - w_{11}^- + w_{11}^+ - a_1\mu_1 + b_1; \qquad (9)$$

$$\mu_1 = (1 + a_1)^{-1}(\mu - w_{11}^- + w_{11}^+ + b_1).$$

Після запуску переходу t_2 з маркування μ_1 маємо рівняння

$$\mu_{12} = \mu_1 - w_{12}^- + w_{12}^+ + f_2(\mu_{12}) = \mu_1 - w_{12}^- + w_{12}^+ - a_2\mu_{12} + b_2; \qquad (10)$$

$$\mu_{12} = (1 + a_2)^{-1}((1 + a_1)^{-1}(\mu - w_{11}^- + w_{11}^+ + b_1) - w_{12}^- + w_{12}^+ + b_2).$$

Аналогичним чином отримуємо

$$\mu_{21} = (1+a_1)^{-1}((1+a_2)^{-1}(\mu - w_{12}^- + w_{12}^+ + b_2) - w_{11}^- + w_{11}^+ + b_1).$$

Тоді

$$(\mu_{12} = \mu_{21}) \Leftrightarrow$$

$$\Leftrightarrow ((1+a_1)^{-1}(1+a_2)^{-1}(-w_{11}^- + w_{11}^+ + b_1) + (1+a_2)^{-1}(-w_{12}^- + w_{12}^+ + b_2) =$$

$$= (1+a_1)^{-1}(1+a_2)^{-1}(-w_{12}^- + w_{12}^+ + b_2) + (1+a_1)^{-1}(-w_{11}^- + w_{11}^+ + b_1)) \Leftrightarrow$$

$$\Leftrightarrow (-w_{11}^- + w_{11}^+ + b_1 + (1+a_1)(-w_{12}^- + w_{12}^+ + b_2) =$$

$$= -w_{12}^- + w_{12}^+ + b_2 + (1+a_2)(-w_{11}^- + w_{11}^+ + b_1)) \Leftrightarrow$$

$$\Leftrightarrow (a_1(-w_{12}^- + w_{12}^+ + b_2) = a_2(-w_{11}^- + w_{11}^+ + b_1)) \Leftrightarrow \left(\begin{vmatrix} a_1 & -w_{11}^- + w_{11}^+ + b_1 \\ a_2 & -w_{12}^- + w_{12}^+ + b_2 \end{vmatrix} = 0 \right).$$

Зазначимо, що в класичній мережі Петрі без антисипації, яка зображена на рис. 3, рівність $\mu_{12} = \mu_{21}$ виконується завжди, якщо обидві послідовності запусків переходів t_1t_2 і t_2t_1 є дозволеними, це співвідноситься з рівностями $a_i = b_i = 0$ для всіх $1 \le i \le n$. Твердження 3 важливе тим, що надає достатні умови рівності $\delta(\mu, t_1t_2) = \delta(\mu, t_2t_1)$ у випадку лінійних функцій.

Також відслідкуємо умови, за яких обидві послідовності запусків переходів t_1t_2 і t_2t_1 є дозволеними у твердженні 3:

1) якщо $\mu \ge w_{11}^-$, перехід t_1 є дозволеним (див. рівність (9));

2) якщо $\mu_1 = (1+a_1)^{-1}(\mu - w_{11}^- + w_{11}^+ + b_1) \ge w_{12}^-$, послідовність переходів t_1t_2 є дозволеною (див. рівність (10));

3) отже, з урахуванням двох аналогічних нерівностей достатньо вимагати

$$\mu \ge \max(w_{11}^-; w_{12}^-; w_{11}^- + w_{12}^- - w_{11}^+ - b_1 + a_1 w_{12}^-; w_{11}^- + w_{12}^- - w_{12}^+ - b_2 + a_2 w_{11}^-).$$

ВИСНОВКИ

Запропоновано дві модифікації класичних мереж Петрі: із сильною антисипацією за позицією та дійсними функціями та із сильною антисипацією за переходом та дійсними функціями. Розглянуто виконання таких мереж, указано важливі властивості, а також досліджено умови виконання рівності маркувань для послідовностей запусків переходів t_jt_k і t_kt_j.

ЛІТЕРАТУРА

1. J. Peterson, *Theory of Petri nets and system modeling*. M.: Mir, 1984, 264 p.
2. T. Murata, “Petri nets: Properties, analysis, applications,” *TIIER*, vol. 77, no. 4, pp. 41–85, 1989.
3. R. David and H. Alla, *Discrete, Continuous, and Hybrid Petri Nets*. Springer, Berlin, Heidelberg, 2005.
4. R. David and H. Alla, “Continuous Petri Nets”, *8th European Workshop on Application and Theory of Petri Nets, Zaragoza, Spain*, pp. 275–294, 1987.

5. C.R. Vazquez, C. Mahulea, J. Julvez, and M. Silva, “Introduction to Fluid Petri nets”, Chapter in Book: “Control of Discrete-Event Systems”, *Lecture Notes in Control and Information Sciences*, vol. 433, Eds. C. Seatzu, M. Silva, J.H. van Schuppen, Springer-Verlag, London, pp. 365–386, 2013. doi: 10.1007/978-1-4471-4276-8_18.
6. R. Rosen, *Anticipatory Systems: Philosophical, Mathematical and Methodological Foundations*. Pergamon Press, 1985. doi: 10.1016/C2009-0-07769-1.
7. D. Dubois, “Incursive and hyperincursive systems, fractal machine and anticipatory logic,” *Computing Anticipatory Systems: CASYS 2000 — Fourth International Conference. AIP Conference Proceedings*, vol. 573, pp. 437–451, 2001. doi: 10.1063/1.1388710.
8. D. Dubois, “Generation of fractals from incursive automata, digital diffusion and wave equation systems,” *Biosystems*, vol. 43, pp. 97–114, 1997. doi: 10.1016/S0303-2647(97)01692-4.
9. A. Makarenko, “Multivaluedness Aspects in Self-Organization, Complexity and Computations Investigations by Strong Anticipation”, *Chapter in Book: Recent Advances in Nonlinear Dynamics and Synchronization*; Eds. K. Kyamakya, W. Mathis, R. Stoop, J. Chedjou, Z. Li, Springer, Cham, pp. 33–54, 2018. doi: 10.1007/978-3-319-58996-1_3.
10. A. Makarenko, “Toward Multivaluedness Aspects in Self-Organization, Complexity and Computations Investigations,” *Forth International Workshop on Nonlinear Dynamics and Synchronization INDS’15, Klagenfurt, Austria, Alpen-Adria University, July 31, 2015*, pp. 84–93.
11. V.M. Statkevych, “A modification of Petri nets with anticipation on a position,” *System Research & Information Technologies*, no. 1, pp. 102–112, 2023. doi: 10.20535/SRIT.2308-8893.2023.1.08.
12. S.V. Lazarenko, O.S. Makarenko, *Discrete systems with anticipation*. National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, 2020.

Надійшла 01.07.2023

INFORMATION ON THE ARTICLE

Vitalii M. Statkevych, ORCID: 0000-0001-5210-9890, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: mstatckevich@yahoo.com

DESIGNING PETRI NETS WITH STRONG PLACE AND TRANSITION ANTICIPATION FOR REAL-VALUED FUNCTIONS / V.M. Statkevych

Abstract. We propose extending the classic Petri nets and considering D. Dubois’s strong anticipation in two ways. We propose to add a new term into a transition rule that contains a real-valued function of a new marking in a certain place (strong place anticipation) or of a new marking in the input place of a certain transition (an example of strong transition anticipation). Any integer constraints are not applied either to the weight function or to the marking in contrast to the classic Petri nets (as in continuous Petri nets). The execution of the mentioned nets is investigated, and important properties are stated. Several examples of reachability graphs are given, and differences from classic Petri nets are formulated. We also investigate the conditions of the equality of the markings, which are obtained by firing the sequences of transitions $t_j t_k$ and $t_k t_j$.

