Computer Science and Interdisciplinary Research Journal

Volume 1, Issue 1, 2024

ISSN: 3033-1218



CSIR Journal

https://csirjournal.com

Computer Science and Interdisciplinary Research Journal [CSIR Journal]

ISSN: 3033-1218
Volume 1, Issue 1, 2024.

Published by Science, Education, Technologies Association
Plovdiv, Bulgaria

Licensed under CC BY-NC-ND 4.0.

The views expressed in the articles are those of the authors and do not necessarily reflect the views of the publisher or editorial team.

Editorial Team

Editors-in-Chief

Section COMPUTER SCIENCE

Prof. Asya Stoyanova-Doycheva, Ph.D.

University of Plovdiv Paisii Hilendarski, Plovdiv, Bulgaria

Section MATHEMATICS

Assoc. Prof. Mariyan Milev, Ph.D.

Sofia University St. Kliment Ohridski, Sofia, Bulgaria

Section EDUCATION

Prof. Todorka Glushkova, Ph.D.

University of Plovdiv Paisii Hilendarski, Plovdiv, Bulgaria

Section INTERDISCIPLINARY RESEARCH

Assoc. Prof. Stanka Hadzhikoleva, Ph.D.

University of Plovdiv Paisii Hilendarski, Plovdiv, Bulgaria

Editorial Board

Prof. Aldo Tagliani, Ph.D.

Università di Trento, Trento, Italy

Prof. Emil Hadzhikolev, Ph.D.

University of Plovdiv Paisii Hilendarski, Plovdiv, Bulgaria

Assoc. Prof. Senthil Kumar, Ph.D.

Mepco Schlenk Engineering College, Sivakasi, India

Assoc. Prof. Margarita Terziyska, Ph.D.

University of Food Technology, Plovdiv, Bulgaria

Assoc. Prof. Stoyan Cheresharov, Ph.D.

University of Plovdiv Paisii Hilendarski, Plovdiv, Bulgaria

Assoc. Prof. Velika Kuneva, Ph.D.

Agricultural University, Plovdiv, Bulgaria

Assoc. Prof. Silvia Gaftandzhieva, Ph.D.
University of Plovdiv Paisii Hilendarski, Plovdiv, Bulgaria

Assoc. Prof. Hristina Kulina, Ph.D.
University of Plovdiv Paisii Hilendarski, Plovdiv, Bulgaria

Assoc. Prof. Mikhail Kolev, Ph.D.
University of Architecture, Civil Engineering and Geodesy, Sofia, Bulgaria

Assist. Prof. Shih-Hao Chang, Ph.D.
National Taipei University of Technology, Taiwan

Assist. Prof. Gordana Rakić, Ph.D.
University of Novi Sad, Serbia



English Proofreader

Assoc. Prof. Vanya Ivanova, Ph.D.
University of Plovdiv Paisii Hilendarski, Plovdiv, Bulgaria



Technical Assistance

Ph.D. student Maria Gorgorova
University of Plovdiv Paisii Hilendarski, Plovdiv, Bulgaria

Preface

Dear Readers,

It is a great pleasure to present the first issue of the *Computer Science and Interdisciplinary Research* journal. This journal was founded with the vision to provide a platform for sharing scientific research, innovative ideas, and best practices in the fields of computer science and interdisciplinary studies. Our mission is to promote the work and achievements of scientists, professionals, and educators who are dedicated to exploring sustainable solutions that drive economic and societal progress.

The first issue of *CSIR Journal* is very special to us, because it marks a new beginning. The articles included cover a wide range of topics – from innovations in elderly care, modeling of financial derivatives, and suicide risk analysis to studies on life expectancy using artificial intelligence, as well as the application of ontologies in e-learning. These articles reflect our team's commitment to provide high-quality and diverse scientific content.

We are also proud to announce that *CSIR Journal* now holds an ISSN, establishing its status as a scientific publication, enhancing its discoverability in academic circles, and ensuring its long-term preservation. Each article published in the journal is assigned a unique DOI (Digital Object Identifier), which facilitates access to the publications and makes citation easier for other researchers.

We thank all the authors, reviewers, and members of the editorial board who contributed to making this issue a reality. Your dedication and professionalism form the foundation of this new publication's success. We also thank our readers for their trust, and we hope that *CSIR Journal* will become a valuable source of knowledge, inspiration, and new ideas for everyone with an interest in science and technology.

Welcome to the pages of the first issue of *CSIR Journal*!

Enjoy reading!

EDITORS-IN-CHIEF:

Prof. Asya Stoyanova-Doycheva, Ph.D.
Section COMPUTER SCIENCE

Assoc. Prof. Mariyan Milev, Ph.D.
Section MATHEMATICS

Prof. Todorka Glushkova, Ph.D.
Section EDUCATION

Assoc. Prof. Stanka Hadzhikoleva, Ph.D.
Section INTERDISCIPLINARY RESEARCH

Table of contents

Preface 1
Asya Stoyanova-Doycheva, Mariyan Milev, Todorka Glushkova, Stanka Hadzhikoleva

Transforming Nursing Home Care: An Integrated Approach Using Sensors, AI, and Monitoring Technologies 3
Stoyan Cheresharov, Georgi Dragomirov, Georgi Gustinov, Stanka Hadzhikoleva, Kostadin Yotov

Computational and Implementation Problems in Option Pricing with Long Maturity 14
Mariyan Milev

Analysis of Suicide Risk in European Countries Using Artificial Intelligence Methods 18
Valentin Georgiev, Kostadin Yotov

Impact of Global Country Indicators on Life Expectancy 24
Valentin Georgiev, Stanka Hadzhikoleva, Emil Hadzhikolev

Using Ontologies in e-Learning in Software Engineering Course 33
Asya Stoyanova-Doycheva, Georgi Cholakov, Todorka Glushkova

Computer Science and Interdisciplinary Research Journal
Volume 1, Issue 1, ISSN: 3033-1218, DOI: 10.70862/CSIR.2024.0101-01, 2024

Transforming Nursing Home Care: An Integrated Approach Using Sensors, AI, and Monitoring Technologies

Stoyan Cheresharov[1], Georgi Dragomirov[2], Georgi Gustinov[3], Stanka Hadzhikoleva[4], Kostadin Yotov[5]

University of Plovdiv Paisii Hilendarski, 236, Bulgaria Blvd, 4027 Plovdiv, Bulgaria

[1] cheresharov@uni-plovdiv.bg, [2] georgiuk86@gmail.com, [3] jbd30@coolcsn.com, [4] stankah@uni-plovdiv.bg, [5] kostadin_yotov@uni-plovdiv.bg

Abstract – This paper introduces an innovative integrated system leveraging sensors, artificial intelligence (AI), and monitoring technologies to revolutionize care in nursing homes. Despite the availability of various digital tools, the absence of a cohesive system that synergizes these technologies represents a significant gap in current nursing home care. Our approach unifies these technologies into a single, comprehensive environment, addressing the fragmentation in existing solutions and targeting the holistic improvement of care delivery. By automating data collection and analysis, facilitating seamless information sharing, and simplifying user interaction with the system, we propose a model that significantly enhances operational efficiency, care quality, and user experience. The development and implementation of this integrated system promise a shift towards more reliable, efficient, and user-friendly care in nursing homes, setting a new standard for digital healthcare environments.