Keywords: Petri net, strong anticipation, transition rule, reachability graph, real-valued function, next-state function, sequence of transition firings, limit reachability.

UDC 62-50
DOI: 10.20535/SRIT.2308-8893.2024.1.10

INTELLIGENT OPTIMAL CONTROL OF NONLINEAR DIABETIC POPULATION DYNAMICS SYSTEM USING A GENETIC ALGORITHM

EL OUISSARI ABDELLATIF, EL MOUTAOUAKIL KARIM

Abstract. Diabetes is a chronic disease affecting millions of people worldwide. Several studies have been carried out to control the diabetes problem, involving both linear and non-linear models. However, the complexity of linear models makes it impossible to describe the diabetic population dynamic in depth. To capture more detail about this dynamic, non-linear terms were introduced into the mathematical models, resulting in more complicated models strongly consistent with reality (capable of re-producing observable data). The most commonly used methods for control estimation are Pantryagain's maximum principle and Gumel's numerical method. However, these methods lead to a costly strategy regarding material and human resources; in addition, diabetologists cannot use the formulas implemented by the proposed controls. In this paper, the authors propose a straightforward and well-performing strategy based on non-linear models and genetic algorithms (GA) that consists of three steps: 1) discretization of the considered non-linear model using classical numerical methods (trapezoidal rule and Euler–Cauchy algorithm); 2) estimation of the optimal control, in several points, based on GA with appropriate fitness function and suitable genetic operators (mutation, crossover, and selection); 3) construction of the optimal control using an interpolation model (splines). The results show that the use of the GA for non-linear models was successfully solved, resulting in a control approach that shows a significant decrease in the number of diabetes cases and diabetics with complications. Remarkably, this result is achieved using less than 70% of available resources.

Keywords: optimal control, differential equation, diabetes, genetic algorithms, artificial intelligence, intelligent local search.

INTRODUCTION

Diabetes is a major public health problem. Diabetes is a major public health problem and one of the most dangerous common diseases and is characterized by high blood sugar [2]. Diabetes is the root of many diseases and costs many lives. It is a serious chronic disease that occurs when the body does not properly use the insulin it produces or when the pancreas does not produce enough insulin. There are three types of diabetes: Type 1 diabetes, Type 2 diabetes and Gestational Diabetes Mellitus (GDM) [3].

The number of people with diabetes has increased exponentially in recent times. According to the World Health Organization (WHO) and the Interna- tional Diabetes Federation (IDF) [1; 4], 463 million people had diabetes in 2019 and this number is expected to reach 578 million in 2030 and 700 million in 2045 [11; 9].

The mathematical modeling of the phenomenon of diabetes is the subject of several researchers in different mathematical fields. These include ordinary differential equations (ODE) that study, for example, the diabetic population as found in [5–8], there are also research works that study the phenomenon of diabetes by modeling has partial differential equations (PDE). Likewise, studies on this

phenomenon are carried out using delay differential equations (DDE), and others, for example stochastic differential equations (SDE) and integro-differential equations (IDE), Fredholm integral equations (FIE) [15–18].

In [18] the authors have given a thorough review of the delay differential equation models and they are presented with some computational results and brief summaries of the theoretical results for the cases of the insulin ultradian oscillation models and the models for the diagnostic tests.

The modeling of natural phenomena is a very successive task to deal with any phenomenon [40]. In the first modeling of the phenomenon of diabetes, two types of population were considered, pre-diabetic and diabetic. The authors of [5] modeled this natural phenomenon as a system of linear ordinary differential equations of the first degree. Then, in 2007 [10], a modification was made to the above system. This time, another type of population was considered, namely diabetic patients with complications.

Several studies were performed for this model, they show that the system is well defined, also the stability and determination of equilibrium points was done. Then in the year 2014 [6], the authors is thought to try to reduce the neg- ative effect of the phenomenon of diabetes, for this they proposed an approach of optimal control that help to minimize as much as possible the spread of the phenomenon, focusing on the study of the diabetic population. Three types of diabetic patients are considered, patients who become diabetic for different reasons, which may be genetic or related to a negative lifestyle. The other two types are diabetic patients without complications and diabetic patients with complications. The modeling of the mathematical model is well explained in [6], and the existence and uniqueness of solution is well shown, also the existence of control. In the ten years, the studies carried out with this control have shown their effectiveness in reducing the number of diabetic patients, which shows the success of this strategy of searching for an optimal control to control this phenomenon. A dynamic 6-compartment control system was proposed for the study of the diabetes phenomenon in [11], they divided the population in general as follows: healthy people H, pre-diabetic patients due to genetics which is denoted by P, pre-diabetic patients due to lifestyle denoted by E. The other three types are di- abetic patients without complications and diabetic patients with complications, which are denoted by D and C respectively. In [11], they paid attention to the fact that there are several types of consciousness that are directly related to diabetes, there are genetic influences and bad lifestyle, on the other hand there is the psychology of the person. For this reason, they proposed optimal controls that take these influences into account.

In [12], another reformulation of dynamic model of diabetic population was proposed. They resulted that the factors most related to diabetes or to the fact that a person becomes diabetic are the genetic factors and a bad lifestyle. To protect diabetic patients, an awareness program was proposed using media and education; psychological follow-up was also considered and medical treatment, this time they grouped patients according to their age through a continuous dynamic system. They divided the general population into 4 different compartments given in the following form: pre-diabetic patients due to genetics, pre-diabetic patients due to lifestyle, diabetic patients with complication and without complication.

In this work we propose a very sample and performance strategy based on non-linear models and genetic algorithm (GA)s that processes into three steps: 1) discretization of the considered non-linear model using classical numerical methods (trapezoidal rule and Euler–Cauchy algorithm); 2) estimation of the

optimal control, in several points, based on GA with appropriate fitness function and suitable genetic operators (mutation, crossover, and selection); 3) construction of the optimal control using interpolation model.

This paper is organized as follows: The second section provides an introduction on the importance of using genetic algorithms to solve optimal control problems. The third section presents two dynamic control systems, one is the kernel linear system and the other is the nonlinear system proposed in recent years, we have given the theoretical summary that has been done on both models. The fourth section presents the minimization of the objective functions of each control system of the diabetic population by the genetic algorithm and we compare the results found by the results of the classical method. Finally, we end the paper with a general conclusion.

GENETIC ALGORITHM

In [9] Boutayeb et al., in recent years artificial intelligence is invading all fields, it has also become an essential part of human life. For example, in this work, we will use the artificial intelligence methods to treat the optimal control problems associated with a diabetic population [29; 14].

In this article, we deal with a very complicated phenomenon, because several effects come into play, for example genetic influence, age, lifestyle and others.

Genetic algorithms are artificial intelligence methods and are heuristic search techniques that are very simple to handle [30; 31]. Genetic algorithms (GAs) are search optimization algorithms based on three essential operators, selection, crossover and mutation [21]. In general, GAs were first developed by Holland and are derived from Darwin's theory of evolution [20]. In the first step, a population of initial solutions, called chromosomes, is randomly selected, then these solutions are evaluated by the objective function or fitness function, if the solution has a very high performance then this is the solution we are looking for, otherwise, we move to the crossover and mutation step which generates a new solution from the initial solution, likewise this solution generated and evaluated by the objective function, and so on until we find the best solution. Genetic algorithms are very effective in complex optimization problems more than simple methods, which sometimes fail to achieve a certain problem due to its complexity [32].

They are used in many research areas and are very useful in real world applications, because they are very simple and give the best solutions. Classical optimization methods, which are purely computational methods, start with a single initial solution and then search for the optimal solution, but the genetic algorithm starts with an initial population of candidates and then searches for the best optimal solution in the search space [28].

Optimal control problems are also among the optimization problems that have been solved by the GA. The real beginning of the use of GA for dynamic control systems is in the years 1992 by Krishnakumar and Goldberg [22] who gave a start to GA in a very important discipline of applied mathematics [23; 26; 27]. The main objective of our work is to achieve a significant reduction in the number of diabetics without complications and with treatable complications, which involves the search for optimal control to achieve this goal. The dynamic PEDC model is a non-linear mathematical model representing the evolution of the diabetic population and the influence of uncomplicated diabetic patients on pre-diabetics. The advantage of using the GA to solve this type of optimal control

problem is that we do not need to determine the adjoint values or characterize the control in any specific way. We only need to treat the problem well and specify the objective function we need to minimize and its various constraints.

THE MODELS

In this section, we will focus on different dynamic and controlled systems for diabetic patients found in the literature that minimize the negative impact of this phenomenon. First, we will start with the dynamically controlled Kernel system, which allows the study of diabetic patients without entering their bodies.

The EDC model

In [13] Boutayeb et al. have proposed an optimal control approach modeling the progression from pre-diabetes to diabetes with and without complications. Three types of diabetic patients are considered, patients who become diabetic for different reasons, which may be genetic or related to a negative lifestyle, this type is noted by E. The other two types are diabetic patients without complications and diabetic patients with complications, which are denoted by D and C respectively. We conclude that the diabetic population N is divided into three types such that $N = N(t) = E(t) + C(t) + D(t)$ at time t:

$$\begin{cases} \dfrac{dE(t)}{dt} = I - (\mu + \beta_3 + \beta_1)E(t); \\ \dfrac{dD(t)}{dt} = \beta_1 E(t) - (\mu + \beta_2)D(t) + \gamma C(t); \\ \dfrac{dC(t)}{dt} = \beta_3 E(t) + \beta_2 D(t) - (\mu + \gamma + \nu + \delta)C(t). \end{cases}$$

The modeling of the model is explained in [13], the model contains eight parameters: $I, \mu, \beta_1, \beta_2, \beta_3, \gamma, \nu$ and δ, which were estimated in [10]. The control of this phenomenon is the subject of several research works, due to the fact that this phenomenon has social burdens that should be reduced. For this purpose the authors proposed to control the transition rate from diabetes with complications to diabetes without complications.

The controlled model is given by the following system:

$$\begin{cases} \dfrac{dE(t)}{dt} = I - (\mu + (\beta_3 + \beta_1)(1 - u(t)))E(t); \\ \dfrac{dD(t)}{dt} = \beta_1(1 - u(t))E(t) - (\mu + \beta_2(1 - u(t)))D(t) + \gamma C(t); \\ \dfrac{dC(t)}{dt} = \beta_3(1 - u(t))E(t) + \beta_2(1 - u(t))D(t) - (\mu + \gamma + \nu + \delta)C(t), \end{cases} \quad (1)$$

where u is a control. The objective function we are trying to minimize is given by:

$$J(u) = \int_0^T (D(t) + C(t) + Au^2(t))dt,$$

where A is a positive weight that balances the size of the terms. U is the control set defined by:

$$U = \{u / \text{uismeasurable}, \ 0 \le u(t) \le 1, \ t \in [0, T]\}.$$

The goal is to find a control u^* in U that minimizes the objective function:

$$J(u^*) = \min_{u \in U} J(u).$$

The system (1) is well defined, moreover theorems 3.2, 3.3, 3.4 and 3.5 in [13] have shown respectively the existence and uniqueness and the positivity of the solution, the existence of an optimal control and the characterization of this control for this system (1).

Characterisation of the control optimal. The principle of maximum Pontryagin and Hamiltonian has been used to characterize the optimal control, and the adjoint variables. After mathematical calculations and demonstrations that can be found in [13], we find that the adjoint variables are defined by:

$$\begin{cases} \lambda_1' = (\lambda_1 - \lambda_2)(1-u^*)\beta_1 + (\lambda_1 - \lambda_3)\beta_3 + \mu\lambda_1; \\ \lambda_2' = -1 + (\lambda_2 - \lambda_3)(1-u^*)\beta_2 + \mu\lambda_2; \\ \lambda_3' = -1 + (\lambda_3 - \lambda_2)\gamma + \lambda_3(\mu + \nu + \delta) \end{cases}$$

or λ_1, λ_2 and λ_3 are the adjoint variables With conditions: $\lambda_1(T)=\lambda_2(T)= =\lambda_3(T)=0$ and E^*, D^* and C^* are the solutions of system (1) and the control is defined as :

$$u^* = \min\left(1, \max\left(0, \frac{1}{2A}[E^*\beta_1(\lambda_2 - \lambda_1) + E^*\beta_3(\lambda_3 - \lambda_1) + D^*\beta_2(\lambda_3 - \lambda_2)]\right)\right).$$

This study, which was conducted over a period of ten years, summarizes that the optimal control approach is very effective in reducing the effect of this phenomenon, which is demonstrated in the experimental results found.

The $HP\,EDC_SC_T$ model

An improvement of the EDC model was proposed in 2020 by Kouidere et al. [11] noted $HPEDC_TC_S$. This time by a nonlinear controlled dynamic system. Six types of populations are considered, people who are healthy H, pre-diabetic patients due to genetics which is denoted by P , pre-diabetic patients due to lifestyle denoted by E. The other three types are diabetic patients without complications and diabetic patients with complications, which are denoted by D and C respectively. We conclude that the diabetic population N is divided into six types such that $N = N(t) = H(t) + P(t) + E(t) + D(t) + C_S(t) + C_T(t)$ at time t:

$$\begin{cases} \dfrac{dH(t)}{dt} = I - (\mu + \theta_1 + \theta_2)H(t); \\ \dfrac{dP(t)}{dt} = \theta_1 H(t) - (\mu + \beta_1 + \beta_3)P(t); \\ \dfrac{dE(t)}{dt} = \theta_2 H(t) - (\mu + \gamma)E(t); \\ \dfrac{dD(t)}{dt} = \beta_1 P(t) + \gamma E(t) - \alpha_1 \dfrac{D(t)E(t)}{N} - (\mu + \beta_2 + \eta_2)D(t); \\ \dfrac{dC_T(t)}{dt} = \beta_3 P(t) + \beta_2 D(t) + \alpha_1 \dfrac{D(t)E(t)}{N} - \alpha_2 \dfrac{C_T(t)E(t)}{N} - (\mu + \eta_1)C_T(t); \\ \dfrac{dC_S(t)}{dt} = \eta_2 D(t) + \eta_1 C_T(t) + \alpha_2 \dfrac{C_T(t)E(t)}{N} - (\mu + \delta_1)C_S(t); \end{cases} \quad (2)$$

with $H(0) \geq 0,\ P(0) \geq 0,\ E(0) \geq 0,\ D(0) \geq 0,\ C_T(0) \geq 0$, and $C_S(0) \geq 0$.

The modeling of the model is explained in [11], the model contains twelve parameters: $I, \mu, \theta_1, \theta_2, \eta_1, \eta_2, , \alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3, \gamma, \nu$ and δ which were estimated in [11]. The objective remains to control the transition rate from prediabetes to diabetes with complications and transition from diabetes with complications to diabetes without complications.