Keywords – Nursing Homes, Digitalization, Integrated Care System, AI, Sensor Technologies, Internet of Things.

DOI: https://doi.org/10.70862/CSIR.2024.0101-01

Corresponding author: Stoyan Cheresharov, University of Plovdiv Paisii Hilendarski
Email: cheresharov@uni-plovdiv.bg

Received: 12 June 2024.
Accepted: 23 June 2024.
Published: 13 July 2024.

© 2024 Stoyan Cheresharov, Georgi Dragomirov, Georgi Gustinov, Stanka Hadzhikoleva & Kostadin Yotov. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

The article is published with Open Access at http://csirjournal.com/

1. Introduction

Nursing homes provide care and support elderly individuals who require constant assistance and medical attention. Challenges associated with this include limited staffing, the continuous monitoring of residents' health and well-being, addressing their social isolation, maintaining a high standard of living, etc. Caring for residents often demands significant time and effort from the staff. Some residents require periodic repositioning to prevent bedsores, while others need frequent checks to ensure their bedding is dry. Continuous physical observations monitor parameters such as heart rate, respiratory rate, and signs of discomfort or pain. Ensuring residents' unique needs and preferences is vital to their sense of comfort, particularly for those requiring specialized care. Staff actions can disrupt residents' sleep and overall comfort, especially during night shifts.

Although many vital sign observations can be automated, nursing homes often use multiple fragmented systems, which complicates coordination between them and hinders adaptation to increasing demands. Traditionally, caregivers rely on pagers, monitoring systems, bells, and other signals from residents. This model requires physical visits to provide care and services. An additional responsibility for staff is medication management, which includes organizing and supervising medication intake and maintaining documentation. Documentation of visits, observations, medication administration, and other interventions must be meticulously maintained, which is time-consuming and can divert staff from direct patient care.

This study presents various aspects of developing an integrated system tailored to the unique needs of nursing homes. A unified solution is proposed,

incorporating sensors, the Internet of Things (IoT), artificial intelligence (AI), and monitoring technologies. Key features of the system include unobtrusive monitoring of residents' health, safety, security, and privacy, along with a comprehensive reduction in redundant operational tasks and staff effort to improve efficiency and resident care. Additional functions of the system include scheduling and managing staff shifts and duties, among other features. The system utilizes a range of advanced sensors and technologies to transform care delivery. Key components include facial recognition cameras, vital sign monitors, and innovative bed sensors, all working together to enhance residents' safety and comfort. The sensors monitor health metrics silently and non-invasively, thus minimizing disruptions during rest periods. Additionally, our system incorporates a novel call bell system based on radio frequency (RF) technology, which ensures prompt communication between residents and caregivers without the constraints of traditional wired systems.

A key element of our approach is the integration with a patient care system, which allows for seamless aggregation and analysis of data collected by the sensors. The incorporation of AI further distinguishes our system. The AI engine analyzes the data to provide real-time insights and alerts to predict potential health issues based on behavioral and physiological parameters. The integration of the system in a real environment enhances the quality of services provided by staff, automates part of the activities, and increases residents' sense of freedom, dignity, security, and happiness. The proposed system model has the potential to revolutionize the standard of care in nursing homes, providing the comfort of a homelike atmosphere and making them more efficient and responsive to the individual needs of residents.

2. Related work. Literature review

The use of digital technologies in nursing homes has the potential to enhance resident care, clinical support, and administrative activities [1]. This is particularly important in the context of an aging population, where technology can help address the challenges of healthcare for older adults [2]. However, the acquisition of digital skills by staff and residents in care homes is influenced by psychosocial and environmental factors [3]. Despite the potential benefits, challenges in the use of digital technology in nursing care persist, highlighting the need for ongoing training and support [4].

2.1. Evolution of Nursing Home Care

Wan [5] provides a comprehensive overview of the current state of care in nursing homes, emphasizing the need for a service-oriented approach and the integration of technology. Karsch-Völk [6] examines the development of criteria for effective medical care in German nursing homes, highlighting the importance of collaboration between nursing home staff and physicians. Together, these studies underscore the need for a holistic approach to care in nursing homes, addressing both the physical and psychosocial well-being of residents.

The field of elderly care technology has advanced significantly in recent years, focusing on smart home solutions [7], assistive technology [8], and information technology [9]. These innovations aim to enhance the safety, security, and independence of elderly individuals while alleviating the burden on healthcare systems. Robotics in elderly care is another emerging area of interest [10], with the potential to further improve the quality of care provided to older adults.

2.2. Current Technologies in Nursing Homes

Numerous studies have explored the use of monitoring systems in nursing homes, with a focus on improving patient care and outcomes. Motwani [11] and Frech [12] highlight the potential of these systems to detect and address health issues in a timely manner, with Motwani specifically noting the use of sensors for data collection and wireless transmission to caregivers. Cislo [13] and Chang [14] further support the feasibility and benefits of these systems, with Cislo emphasizing the use of innovative sensors and a self-learning decision process, and Chang demonstrating the successful application of a wireless sensor network for vital signs monitoring. These studies collectively underscore the potential of monitoring systems, particularly those utilizing digital technologies, in enhancing the quality of care in nursing homes.

Passive health monitoring systems have been shown to significantly reduce care costs and improve the efficiency of professional caregivers in assisted living facilities [15]. These systems also enable informal caregivers to provide more support without increasing their burden or negatively impacting their quality of life [16]. Additionally, these systems are generally well-accepted by residents and may positively influence their quality of life [16]. However, their effect on life satisfaction among residents in independent senior housing remains inconclusive [16].

Various communication tools and technologies have been developed to support interactions among residents, caregivers, and families. Cranley [17] emphasizes the value of engaging residents and families in the development of a communication tool,

while Hoel [18] highlights the potential of technology-driven solutions, such as tablet computers, social robots, and computer systems, to enhance communication and interactions, particularly for people with dementia and their caregivers. Généreux [19] further supports the use of communication plans, which contain information about a resident's specific characteristics and communication disorder, to facilitate interactions with communication-impaired residents. These studies collectively underscore the importance of person-centered communication tools and technologies in long-term care.

Effective communication tools in nursing homes have been found to significantly impact resident health outcomes [20]. This is particularly important in the medication process, where an online communication platform has been shown to improve cooperation and systematize the process [21]. User and family satisfaction with nursing homes is also crucial, with determinants including quality of life, anxiety, and social and health factors [22]. Therefore, the use of effective communication tools can not only improve resident health outcomes but also contribute to user and family satisfaction in nursing homes.