The controlled model is given by the following system:

$$\begin{cases} \dfrac{dH(t)}{dt} = I - (\mu + \theta_1 + \theta_2)H(t); \\ \dfrac{dP(t)}{dt} = \theta_1 H(t) - (\mu + \beta_1 + \beta_3)P(t); \\ \dfrac{dE(t)}{dt} = \theta_2 H(t) - (\mu + (1 - u_4(t))\gamma)E(t); \\ \dfrac{dD(t)}{dt} = \beta_1 P(t) + (1 - u_4(t))\gamma E(t) - (1 - u_2(t))\alpha_1 \dfrac{D(t)E(t)}{N} - \\ \qquad\qquad - (\mu + \beta_2 + \eta_2)D(t) + u_1(t)C_T(t); \\ \dfrac{dC_T(t)}{dt} = \beta_3 P(t) + \beta_2 D(t) + (1 - u_2(t))\alpha_1 \dfrac{D(t)E(t)}{N} - \\ \qquad\qquad - (\mu + \eta_1 + u_1(t))\alpha_2 \dfrac{C_T(t)E(t)}{N} - (\mu + \eta_1 + u_1(t))C_T(t); \\ \dfrac{dC_S(t)}{dt} = \eta_2 D(t) + \eta_1 C_T(t) + (1 - u_3(t))\alpha_2 \dfrac{C_T(t)E(t)}{N} - (\mu + \delta_1)C_S(t), \end{cases}$$

where the controls $u_1(t),\ u_2(t), u_3(t)$, and C are the proposed controls.

The objective function we are trying to minimize is given by :

$$J(u_1, u_2, u_3, u_4) = C_T(T_f) - D(T_f) - E(T_f) +$$

$$+ \int_0^{T_f} \left[C_T(t) - D(t) - E(t) + \frac{A}{2}u_1^2(t) + \frac{B}{2}u_2^2(t) + \frac{F}{2}u_3^2(t) + \frac{G}{2}u_4^2(t) \right] dt,$$

where $A > 0,\ B > 0,\ F > 0$, and $G > 0$.

The system (4) is well defined, moreover theorems 1, 2, 3, 4, and 5 in [11] have shown respectively the existence and uniqueness and the positivity of the solution, the existence of an optimal control and the characterization of this control for this system (4):

$$\begin{cases} \dfrac{(u_1, u_2, u_3, u_4)}{0} \leq u_{1min} \leq u_1(t) \leq u_{1max} \leq 1; \\ 0 \leq u_{2min} \leq u_2(t) \leq u_{2max} \leq 1; \\ 0 \leq u_{3min} \leq u_3(t) \leq u_{3max} \leq 1; \\ 0 \leq u_{4min} \leq u_4(t) \leq u_{4max} \leq 1 \ \ / t \in [0, T_f]. \end{cases} \tag{3}$$

Characterisation of the control optimal. The principle of maximum Pontryagin and Hamiltonian has been used to characterize the optimal control, and the adjoint variables. After mathematical calculations and demonstrations that can be found in [11], we find that the adjoint variables are defined by:

$$\lambda_1' = \lambda_1(\mu + \theta_1 + \theta_2) - \lambda_2\theta_1 - \lambda_3\theta_2,$$

$$\lambda_2' = \lambda_2(\mu + \beta_1 + \beta_3) - \lambda_4\beta_1 - \lambda_5\beta_3,$$

$$\lambda_3' = 1 + \lambda_3(\mu + \gamma(1-u_4(t))) - \lambda_4\left[\gamma(1-u_4(t)) - \alpha_1(1-u(t))\frac{D(t)}{N}\right] -$$

$$-\lambda_5\left[\alpha_1(1-u_2(t))\frac{D(t)}{N} - \alpha_2(1-u_3(t))\frac{C_T(t)}{N}\right] - \lambda_6\left[\alpha_2(1-u_3(t))\frac{C_T(t)}{N}\right],$$

$$\lambda_4' = 1 - \lambda_4\left[\left(\alpha_1(1-u_2(t))\frac{E(t)}{N} + (\mu + \beta_2 + \eta_2)\right)\right] -$$

$$-\lambda_5\left[\beta_2 + \alpha_1(1-u_2(t))\frac{E(t)}{N}\right] - \lambda_6\eta_2, \qquad (4)$$

$$\lambda_5' = -1 - \lambda_4 u_1(t) + \lambda_5\left[-\alpha_2(1-u_3(t))\frac{E(t)}{N} - (\mu + \eta_1 + u_1(t))\right] -$$

$$-\lambda_6\left[\eta_1 + \alpha_2(1-u_3(t))\frac{E(t)}{N}\right],$$

$$\lambda_6' = \lambda_6(\mu + \delta).$$

With the transversality conditions at time T_f, $\lambda_1(T_f) = 0$, $\lambda_2(T_f) = 0$, $\lambda_3(T_f) = 1$, $\lambda_4(T_f) = 1$, $\lambda_5(T_f) = -1$ and $\lambda_6(T_f) = 0$.

Furthermore, for $t \in [0, T_f]$, the optimal controls u_1^*, u_2^*, u_3^* and u_4^* are given by:

$$u_1^* = \min\left(1, \max\left(0, \frac{(\lambda_5 - \lambda_4)}{A}C_T(t)\right)\right),$$

$$u_2^* = \min\left(1, \max\left(0, \alpha_1 \times \frac{\lambda_5 - \lambda_4}{B} \times \frac{D(t)E(t)}{N}\right)\right),$$

$$u_3^* = \min\left(1, \max\left(0, \alpha_2 \times \frac{\lambda_6 - \lambda_5}{F} \times \frac{C_T(t)E(t)}{N}\right)\right),$$

$$u_4^* = \min\left(1, \max\left(0, \frac{(\lambda_4 - \lambda_3)}{G} \times \gamma E(t)\right)\right).$$

EXPERIMENTAL RESULTS

In this section, we represent a comparison between the classical numerical method used to solve the mathematical models proposed in the literature and the intelligent genetic algorithm method. In the first part we will start with the kernel model (1) and in the second part we will make the comparison in the nonlinear model (4).

In this research paper, we used the genetic algorithm to adjust the different controls and the different parameters of the system, knowing that it is a powerful evolutionary algorithm.

The advantage of using the genetic algorithm to solve this type of optimal control problem is that we do not need to determine the adjoint values or characterize the control in any specific way. We only need to treat the problem well and specify the objective function we need to minimize and its various constraints. We present the results obtained by solving numerically the optimality systems (1) and (4) using the genetic algorithm method. In this control problems, we used the (GA) to find the most accurate initial conditions for the state variables.

In both models, which we will study, the quality of the controls of each of them, we have used the same values of the parameters that were used in the work [11; 13]. MatLab was used to write and compile the code, and the following data was used:

Genetic algorithm configuration:

Crossover : multiple;

Crossover : 0.8;

Initialization : random;

Number of iteration: $100 \times dim$;

Mutation : gaussian;

Population size : 200;

Selection function : stochastic(uniform).

The choice of the parameters of the genetic algorithm operators (selection, crossover, mutation) is a difficult task. In our work, our configuration is based on the following information:

a) Crossover: this step allows the production of new solutions, from the previous individuals. Each bit is chosen from either parent with the same probability [35; 36].

b) Mutation: it is sometimes found that all solutions are weak, which means the crossover operator does not lead to a new gene. The mutation causes random changes of rate pm in the genes, but this rate should not be larger, to avoid loss of the principle of selection and evolution [37].

c) Population size: the size of the population has a direct influence on the optimal control capability of our problem. Our choice of this parameter is based on [38; 39].

Linear EDC model

Let's start with the linear model (1). The use of single-objective optimization methods is not obvious on all this type of problem, first of all the understanding of the phenomenon is essential to put the pseudo-code simple to execute. Secondly, for a very good choice of the initial population that an important step in the method of genetic algorithm.

The Boutayeb model (1) parameters: $I = 2000000; \mu = 0.02; \gamma = 0.08; \nu = 0.05;$ $\delta = 0.05; \beta_1 = 0.5; \beta_2 = 0.1; \beta_3 = 0.5;$ $dim = 0 : .1 : 10; A = 3550000.$

The Boutayeb model (2) compartiments initialization: $E(0) = 66.6 \times 105;$ $D(0) = 102 \times 105;$ $C(0) = 55 \times 105.$

As the linear case [33; 34], we used least square linear method to estimate the values of the parameters of the dynamic system of population diabetic and we found values close to those of the research paper [13; 34], which was based on the values of the World Diabetes Federation [1; 4].