Jiang [23] proposed a health monitoring system for nursing homes that utilizes lightweight security and privacy protection, including dual-band RFID and a mobile authentication protocol. This system offers real-time monitoring and high security, meeting the requirements of nursing home health monitoring. Levy [24] discussed the challenges of privacy and security in nursing home monitoring, particularly in the context of consumer-purchased monitoring devices. State regulations play a crucial role in balancing these interests. Samadbeik [25] emphasized the importance of user authentication and security controls in managing nursing data in the Electronic Health Record, with a focus on the viewpoints of IT experts and computer companies. These studies collectively highlight the need for robust security and privacy measures in nursing home monitoring systems, with a particular emphasis on user authentication and state regulations.

A range of studies have explored the benefits and limitations of security and access control systems in nursing homes. Li [26] and Jiang [23] both highlight the potential of these systems to enhance safety and efficiency, with Li's system focusing on real-time monitoring and Jiang's on lightweight security and privacy protection. Chinnasamy [27] emphasizes the importance of access control in protecting healthcare data, while Samadbeik [25] underscores the need for strong authentication and authorization methods. These studies only underline the need for reliable security measures and access control.

2.3. Challenges in Integrating Technology in Nursing Homes

The integration of technology in nursing homes is hindered by several challenges. Cost, resistance to change, and training requirements are key barriers [28]. There is a lack of information about the cost-effectiveness and other aspects of technologies, as well as limited resources for their purchase and development [29]. Management attitudes also play a role, with some administrators lacking interest in change and a systematic approach to technology adoption [30]. Additionally, the delivery of high-tech home care by hospital-based nursing units presents clinical and technical challenges [31]. These findings highlight the need for evidence-based decision-making, participatory planning, and professional training initiatives to overcome these barriers and improve the integration of technology in nursing homes.

Integrating technology in nursing homes is a complex process, with various challenges and limitations. Weernink [32] and Lin [33] both highlight the need for careful implementation and integration of new technologies, particularly in the context of lost and misplaced items and real-time positioning and monitoring services for the elderly. However, Freedman [29] identifies several barriers to the adoption of technology in U.S. residential long-term care settings, including lack of information, limited resources, concerns about liability, and regulatory constraints. Bourbonnais [34] further emphasizes the importance of understanding the perceptions and needs of care managers, formal and family caregivers in nursing homes, in order to develop useful technologies that can improve the quality of clinical practice.

Some ethical and privacy-related challenges need to be taken into consideration when using technology in nursing homes. Berridge [35] highlights the risks to residents' privacy and dignity, as well as the potential for staff and resident monitoring. Bourbonnais [36] emphasizes the need for pilot testing and coaching activities to facilitate technology adoption, while also addressing ethical risks and prejudices. Sifford [37] discusses the benefits and challenges of electronic surveillance in LTC research, particularly in investigating resident-to-resident violence. Hall [38] underscores the importance of involving all stakeholders in decision-making and training to ensure a deeper understanding of the potential benefits and challenges of monitoring technologies.

The integration of monitoring technologies in nursing homes, while beneficial for safety, presents ethical and privacy concerns [38]. These concerns are further complicated by the need for pilot testing and

coaching activities to facilitate adoption [36]. The use of surveillance cameras in nursing home rooms raises additional ethical issues, including privacy and dignity risks for residents and the potential for staff and resident monitoring [35]. Despite these challenges, electronic surveillance has been proposed as a method to address resident-to-resident violence, particularly in the context of dementia [37]. These studies highlight the need for a balanced approach to the integration of technology in nursing homes, one that prioritizes safety while addressing ethical and privacy concerns.

A qualitative study exploring nursing home care environments where nurse practitioners work was conducted by Gregory in [39]. The research supports the growing nurse practitioner workforce to assure higher quality care for older adults in nursing homes. The author claims that nursing homes with optimal care environments that support nurse practitioner roles, increased visibility, independence, and relationships are better positioned to support care of older adults. The study reports findings of thirteen qualitative interviews with nurse practitioners to explore facets of nursing home care environments and adapt a tool to measure care environments. The authors incorporated deductive and inductive coding to identify three major emerging themes impacting care environments: 1) nurse practitioner practice in nursing homes, 2) overall goals of practice, 3) workplace challenges. Some of the most highly important survey items measured nurse practitioner and physician relationships. Since our integrated system is aiming of improving the performance of the caregivers the mentioned study is important to us.

In the interesting study of Marjanovich et. al. the claim is made that between 2% and 5% of the population in industrialized countries lives in nursing homes [40]. The author noticed that the residents often lack medical referrers, especially at night and on weekends, and emergency departments (EDs) are often the last answer to medical questions. In the research have been included 3103 calls in the final analysis (355 from equipped nursing homes and 2748 from unequipped nursing homes). The proportion of patients dispatched to an ED after the first call was lower among telemedicine-equipped than among telemedicine-unequipped nursing homes (41% vs 50%; odds ratio, 0.71; 95% CI, 0.56–0.90). Telemedicine and using modern technologies is obviously improving the quality of service in the nursing homes.

The prevailing systems are predominantly reliant on manual checks and suffers from an absence of real-time data analysis capabilities. The literature review underscores a discernible void in the current methodologies—specifically, the absence of a comprehensive, AI-enhanced system that integrates diverse data streams effectively. This deficiency highlights the critical need for developing advanced technological frameworks that not only consolidate data from multiple sources but also utilize artificial intelligence to facilitate immediate and actionable insights.

3. Methodology, System Design and Architecture

We present a conceptual model for the development of a digital system tailored specifically for nursing homes. This model delineates the primary modules and abstract concepts integral to the system's architecture. Utilizing a cloud-native, service-oriented architectural framework, the model facilitates scalable and efficient data management and service delivery. The main conceptual framework and the corresponding modules are depicted in Figure 1, illustrating the interconnectivity and functional flow among the system components.

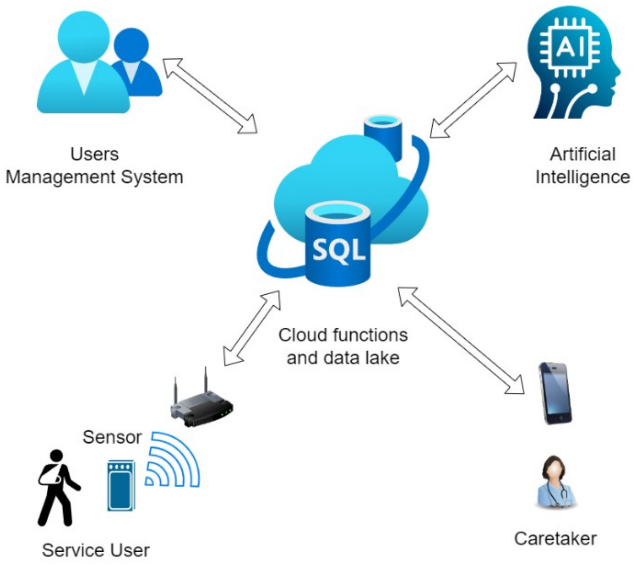


Figure 1. The formal conceptual model

The overall system architecture is illustrated in Figure 1. Within this infrastructure, sensors deployed across various rooms may originate from diverse manufacturers, contributing to variability in system integration. The simplest among these are the call bell buttons, which differ in connectivity; some are tethered through wired connections while others communicate wirelessly with a central hub. Notably, even identical hardware configurations may employ disparate communication protocols. Figure 2 provides a streamlined diagram of the sensor data flow within the system. While not exhaustive, this representation captures the fundamental pathways through which data is transmitted and integrated into the cloud

architecture, emphasizing the complexity and diversity of sensor integration in the proposed model.

One notable dataflow configuration illustrated in Figure 2. involves the direct connection of sensors, designated as "Wi-Fi sensors," to a router. These sensors, characteristic of Internet of Things (IoT) devices, possess the capability to autonomously establish a Wi-Fi connection and transmit data directly to a designated endpoint. Although this configuration is favored for its straightforward installation and operational simplicity, it incurs a substantial power demand, typically around 120 mA. Consequently, such arrangements generally necessitate a continuous power supply from the main electricity grid. Reliance on battery power alone would lead to rapid depletion, typically within two to three days, rendering it impractical for sustained use without frequent battery replacements or recharging.

The configuration denoted as "Wired sensor" in Figure 2 presents its own set of advantages and disadvantages. This setup ensures a stable and reliable connection due to the physical wiring to each sensor. However, the installation of these cables introduces logistical challenges. Although employing Power over Ethernet (PoE) can obviate the need for separate power supplies or batteries, in its absence, each sensor must either be connected to the main power grid or equipped with batteries. This necessity may complicate installation and increase maintenance requirements, especially in environments where access to power sources is constrained.

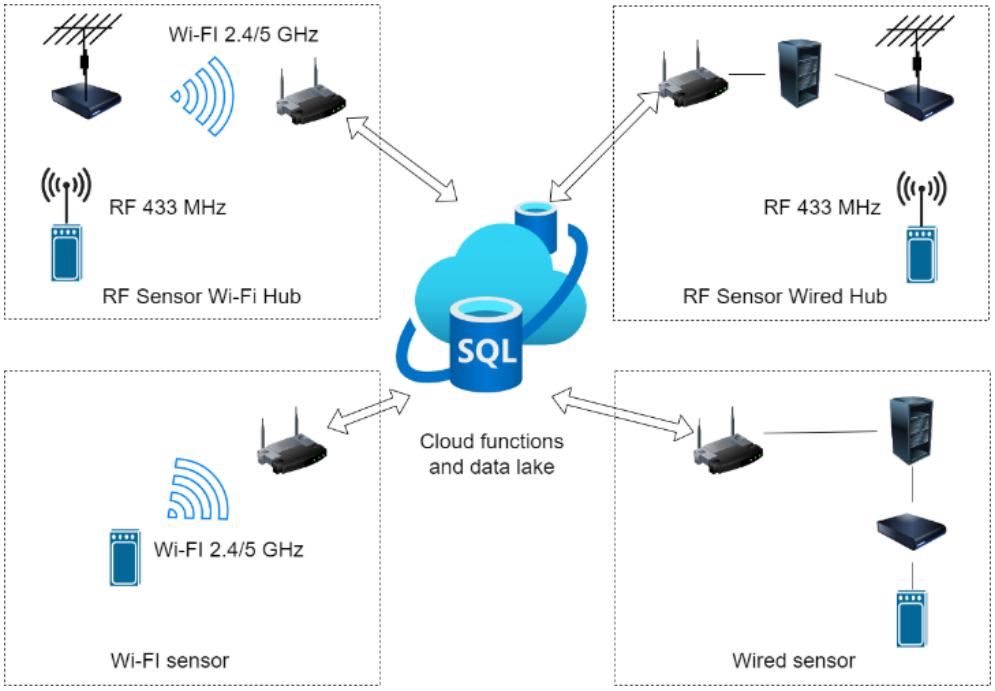


Figure 2. Sensors data flows

The dataflow configuration labeled as "RF Sensor Wired Hub" in Figure 2. employs a central hub that receives signals from all connected sensors and transmits these via a wired connection to a server. This setup facilitates system maintenance by centralizing signal reception and enabling remote server access, thereby streamlining administrative tasks. However, the scalability of this configuration poses potential challenges; the central hub might become overwhelmed with data processing demands if the number of sensors is large. Additionally, the risk of signal interference increases when many sensors transmit simultaneously on the same frequency, potentially leading to signal jamming and data loss.

The configuration termed "RF sensor Wi-Fi Hub" depicted in Figure 2, implements a decentralized approach where each room is equipped with a receiving hub. This hub collects signals from all sensors within the room and subsequently transmits the data to the cloud via a Wi-Fi connection to the router. This arrangement enhances the system's resilience and scalability by distributing the data collection points, thereby reducing the load on any single device and facilitating robust signal transmission even in complex environments.

In each of these configurations, the deployment of repeaters for either lower frequency or Wi-Fi signals can enhance system performance. These devices are instrumental in extending the range of signal transmission, ensuring consistent and reliable communication across larger or structurally complex environments.

Additionally, cameras have been installed in each room as part of a strategic upgrade, paving the way for the next generation of the system. This future enhancement will incorporate facial and expression recognition technologies, enabling the system to generate events based on nuanced visual data. By analyzing facial expressions, the system will offer a more sophisticated level of monitoring, capable of detecting subtle cues that may indicate changes in a resident's emotional or physical state. This advancement is expected to significantly enrich the system's responsiveness and accuracy, providing caregivers with timely and pertinent alerts based on real-time visual analytics.

Reliability is paramount in the design and operation of this system. To ensure ongoing functionality, a heartbeat monitoring system is integral. This system periodically sends signals through all components to verify their operational status. Alongside these heartbeat checks, payload data carrying useful information may be concurrently transmitted, enriching the system's utility. However, the primary value of the heartbeat signal lies in its role as a diagnostic tool; receipt of this signal confirms the proper functioning of all system elements. Conversely, if a heartbeat signal is not received within a domain-specific predefined period, it is presumed that there is a malfunction. Consequently, an alert is triggered to notify maintenance personnel of the issue, prompting immediate attention to prevent system failure.

As outlined earlier, the four data flow models presented serve as a general framework for understanding potential configurations within the system. Detailed exploration of each configuration is beyond the scope of this publication, as each model warrants an in-depth study due to its unique technical nuances. Various communication protocols, such as EV1527, and modulation techniques, including Amplitude Shift Keying (ASK) and On-Off Keying (OOK), could be employed depending on specific requirements. Additionally, for applications requiring transmission over greater distances, Long Range (LoRa) communication technology offers a viable solution, enhancing the flexibility and reach of the system.

Data originating from various sensors, each utilizing a distinct data flow model as depicted in Figure 2. , is typically transmitted in binary or simple text formats. These data streams are captured by specialized hardware devices that interface with server computers via RS232 or USB connections. On these servers, dedicated services are continuously operational, tasked with the critical functions of reading and interpreting the incoming data. Following this initial processing, the data is converted into JSON format, which standardizes the information for enhanced interoperability and ease of handling. Subsequently, this formatted data is transmitted to designated endpoints within the cloud infrastructure, ensuring timely and accurate data availability for further analysis and application.