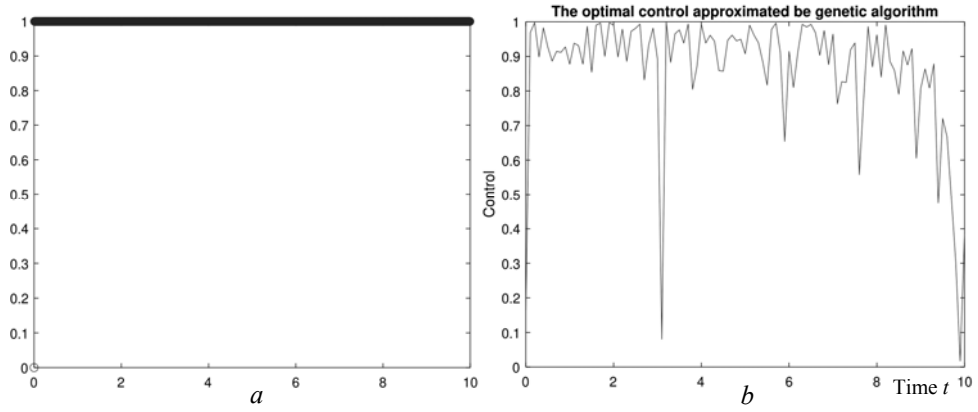


Fig. 1. The comparison between the control given by the GA and Gumel methods: *a* — the control *u* given by system (1) using Gumel method; *b* — the control *u* using GA

From the Fig. 1, *a*, we can notice that the use of the Gauss–Seidel type implicit finite difference method developed by Gumel [19] is give an estimate and does not give an optimal control u^*.

For that we propose to solve this problem with the genitic algorithm (GA), we can see in the Fig. 2, *a*, *b*, *c* and Fig. 1, *b* that GA gives an optimal control but does not give a good effect.

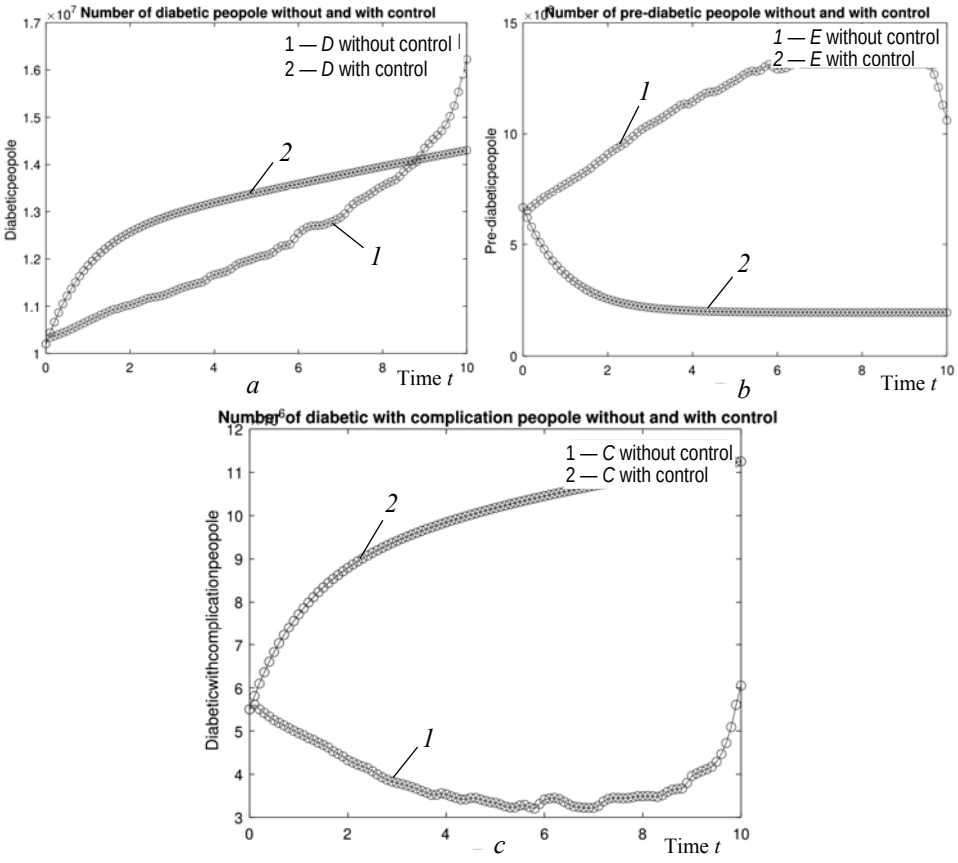


Fig. 2. The compartment *E*, *D* and *C* with and without control: *a* — the compartment *D* with and without control using GA; *b* — the compartment *E* with and without control using GA; *c* — the compartment *C* with complications withand without control using GA

It can be noticed from Fig. 1, *a* that the control given by Gumel's method takes almost a fixed value which is exactly 1. Of course, this allows a com- plete control of the dynamics of the phenomenon, throughout the five years, but consumes the totality of human and material resources. The proposed solution, based on a nonlinear model and a genetic algorithm, produces reasonable controls with very low cost (in fact, we never need all the resources in a given period to achieve our goal) (see Fig. 2). Furthermore, the genetic algorithm shows a great ability to produce controls in only a few iterations.

A nonlinear $HP\,EDC_SC_T$ model

We now turn to the non-linear model (4) which contains 6 compartments. We have seen that four compartments have been controlled by four different controls. Our objective is to learn about the controls given by the intillegent method and their quality.

The model 4 shows that the number of pre-diabetic population by the negative effect of behavioral factors on diabetic patients and other diabetics E and D without complications are those who control blood glucose by diet and exercise before it is too late, C_T diabetics with treatable complication and C_S diabetics with severe complications is decreased by the use of controls u_1, u_2, u_3 and u_4 respectively. This can be seen clearly in the Figs. 3, *a*, *b*, 4, *a*, *b*, 5, *a*, *b*.

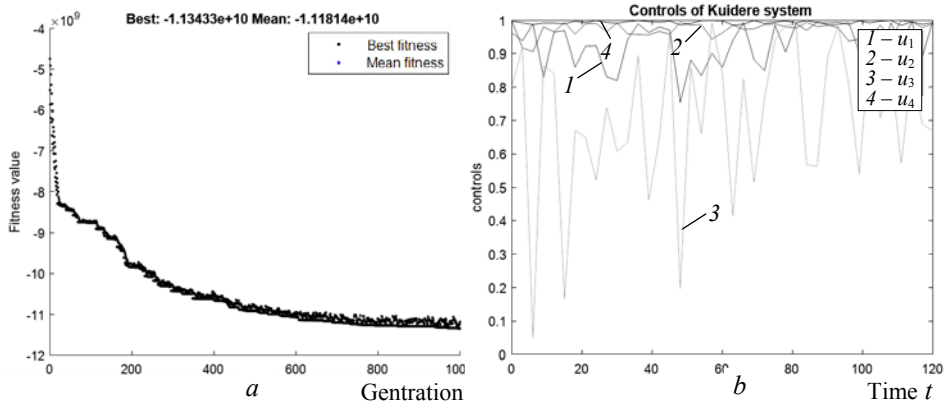


Fig. 3. The controls of non-linear system: *a* — objective function of GA; *b* — the controls of non-linear system

In genetic algorithm, there exist several types of convergences: a) the constructed control sequence becomes very close to the optimal control sought; b) the constructed control sequence becomes stagnant; c) the cost sequence of the constructed control sequence becomes stagnant; d) the maximum number of iterations, set by the user, is exhausted. It is impossible to refer to the optimal control sought because it is not known; therefore the first type of convergence remains impractical. The second and third type of convergence is directly influenced by the size of the population, the crossover rate, and the mutation rate; indeed, a bad choice of these can lead the genetic search to very bad local minima. The fourth type of convergence is called artificial convergence because the user is satisfied with a reasonable improvement of the quality of the initial population especially when the studied phenome is very complex and calls for several algorithmic components. We opted for the artificial convergence because the studied system is very complex and the control estimation is part of a very large research project involving several components. Moreover, we can notice

that a few iterations were able to achieve a very good estimation; of course, we can increase the number of iterations to improve the obtained control even if the one obtained is very satisfactory (see Fig. 3, *a*).