The endpoints within the system consist of functional applications, each equipped with HTTP-triggered functions that facilitate responsive data processing upon receipt of incoming data. Additionally, certain functions within these applications are configured to activate on a time-triggered basis, systematically processing data that has been previously collected and stored. This architecture allows for the deployment of multiple functional applications, each tailored with specific functions designed to meet diverse operational needs. This modular approach not only enhances the system’s flexibility but also optimizes the efficiency of data processing tasks across different segments of the network.

The consolidation of business logic and services within the cloud framework underpins our system architecture, enabling remote management and development capabilities. This cloud-centric approach confers numerous advantages. Primarily, it offloads the responsibility for infrastructure maintenance to cloud service providers, thereby alleviating the need for extensive local IT resources. Such decentralization allows for meticulous control over individual functions and optimizes the utilization of cloud-specific features. Applications developed in a cloud-native environment inherently exhibit enhanced robustness, stability, and scalability. Furthermore, these applications are simpler to maintain and evolve due to their integration with the dynamic, resource-elastic infrastructure provided by cloud platforms.

The cloud functions serve a pivotal role in aggregating data from diverse sensors and sources, recording this information into a cloud-based data lake. This repository houses all sensitive information pertaining to service users, ensuring comprehensive data capture in a secure and scalable environment. The utilization of a data lake not only facilitates robust data storage capabilities but also supports advanced analytics by providing a centralized point of access to a wide array of data types and structures. This architecture enhances the ability to extract actionable insights while rigorously maintaining data privacy and security standards essential for handling sensitive information.

Time-triggered functions within the system are designed to transmit essential data to third-party APIs for analysis and processing, facilitating a seamless

integration of external computational resources. Additionally, some sensor providers enable access to sensor data via web sockets or RESTful APIs. This arrangement allows for two modes of data retrieval: active polling by our system at regular intervals, or passive reception where data is pushed directly to our endpoints by the providers. Both methods ensure that the system remains consistently updated with the latest sensor data, enhancing the timeliness and accuracy of the information available for processing and decision-making.

Certain sensor providers, such as those manufacturing wristwatch devices, utilize available mobile networks to transmit data directly to their respective servers. Subsequently, this data is pushed to designated endpoints within our system. This method facilitates a continuous stream of data flow, ensuring that our system receives timely updates. Such integration allows for real-time monitoring and responsive interaction with the data received, enhancing the overall efficacy of the service provided by these sensors.

Dedicated virtual machines within the cloud infrastructure are configured to manage data transactions with the RESTful APIs of sensor providers. These virtual machines actively pull data from and push data to these APIs, facilitating a dynamic exchange of information. This setup ensures that the data integration is not only seamless but also robust, allowing for continuous synchronization and real-time updates. The use of virtual machines specifically for this task leverages their scalability and reliability, which are critical for maintaining consistent performance and availability in cloud-based systems.

The data lake serves as the principal reservoir of data for Artificial Intelligence (AI) providers. Raw data is routinely pushed from the data lake to these providers at regular intervals, facilitating continuous data analysis. Upon analyzing this data, AI providers generate insights that are encapsulated in events. These events are then transmitted back to our system and directed to specific endpoints designed to handle this information. Simultaneously, these events are also archived within the data lake for future reference and analysis. This cyclical flow of data ensures that the insights derived from AI analysis are both actionable and integrative, enhancing the overall decision-making processes within the system.

Cloud-based functions play a crucial role in analyzing sensor data and information received from Artificial Intelligence (AI) systems. Upon analysis, these functions generate specific events based on predefined criteria tailored to each type of sensor data. For instance, events may include notifications such as "wet bed," activation of a call bell, or alerts indicating that a resident has left the bed. These events are crucial for operational responsiveness and are not broadcast indiscriminately. Instead, they are strategically routed to handheld devices of specifically assigned caregivers. This targeted delivery ensures that alerts are promptly addressed by the responsible personnel, enhancing the efficiency and effectiveness of the caregiving process. Each type of sensor data has distinct boundaries and criteria that define the conditions under which these events are triggered, ensuring that the system's responses are both appropriate and timely.

The mobile application serves as a sophisticated user interface for caregivers within the system. It not only presents the generated events but also categorizes and prioritizes them according to their urgency and relevance. Each type of event is visually distinct within the application to facilitate quick recognition and response. Additionally, the integration of text-to-speech technology enhances the usability of the application; it vocalizes event notifications, allowing caregivers to receive critical updates audibly. This feature is particularly beneficial as it enables caregivers to stay informed without the necessity of continuously monitoring their devices visually, thereby improving their efficiency and allowing them to maintain focus on direct patient care.

To enhance security and operational efficiency, a specialized Mobile Device Management (MDM) system is implemented to manage and secure the mobile devices used by caregivers. This MDM system is configured to grant specific privileges to the mobile application responsible for processing events, ensuring it operates with the necessary authority and access. Furthermore, the MDM restricts the functionality of these devices to only those applications that are pertinent to caregiving tasks. This controlled environment prevents the execution of non-essential apps, thereby safeguarding sensitive data and optimizing the devices' performance for professional use. This strategic application of MDM technology plays a crucial role in maintaining the integrity and confidentiality of patient-related information, while also ensuring that caregivers have seamless access to vital operational tools.

4. Implementation and Validation

As a culmination of our research, the architecture previously delineated was successfully employed to construct a digital system for nursing homes. This system serves as a proof of concept, demonstrating the practical viability and effectiveness of the integrated technologies. By implementing this architecture, we have established a foundational model that not only

addresses the specific needs of nursing home care but also showcases the potential for broader applications in similar healthcare environments. This proof of concept validates the theoretical constructs and operational strategies proposed, paving the way for further development and refinement in future iterations.

The development of the described system was guided by agile software development principles, with continuous input from caregivers and management to ensure relevance and usability. The cloud functions were developed using C#, a modern programming language that benefits from robust support and seamless integration within Microsoft's ecosystem. This integration was a pivotal factor in selecting Azure as the cloud provider, as it offers well-coordinated tools that facilitate rapid application development. Furthermore, the Microsoft environment supports Continuous Integration and Continuous Delivery (CI/CD) pipelines that are well-suited to C#, .NET, and Azure technologies. This integration enhances the development workflow, simplifying the processes of application deployment and iterative updates, thereby ensuring that the system remains adaptable and aligned with evolving user needs.

Within our system, a mix of third-party and bespoke sensors is employed to capture data tailored to specific care requirements. Third-party sensors transmit data to their respective providers' servers, which then reroute this information to our designated endpoints or web sockets, ensuring seamless integration into our architecture. Conversely, many sensors are custom-developed within our organization to align precisely with unique operational needs. These are typically constructed using Single Board Computers (SBCs), Systems on Chip (SoCs), and Microcontrollers (MCUs). Notably, certain microcontrollers and development boards are equipped with advanced connectivity features such as Wi-Fi and Bluetooth, making them ideally suited for our application. Examples of such devices include the Espressif ESP 32 and ESP 8266, which offer robust performance and versatility, crucial for the real-time data communication demands of our system.

A primary challenge in implementing sensors within our system is ensuring their reliability while adhering to several operational constraints: they must be battery-operated, possess long battery life, be easy to install and relocate, and maintain an aesthetically pleasing appearance. Given these requirements, the use of power-intensive technologies such as Wi-Fi and Bluetooth is often impractical. Consequently, our approach has shifted towards utilizing lower, license-free frequencies, which are less demanding on power resources. The transmitters employed in this spectrum are smaller, consume less power, and consequently have a longer operational life. Integrating these with Microcontrollers (MCUs) set to sleep mode, paired with low-power transmitters, forms an optimal solution. This combination ensures extended durability and efficiency, meeting the system's need for continuous, reliable performance without frequent maintenance.

Data transmission within our system can be executed through two distinct methods: encoded and modulated, or solely modulated without encoding. For instances where encoding is utilized, the EV1527 encoding scheme is employed. This specific method allocates 20 bits for the unique address of the transmitting device, ensuring precise identification, while the remaining 4 bits are dedicated to transmitting the actual data. This structure enhances the security and integrity of the data transmission process, enabling reliable and distinct communication between devices within the network.

The system underwent rigorous testing in a controlled environment over a period of one month, followed by a scaled-down implementation in the field to gather real-world data and feedback. During these tests, the system achieved a 99% accuracy rate in recognizing and transmitting events, demonstrating its robustness and reliability. Additionally, the power consumption metrics for sensors equipped with RF transmitters were quantified, showing a draw of 10 mA at a voltage range of 3.2 to 5 V. These results affirm the system's effectiveness in operational conditions and its efficiency in terms of energy usage, critical factors for long-term deployment in dynamic environments.

Aggregating collected data into a data lake facilitates various analytical operations, allowing for diverse slicing and dicing of the data to extract meaningful insights. To visualize and monitor this data effectively, Power BI was employed as the tool of choice. This business analytics service provides robust data representation capabilities, enabling the creation of interactive visualizations and comprehensive dashboards. Utilizing Power BI enhances our ability to interpret complex datasets and monitor system performance in real time, making it an ideal solution for overseeing and refining the operational processes within our system.

5. Discussion

The proposed system was implemented to enhance the organization and performance across 14 nursing homes. It has proven effective in its core functions, detecting events and promptly alerting caregivers. This capability ensures that critical information is communicated swiftly and accurately, facilitating

timely interventions and improving the overall care process. The effectiveness of the system in these settings demonstrates its utility in enhancing operational efficiency and the quality of care provided to residents, thereby affirming its value as a significant technological advancement in the healthcare sector.

The system provides an integrated approach by seamlessly combining third-party software and artificial intelligence with sensor data from various providers. It efficiently processes and sends event notifications to the handheld devices of caregivers, ensuring that they receive timely and relevant information. Furthermore, the system supports integration with multiple data providers while offering a single sign-on feature, which simplifies the authentication process and enhances user convenience. This capability allows caregivers to access a unified platform that aggregates diverse data streams, facilitating a comprehensive overview and efficient management of care activities. This holistic integration not only improves operational efficiency but also enhances the responsiveness and adaptability of caregiving practices in dynamic healthcare environments.

The implementation of the system has had a demonstrably positive impact on several key aspects of nursing home operations. Specifically, it has streamlined caregiver workflows, significantly enhancing their efficiency and reducing response times to resident needs. Additionally, the system has bolstered resident safety by enabling more precise and timely monitoring, thus facilitating early intervention in potential issues. Overall, the quality of care provided within these facilities has seen marked improvements, evidenced by increased satisfaction among residents and staff alike. These enhancements collectively demonstrate the system's effectiveness in transforming the care environment through advanced technological integration.

Future enhancements of the system could involve the integration of cameras equipped with facial recognition technology to bolster security measures. This addition would allow caregivers to remotely verify the status of residents without physical presence, enhancing efficiency and response times. Communication capabilities could be augmented through microphones and loudspeakers embedded in the cameras, facilitating direct interaction with residents. Additionally, these cameras could ensure that only authorized individuals gain access to the facility and specific rooms, further enhancing security.

The potential for expanding the role of Artificial Intelligence (AI) is considerable. Future iterations of the system could incorporate expression recognition to generate alerts based on residents' emotional or physical distress signals. AI could also play a more significant role in medical management, assisting with medication prescriptions and preliminary diagnostics. Moreover, AI-driven workflows could optimize operational tasks, assigning responsibilities based on real-time needs to maintain optimal living conditions for residents.

With the rapid advancements in AI and robotics, coupled with continuous improvements in medical technology, the possibilities for enhancing care through technology are virtually limitless. These advancements promise not only to improve the efficiency of care delivery but also to significantly elevate the quality of life for residents by providing more personalized and proactive care solutions.

Acknowledgements

This work was partially funded by the MUPD23-FMI-021 project of the Research Fund of the University of Plovdiv “Paisii Hilendarski”.

References

[1] G. L. Alexander and D. S. Wakefield, "Information Technology Sophistication in Nursing Homes," J. Am. Med. Dir. Assoc., vol. 10, no. 6, pp. 398-407, 2009. https://doi.org/10.1016/j.jamda.2009.02.001

[2] M. J. Rantz et al., "Improving nurse care coordination with technology," CIN - Comput. Informatics Nurs., vol. 28(6), pp. 325-332, 2010. https://doi.org/10.1097/NCN.0b013e3181f69838

[3] D. Wild, A. Kydd, and A. Szczepura, "Implementing digital skills training in care homes: A literature review: In the first of two articles, deidre wild and colleagues look at the use of technology to increase learning, and possible barriers to its uptake by residents and staff," Nurs. Older People, vol. 28(4), pp. 26-29, 2016. https://doi.org/10.7748/nop.28.4.26.s22

[4] N. T. Chaharsoughi, M. Ahmadifaraz, and L. S. Kahangi, "The impact of digital technologies in nursing care and their application: A narrative review," J. Multidiscip. Care, vol. 11(3), pp. 149-156, 2022. https://doi.org/10.34172/jmdc.2022.1127

[5] T. Wan, "Improving the quality of care in nursing homes: an evidence-based approach," 2010. https://doi.org/10.56021/9780801897184

[6] M. Karsch-Völk, J. Lüssenheide, K. Linde, E. Schmid, and A. Schneider, "Entwicklung eines Kriterienkatalogs für eine gelungene ärztliche Versorgung in Pflegeeinrichtungen," Z. Evid. Fortbild. Qual. Gesundhwes., vol. 109(8), pp. 570-577, 2015. https://doi.org/10.1016/j.zefq.2015.06.008

[7] S. Majumder et al., "Smart Homes for Elderly Healthcare-Recent Advances and Research Challenges," Sensors 2017, Vol. 17(11), p. 2496, 2017. https://doi.org/10.3390/s17112496

[8] F. G. Miskelly, "Assistive technology in elderly care," Age Ageing, vol. 30(6), pp. 455-458, 2001. https://doi.org/10.1093/ageing/30.6.455

[9] J. Soar and Y. Seo, "Health and Aged Care Enabled by Information Technology," Ann. N. Y. Acad. Sci., vol. 1114(1), pp. 154-161, 2007. https://doi.org/10.1196/annals.1396.040

[10] E. G. Christoforou, A. S. Panayides, S. Avgousti, P. Masouras, and C. S. Pattichis, "An Overview of Assistive Robotics and Technologies for Elderly Care," IFMBE Proc., vol. 76, pp. 971-976, 2020. https://doi.org/10.1007/978-3-030-31635-8_118

[11] K. Motwani, D. Mirchandani, Y. Rohra, H. Tarachandani, and A. Yeole, "Smart Nursing Home Patient Monitoring System," Imp. J. Interdiscip. Res., 2016.