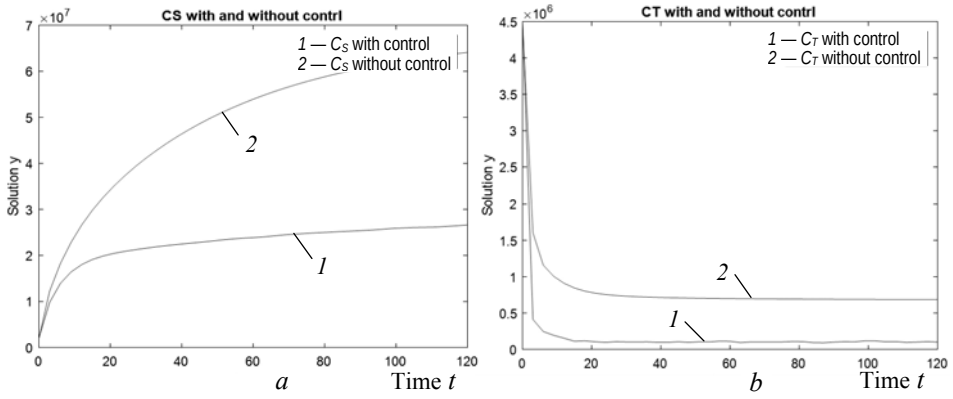


Fig. 4. The compartment (*a*) C_T and (*b*) C_S with and without control given by system (7) and (10) using GA

In this work, we present two types of models. A linear dynamic model and a non-linear model. In the first model, we considered a single control on both behaviors, which does not correspond to human nature, because each diabetic patient has his own needs. But the second non-linear model takes into account 6 different behaviors, that is, the most possible cases of diabetes, moreover four controls were considered, which helps us to treat each behavior with its needs. In conclusion, the dynamic controlled model $HPEDC_TC_S$ gives us more infor- mation because of the diversity of diabetes types, and the control strategy for each type.

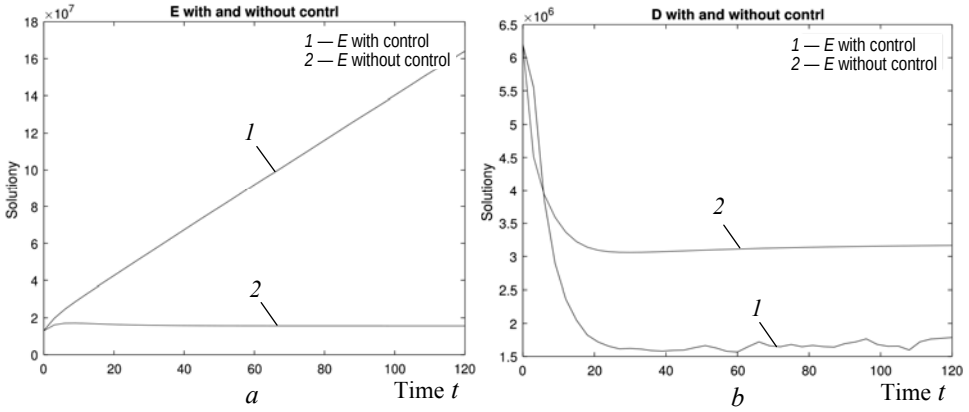


Fig. 5. The compartment E (*a*) and D (*b*) with and without control given by systems (2) and (3) using GA

In general, we introduced a comparison between the following systems: a) Gumel and linear model; b) Gumel and non-linear model; and c) Genetic Algorithm and non-linear model.

Solved by Gumel method, the linear and non-linear models, the produced control recommends always the implementation of all disponible resources because of the formula of the control that implements the operators max-min which leads to the value 1. Solved by genetic algorithm, the linear and non-linear models, the produced control is capable to reduce significantly the number of diabetic and diabetics with complications using less than 70% (mean) of resources (see

Figs. 1 and 3). In addition, the non-linear system is more suitable to understand and control the phenomenon under study.

The estimated control are in the hand of diabetologists and it will be traduced in terms of medication, exercise, and diet.

ACKNOWLEDGMENT

This work was supported by Ministry of National Education, Professional Training, Higher Education and Scientific Research (MENFPESRS) and the Digital Development Agency (DDA) and CNRST of Morocco (Nos. Alkhawarizmi/2020/23).

CONCLUSION

In this paper we have elaborated a state of the art on the different dynamic systems with economic function proposed in the literature controlling diabetes in order to alleviate the socio-economic damage caused by it. Then, we have used of the best performing local search artificial intelligence methods (metaheuristics) to solve the kernel model dealing with the common compartments between all these models. Controlling diabetes is a major challenge, and studies have highlighted the lim- itations of linear models in describing the complexity of this chronic disease. Non-linear models, by introducing more realistic terms, have enabled a better understanding of the dynamics of the diabetic population. However, traditional methods of order estimation, such as Pantryagain's maximum principle and Gumel's numerical method, have proved costly and difficult for diabetologist professionals to implement. In this paper, we have proposed an innovative strategy based on nonlinear models and the use of GAs. This three-step approach discretizes the nonlinear model, estimates the optimal control using GAs, and constructs the final control using an interpolation model. The results obtained demonstrated the success of this approach, with a significant reduction in the number of cases of diabetes and complications in diabetic patients, while using less than 70% of available resources. Despite the success of GAs in diabetes control, there are a number of limitations to consider. Firstly, model resolution time can be a challenge, not least due to the complexity of the non-linear models used. In addition, the use of integer derivatives may restrict access to certain relevant information. In future work, it is proposed to explore two avenues to overcome these limitations, 1) the use of parallelism techniques could speed up model solution time; 2) the use of fractional equations rather than ordinary differential equations could be considered.

REFERENCES

1. "Definition and diagnosis of diabetes mellitus and inter mediate hyperglycaemia," *World Health Organisation (WHO), Geneva, 2016*.
2. A. El Ouissari, K. El Moutaouakil,"Density based fuzzy support vector machine: application to diabetes dataset," *Mathematical Modeling and Computing*, vol. 8, no. 4, pp. 747–760, 2021.
3. K. El Moutaouakil, M. Cheggour, S. Chellak, and H. Baizri, "Meta- heuristics Optimization Algorithm to an Optimal Moroccan Diet," in *2021 7th*

Annual International Conference on Network and Information Systems for Computers (ICNISC), IEEE, 2021, pp. 364–368.
4. *IDF DIABETES ATLAS;* 9th edition. International Diabetes Federation (IDF), 2019.
5. A. Boutayeb, E.H. Twizell, K. Achouayb, and A. Chetouani, "A mathematical model for the burden of diabetes and its complications," *Biomedical engineering online*, 3(1), pp. 1–8, 2004.
6. A. Boutayeb, A. Chetouani, K. Achouayb, and E.H. Twizell, "A non- linear population model of diabetes mellitus," *Journal of Applied Mathematics and Computing*, 21(1), pp. 127–139, 2006.
7. A. Nakonechnyj, V.Marzenyuk, "Problems of controllability for differential Gompertzian dynamics equations," *Cybernetics and Systems Analysis*, 40(2), pp. 123–133, 2004.
8. K. Marti, Y. Ermoliev, M. Makowski, G. Pflug, A.G. Nakonechny, and V.P. Marzeniuk, "Uncertainties in medical processes control," in *Coping with Uncertainty: Modeling and Policy Issues*, pp. 185–192. Springer Berlin Heidelberg, 2006.
9. A. Kouidere, O. Balatif, H. Ferjouchia, A. Boutayeb, and M. Rachik, "Optimal control strategy for a discrete time to the dynamics of a population of diabetics with highlighting the impact of living environment," *Discrete Dynamics in Nature and Society*, 2019.
10. A. Boutayeb, A. Chetouani, "A population model of diabetes and pre-diabetes," *International Journal of Computer Mathematics*, 84(1), pp. 57–66, 2007.
11. A. Kouidere, A. Labzai, H. Ferjouchia, O. Balatif, and M. Rachik, "A New Mathematical Modeling with Optimal Control Strategy for the Dynamics of Population of Diabetics and Its Complications with Effect of Behavioral Factors," *Journal of Applied Mathematics*, 2020.
12. A. Kouidere, B. Khajji, O. Balatif, and M. Rachik, "A multi-age math-ematical modeling of the dynamics of population diabetics with effect of lifestyle using optimal control," *Journal of Applied Mathematics and Computing*, vol. 67, pp. 375–403, 2021.
13. M. Derouich, A. Boutayeb, W. Boutayeb, and M. Lamlili, "Optimal control approach to the dynamics of a population of diabetics," *Applied Mathematical Sciences*, 8(56), pp. 2773–2782, 2014.
14. K. El Moutaouakil, M. Roudani, and A. El Ouissari, "Optimal Entropy Genetic Fuzzy-C-Means SMOTE (OEGFCM-SMOTE)," *Knowledge-Based Systems*, 262, 110235, 2023.
15. A. Mahata, S.P. Mondal, S. Alam, A. Chakraborty, S.K. Dey, and A. Goswami, "Mathematical model for diabetes in fuzzy environment with stability analysis," *Journal of Intelligent and Fuzzy Systems*, 36(3), pp. 2923–2932, 2019.
16. R.L. Ollerton, "Application of optimal control theory to diabetes mellitus," *International Journal of Control*, 50(6), pp. 2503–2522, 1989.
17. G.W. Swan, "An optimal control model of diabetes mellitus," *Bulletin of Mathematical Biology*, 44(6), pp. 793–808, 1982.
18. A. Makroglou, I. Karaoustas, J. Li, and Y. Kuang, "Delay differential equation models in diabetes modelling," *Theoretical Biology and Medical Modelling*, 2009.
19. A.B. Gumel, P.N. Shivakumar, and B.M. Sahai, "A mathematical model for the dynamics of HIV-1 during the typical course of infection," in *Proceedings of the 3rdWorld Congress of Nonlinear Analysts*, 47 (2011), pp. 2073–2083.
20. F. Hayes-Roth, Review of "Adaptation in Natural and Artificial Systems by John H. Holland", The U. of Michigan Press, 1975, *ACM SIGART Bulletin*, issue 53, pp. 15, 1975.
21. L.B. Booker, D.E. Goldberg, and J.H. Holland, "Classifier systems and genetic algorithms," *Artificial Intelligence*, 40(1-3), pp. 235–282, 1989.
22. K. Krishnakumar, D.E. Goldberg, "Control system optimization using genetic algorithms," *Journal of Guidance, Control, and Dynamics*, 15(3), pp. 735–740, 1992.

23. Z. Michalewicz, C.Z. Janikow, and J.B. Krawczyk, „A modified genetic algo- rithm for optimal control problems," *Computers & Mathematics with Applications*, 23(12), pp. 83–94, 1992.
24. N. Marco, C. Godart, J.A. Desideri, B. Mantel, and J. Periaux, *A genetic algorithm compared with a gradient-based method for the solution of an active- control model problem (Doctoral dissertation, INRIA*), 1996
25. I.C. Lerman, R. Ngouenet, *Algorithmes genetiques sequentiels et paralleles pour une representation affine des proximites (Doctoral dissertation, INRIA),* 1995.
26. L. Rarita, I. Stamova, and S. Tomasiello, "Numerical schemes and genetic algorithms for the optimal control of a continuous model of supply chains," *Applied Mathematics and Computation*, 388, 125464, 2021.
27. D. Castaldo, M. Rosa, and S. Corni, "Quantum optimal control with quantum computers: A hybrid algorithm featuring machine learning optimization," *Physical Review A*, 103(2), 022613, 2021.
28. J. Wang, J. Hou, J. Chen, Q. Fu, and G. Huang, "Data mining approach for improving the optimal control of HVAC systems: An event-driven strategy," *Journal of Building Engineering*, 39, 102246, 2021
29. H.H. Mehne, S. Mirjalili, "A parallel numerical method for solving optimal control problems based on whale optimization algorithm," *Knowledge-Based Systems*, 151, pp. 114–123, 2018.
30. A.K. John, K. Krishnakumar, "Performing multiobjective optimization on perforated plate matrix heat exchanger surfaces using genetic algorithm," *International Journal for Simulation and Multidisciplinary Design Optimization*, 8, A3, 2017.
31. I.M. Fanuel, A. Mushi, and D. Kajunguri, "Irrigation water allocation optimization using multi-objective evolutionary algorithm (MOEA) a review," *International Journal for Simulation and Multidisciplinary Design Optimization*, 9, A3, 2018.
32. N.R. Nagaiah, C.D. Geiger, "Application of evolutionary algorithms to optimize cooling channels," *International Journal for Simulation and Multidisciplinary Design Optimization*, 10, A4, 2019.
33. K. El Moutaouakil, A. El Ouissari, B. Hicham, C. Saliha, and M. Cheggour, "Multi-objectives optimization and convolution fuzzy C-means: control of diabetic population dynamic," *RAIRO-Operations Research*, 56(5), pp. 3245–3256, 2022. doi: 10.1051/ro/2022142.
34. Abdellatif el Ouissari, Karim el Moutaouakil, Baizri Hicham, and Chellak Saliha, "Intelligent Local Search for an Optimal Control of Diabetic Population Dynamics," *Mathematical Models and Computer Simulations*, 14, pp. 1051–1071, 2022. doi: 10.1134/S2070048222060047.
35. Darrell Whitley, "A genetic algorithm tutorial," *Statistics and computing*, vol. 4, no. 2, pp. 65–85, 1994,
36. David E. Goldberg, *Genetic Algorithm in Search, Optimization and Machine Learning*. New Jersey, Addison-Wesley, 1989.
37. A.E. Eiben, J.E. Smith, *Introduction to evolutionary computing*. Berlin: Springer, 2003.
38. Masaaki Horie, Naoki Fukuta, "An Evolutionary Approach for Simulating Contin- uous Coalition Formation", *2018 7th International Congress on Advanced Applied Informatics (IIAI-AAI)*, pp. 588–593.
39. B.R. Rajakumar, A. George, "APOGA: An adaptive population pool size based genetic algorithm," *AASRI Procedia*, 4, pp. 288–296, 2013.
40. V.P. Martsenyuk, I.Y. Andrushchak, and A.M. Kuchvara, "On conditions of asymptotic stability in SIR-models of mathematical epidemiology," *Journal of Automation and Information Sciences*, 43(12), 2011.