[12] J. Frech, D. Jackson, J. Naber, A. Masukevitch, and T. King, "Nursing Home Patient Monitoring System for Improved Healthcare and Outcomes," IEEE Int. Conf. Consum. Electron., 2024. https://doi.org/10.1109/ICCE59016.2024.10444209

[13] N. Cislo et al., "A System for Monitoring Elderly and Dependent People in Nursing Homes: the E-monitor'age Concept," Stud. Inform. Univ., 2013.

[14] Y. J. Chang, C. H. Chen, L. F. Lin, R. P. Han, W. T. Huang, and G. C. Lee, "Wireless Sensor Networks for Vital Signs Monitoring: Application in a Nursing Home," vol. 2012, 2012. https://doi.org/10.1155/2012/685107

[15] M. Alwan et al., "Impact of Passive Health Status Monitoring to Care Providers and Payers in Assisted Living," https://home.liebertpub.com/tmj, vol. 13(3), pp. 279-285, 2007. https://doi.org/10.1089/tmj.2006.0056

[16] M. Alwan, S. Kell, B. Turner, S. Dalal, D. Mack, and R. Felder, "Psychosocial Impact of Passive Health Status Monitoring on Informal Caregivers and Older Adults Living in Independent Senior Housing," pp. 808-813, 2006. https://doi.org/10.1109/ICTTA.2006.1684477

[17] L. Cranley et al., "Development of communication tool for resident- and family-led care discussions in long-term care through patient and family engagement," Int. J. Older People Nurs., vol. 17(2), p. e12429, 2022. https://doi.org/10.1111/opn.12429

[18] V. Hoel, C. M. Feunou, and K. Wolf-Ostermann, "Technology-driven solutions to prompt conversation, aid communication and support interaction for people with dementia and their caregivers: a systematic literature review," BMC Geriatr., vol. 21(1), p. 157, 2021. https://doi.org/10.1186/s12877-021-02105-0

[19] S. Généreux, M. Julien, C. Larfeuil, V. Lavoie, O. Soucy, and G. Le Dorze, "Using communication plans to facilitate interactions with communication-impaired persons residing in long-term care institutions," Aphasiology, vol. 18(12), pp. 1161-1175, 2004. https://doi.org/10.1080/02687030444000507

[20] P. A. Arling, K. Abrahamson, E. J. Miech, T. S. Inui, and G. Arling, "Communication and effectiveness in a US nursing home quality-improvement collaborative," Nurs. Health Sci., vol. 16(3), pp. 291-297, 2014. https://doi.org/10.1111/nhs.12098

[21] N. Nestler, A. Mahlknecht, U. Bauer, L. Krisch, J. Osterbrink, and M. Flamm, "Optimization of Communication to Improve the Medication Process in Nursing Home Care - a Qualitative Evaluation," 2021. https://doi.org/10.21203/rs.3.rs-149559/v1

[22] V. Vassimon-Barroso, C. B. F. Pantoni, M. S. Zazzetta, D. L. Ferreira, F. A. Vasilceac, and J. H. Ansai, "User and family satisfaction with nursing homes: a systematic review," Int. J. Qual. Heal. Care, vol. 33(1), pp. 1-10, 2021. https://doi.org/10.1093/intqhc/mzaa160

[23] Y. Jiang and J. Liu, "Health Monitoring System for Nursing Homes with Lightweight Security and Privacy Protection," J. Electr. Comput. Eng., vol. 2017, 2017. https://doi.org/10.1155/2017/1360289

[24] K. Levy, L. Kilgour, and C. Berridge, "Regulating Privacy in Public/Private Space: The Case of Nursing Home Monitoring Laws," Elder Law J., vol. 26, 2018.

[25] M. Samadbeik, Z. Gorzin, M. Khoshkam, and M. Roudbari, "Managing the Security of Nursing Data in the Electronic Health Record," Acta Inform. Medica, vol. 23(1), p. 39, 2015. https://doi.org/10.5455/aim.2015.23.39-43

[26] Y. Li, Y. Xu, Y. Lv, and J. Wang, "A Web Based Security Monitoring and Information Management System for Nursing Homes," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 11742 LNAI, pp. 72-81, 2019. https://doi.org/10.1007/978-3-030-27535-8_7

[27] P. Chinnasamy, P. Deepalakshmi, and K. Shankar, "An analysis of security access control on healthcare records in the cloud," Intell. Data Secur. Solut. e-Health Appl., pp. 113-130, 2020. https://doi.org/10.1016/B978-0-12-819511-6.00006-6

[28] W. Broughton, "Health Information Technology: A New World of Nursing Homes," J. Gerontol. Geriatr. Res., vol. 2(2), 2013. https://doi.org/10.4172/2167-7182.1000122

[29] V. Freedman, M. Calkins, and K. Van Haitsma, "An exploratory study of barriers to implementing technology in U.S. residential long-term care settings," Gerontechnology, vol. 4(2), 2005. https://doi.org/10.4017/gt.2005.04.02.004.00

[30] K. C. Bezboruah, D. Paulson, and J. Smith, "Management attitudes and technology adoption in long-term care facilities," J. Health Organ. Manag., vol. 28(3), pp. 344-365, 2014. https://doi.org/10.1108/JHOM-11-2011-0118

[31] P. Lehoux, L. Richard, R. Pineault, and J. Saint-Arnaud, "Delivery of High-Tech Home Care by Hospital-Based Nursing Units in Quebec: Clinical and Technical Challenges," Nurs. Leadersh., vol. 19(1), pp. 44-55, 2006. https://doi.org/10.12927/cjnl.2006.18048

[32] C. E. Oude Weernink, L. Sweegers, L. Relou, T. J. Van Der Zijpp, and J. Van Hoof, "Lost and misplaced items and assistive devices in nursing homes: Identifying problems and technological opportunities through participatory design research," Technol. Disabil., vol. 29(3), pp. 129-140, 2017. https://doi.org/10.3233/TAD-170179