Received 11.11.2022

INFORMATION ON THE ARTICLE

Abdellatif El Ouissari, ORCID: 0000-0003-0956-2029, Sidi Mohamed Ben Abdellah University; Higher School of Engineering in Applied Sciences, Fez, Morocco, e-mail: abdellatif.elouissari@usmba.ac.ma

Karim El Moutaouakil, ORCID: 0000-0003-3922-5592, Sidi Mohamed Ben Abdellah University, Fez, Morocco, e-mail: karim.elmoutaouakil@usmba.ac.ma

ІНТЕЛЕКТУАЛЬНЕ ОПТИМАЛЬНЕ КЕРУВАННЯ НЕЛІНІЙНОЮ СИСТЕМОЮ ПОПУЛЯЦІЙНОЇ ДИНАМІКИ ХВОРИХ НА ДІАБЕТ ІЗ ВИКОРИСТАННЯМ ГЕНЕТИЧНОГО АЛГОРИТМУ / Абделлатиф Ель Уіссарі, Карім Ель Мутауакіль

Анотація. Цукровий діабет є хронічним захворюванням, яким страждають мільйони людей у всьому світі. Виконано кілька досліджень, спрямованих на контроль проблеми діабету, із використанням як лінійних, так і нелінійних моделей. Однак складність лінійних моделей не в змозі глибинно описати динаміку діабетичного населення. Щоб отримати більше деталей про цю динаміку, до математичних моделей уведено нелінійні терміни, що призвело до більш складних моделей, які повністю відповідають реальності (здатні відтворювати спостережувані дані). Найбільш часто використовуваними методами для оцінювання контролю є принцип максимуму Пантрягейна та числовий метод Гумеля. Однак ці методи призводять до дуже дорогої стратегії з точки зору матеріальних і людських ресурсів; крім того, діабетологи не в змозі використовувати формули, реалізовані запропонованими елементами контролю. Запропоновано вибіркову стратегію та продуктивність, засновану на нелінійних моделях і генетичних алгоритмах (GA), яка виконується в три етапи: 1) дискретизація розглянутої нелінійної моделі за допомогою класичних числових методів (правило трапеції та алгоритм Ейлера–Коші); 2) оцінювання оптимального контролю в кількох точках на основі GA з відповідною функцією пристосованості та відповідними генетичними операторами (мутація, схрещування та відбір); 3) побудова оптимального керування за допомогою інтерполяційної моделі (сплайнів). Отримані результати показали, що використання GA для нелінійних моделей було успішно вирішено, що призвело до контрольного підходу, який демонструє значне зменшення кількості випадків діабету та діабетиків з ускладненнями. Примітно, що цей результат досягається з використанням менше ніж 70% доступних ресурсів.

Ключові слова: оптимальне керування, диференціальне рівняння, діабет, генетичні алгоритми, штучний інтелект, інтелектуальний локальний пошук.

ВІДОМОСТІ ПРО АВТОРІВ

Бабенко Віталій Олегович,
аспірант кафедри біомедичної кібернетики КПІ ім. Ігоря Сікорського, Україна, Київ

Бідюк Петро Іванович,
професор, доктор технічних наук, професор кафедри математичних методів системного аналізу ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Бодянський Євгеній Володимирович,
професор, доктор технічних наук, професор кафедри штучного інтелекту Харківського національного університету радіоелектроніки, Україна, Харків

Давидович Ілля Вікторович,
асистент кафедри біомедичної кібернетики КПІ ім. Ігоря Сікорського, Україна, Київ

Дунаєва Тамара Альбінівна,
кандидат фізико-математичних наук, доцент кафедри технічних та програмних засобів автоматизації КПІ ім. Ігоря Сікорського, Україна, Київ

Єрешко Юлія Олександрівна,
доктор економічних наук, професор кафедри економічної кібернетики КПІ ім. Ігоря Сікорського, Україна, Київ, науковий співробітник Технічного університету Мюнхена, Німеччина

Зайченко Олена Юріївна,
професор, доктор технічних наук, професор кафедри математичних методів системного аналізу ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Зайченко Юрій Петрович,
професор, доктор технічних наук, професор кафедри математичних методів системного аналізу ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Коваленко Олександр Миколайович,
аспірант кафедри електронних пристроїв та систем факультету електроніки КПІ ім. Ігоря Сікорського, Україна, Київ

Ковальчук Дмитро Вікторович,
директор підприємства «Відкрите акціонерне товариство «НВО «Червона хвиля»», Україна, Київ

Кузьменко Олексій Віталійович,
аспірант ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Куссуль Наталія Миколаївна,
професор, доктор технічних наук, завідувачка кафедри математичного моделювання та аналізу даних Навчально-наукового фізико-технічного інституту КПІ ім. Ігоря Сікорського, Україна, Київ

Лавренюк Алла Миколаївна,
доцент, кандидат технічних наук, доцент кафедри математичного моделювання та аналізу даних Навчально-наукового фізико-технічного інституту КПІ ім. Ігоря Сікорського, Україна, Київ

Лазоришинець Василь Васильович,
член-кореспондент НАН України, академік НАМН України, професор, доктор медичних наук, директор Національного інституту серцево-судинної хірургії ім. М.М. Амосова НАМН України, Київ

Левенчук Людмила Борисівна,
доктор філософії, старший викладач кафедри математичних методів системного аналізу ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Мельник Ігор Віталійович,
професор, доктор технічних наук, професор кафедри електронних пристроїв та систем факультету електроніки КПІ ім. Ігоря Сікорського, Україна, Київ

Настенко Євген Арнольдович,
професор, доктор біологічних наук, професор кафедри біомедичної кібернетики КПІ ім. Ігоря Сікорського, Україна, Київ

Перепека Євген Олександрович,
аспірант, лікар-хірург відділення складних аритмій з рентгеноопераційною ДУ «НІССХ ім. М.М. Амосова НАМН України», Київ

Салій Євгеній Валерійович,
студент Навчально-наукового фізико-технічного інституту КПІ ім. Ігоря Сікорського, Україна, Київ

Скрипка Михайло Юрійович,
аспірант кафедри електронних пристроїв та систем факультету електроніки КПІ ім. Ігоря Сікорського, Україна, Київ

Статкевич Віталій Михайлович,
кандидат фізико-математичних наук, науковий співробітник відділу прикладного нелінійного аналізу ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Суржиков Микола Сергійович,
магістрант кафедри електронних пристроїв та систем факультету електроніки КПІ ім. Ігоря Сікорського, Україна, Київ

Тимощук Оксана Леонідівна,
доцент, кандидат технічних наук, завідувачка кафедри математичних методів системного аналізу ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Тугай Сергій Борисович,
доцент, кандидат технічних наук, доцент кафедри електронних пристроїв та систем факультету електроніки КПІ ім. Ігоря Сікорського, Україна, Київ

Цеслів Олександр Степанович,
магістр Київського національного університету імені Тараса Шевченка, Україна, Київ

Цеслів Ольга Володимирівна,
кандидат технічних наук, доцент кафедри економічної кібернетики КПІ ім. Ігоря Сікорського, Україна, Київ

Швед Ірина Сергіївна,
магістрант кафедри електронних пристроїв та систем факультету електроніки КПІ ім. Ігоря Сікорського, Україна, Київ

Abdellatif El Ouissari,
Engineering Science Laboratory (LSI), Polydisciplinary Faculty of Taza, Sidi Mohamed Ben Abdellah University, Fez, Morocco; ESISA Analytica Laboratory, Higher School of Engineering in Applied Sciences, Fes, Morocco

Izz K. Abboud,
Master of Engineering, Computer Engineering Department, College of Engineering, Mustansiriyah University, Baghdad, Iraq

Karim El Moutaouakil,
Professor, Engineering Science Laboratory (LSI), Polydisciplinary Faculty of Taza, Sidi Mohamed Ben Abdellah University, Fez, Morocco

Muaayed F. Al-Rawi,
Researcher, Computer Engineering Department, College of Engineering, Mustansiriyah University, Baghdad, Iraq

Nasir A. Al-Awad,
Professor, Computer Engineering Department, College of Engineering, Mustansiriyah University, Baghdad, Iraq

Punitha Mahadevappa,
Research Scholar, Assistant Professor, Department of Information Science and Engineering, JSS Academy of Technical Education, Bangalore, Karnataka, India

Rekha Puranic Math,
Professor and Head of the Department of Information Science and Engineering, JSS Academy of Technical Education, Bangalore, Karnataka, India