[33] Y. Lin, "Design of Nursing Home based on IoT and Blockchain Technology," Highlights Sci. Eng. Technol., vol. 24, pp. 58-61, 2022. https://doi.org/10.54097/hset.v24i.3886

[34] A. Bourbonnais, J. Rousseau, M. H. Lalonde, J. Meunier, N. Lapierre, and M. P. Gagnon, "Perceptions and needs regarding technologies in nursing homes: An exploratory study," vol. 26(3), pp. 1714-1727, 2019. https://doi.org/10.1177/1460458219889499

[35] C. Berridge, J. Halpern, and K. Levy, "Cameras on beds: The ethics of surveillance in nursing home rooms," AJOB Empir. Bioeth., vol. 10(1), pp. 55-62, 2019. https://doi.org/10.1080/23294515.2019.1568320

[36] A. Bourbonnais, J. Rousseau, M. H. Lalonde, J. Meunier, N. Lapierre, and M. P. Gagnon, "Conditions and ethical challenges that could influence the implementation of technologies in nursing homes: A qualitative study," Int. J. Older People Nurs., vol. 14(4), p. e12266, 2019. https://doi.org/10.1111/opn.12266

[37] K. Susan Sifford and A. Bharucha, "Benefits and challenges of electronic surveillance in nursing home research," Res. Gerontol. Nurs., vol. 3(1), pp. 5-10, 2010. https://doi.org/10.3928/19404921-20090706-01

[38] A. Hall, C. B. Wilson, E. Stanmore, and C. Todd, "Implementing monitoring technologies in care homes for people with dementia: A qualitative exploration using Normalization Process Theory," Int. J. Nurs. Stud., vol. 72, pp. 60-70, 2017. https://doi.org/10.1016/j.ijnurstu.2017.04.008

[39] G. L. Alexander, S. Kueakomoldej, C. Congdon, and L. Poghosyan, "A qualitative study exploring nursing home care environments where nurse practitioners work," Geriatr. Nurs. (Minneap)., vol. 50, pp. 44-51, 2023. https://doi.org/10.1016/j.gerinurse.2022.12.020

[40] N. Marjanovic, M. Jonchier, J. Guenezan, H. Delelis-Fanien, P. G. Reuter, and O. Mimoz, "Telemedicine in Nursing Home Residents Requiring a Call to an Emergency Medical Communication Center," J. Am. Med. Dir. Assoc., vol. 25(2), pp. 195-200, 2024. https://doi.org/10.1016/j.jamda.2023.09.019

Computer Science and Interdisciplinary Research Journal
Volume 1, Issue 1, ISSN: 3033-1218, DOI: 10.70862/CSIR.2024.0101-02, 2024

Computational and Implementation Problems in Option Pricing with Long Maturity

Mariyan Milev

Sofia University "St. Kliment Ohridski", Faculty of Economics and Business Administration,
125 Tsarigradsko Shosse Blvd., bl. 3, 1113 Sofia, Bulgaria

m.milev@feb.uni-sofia.bg

Abstract – In this paper we explore the significance of maturity parameter when the Fourier method is applied in option pricing by using two different analytical solutions of the Black-Scholes equation, that are expressed as infinite series of reflections and Fourier series. Even for fixed barriers contracts different solutions give the same answer when all the terms have been added up but the rate of convergence of the sum to the solution can be quite different, depending on the time to expiry (maturity), i.e. the duration of the contract. Thus, we demonstrate the significance of the maturity parameter in option contracts when analytical methods like Fourier one is applied not only in jump-diffusion models.

Keywords – Furrier transformation and coefficients, Black-Scholes PDE, maturity size parameter, discrete barrier options.

1. Introduction

Options are generally defined as a 'contract' between two parties in which one party has the right but not the obligation to do something in a future moment, usually to buy or sell shares of some commodity or stock. The essential of options is that having rights without obligations has financial value because this deal has no downside (risk) for the holder, i.e. he either makes a profit or quits without a loss. Thus, he has a potential gain, if he manages to foresee the rising or falling values of the underlying asset buying shares, or otherwise he does not exercise this option. To enter in such contract without obligations, the holder should pay some money in advance, i.e. the option price or often known as *premium*.

In the market of financial derivatives, the most important problem is the so-called option valuation problem, i.e. to compute a fair value for the option, i.e. the premium. Usually in financial literature, as a mathematical model for the movement of the asset price St under risk-neutral measure is considered a standard geometric Brownian motion diffusion process with constant coefficients r (interest rate), q (constant dividend) and σ (volatility), i.e. satisfies the following stochastic differential equation:

$$\frac{dS_t}{S_t} = (r-q)dt + \sigma dW_t,$$

where dW_t are the increments of Gauss-Wiener process. By the fundamental Itô's lemma [3] for stochastic equations, the following linear parabolic partial differential equation with non-constant coefficients is derived

$$\frac{\partial V}{\partial t} + (r-q)S\frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} - rV = 0,$$

where t is the current time and the value of the option price is $V(S,t)$ at time t before the expiry T. The option price $V(S,t)$ is specified as a solution of this equation according to the boundary conditions that are used. The last equation is known in literature as the Black-Scholes partial differential equation. It has many solutions, corresponding to all the different options contracts that can be defined with the underlying

DOI: https://doi.org/10.70862/CSIR.2024.0101-02

Corresponding author: Mariyan Milev,
Sofia University "St. Kliment Ohridski"
Email: m.milev@feb.uni-sofia.bg

Received: 7 June 2024.
Revised: 4 July 2024.
Accepted: 14 July 2024.
Published: 14 July 2024.

© 2024 Mariyan Milev. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

The article is published with Open Access at http://csirjournal.com/

variable *S* and initial and boundary conditions. A very interesting example of boundary conditions is:

Definition: *A discrete double barrier knock-out call option is an option with a payoff condition equal to max(S-K,0) which expires worthless if before the maturity the asset price has fallen outside the barrier corridor [L,U] at the prefixed monitoring dates: at these dates the option becomes zero if the asset falls out of the corridor. If one of the barriers is touched by the asset price at the prefixed dates then the option is canceled, i.e. it becomes zero, but the holder may be compensated by a rebate payment.*

Here are the initial (actual payoff) and boundary conditions of the respective Black-Scholes partial differential equation in case of a discrete double barrier knock-out call options:

$$V(S,0) = max(S-K,0)1_{[L,U]}(S),$$

$$V(S,t) \to 0 \text{ when } S \to 0 \text{ or } S \to \infty,$$

with updating the initial conditions at the monitoring dates $t_{i,}$ $i=1,\ldots,F$:

$$V(S,t_i^+) = V(S,t_i^-)1_{[L,U]}(S),$$

for $0 = t_0 < t_1 < \cdots < t_F = T$, where $1_{[L,U]}(S)$ is the indicator function, i.e. $1_{[L,U]}(S) = 1$ if $S \in [L,U]$ and $1_{[L,U]}(S) = 0$ otherwise. It should be noted that away from the monitoring dates, the option price can move on the positive real axis interval $[0,+\infty)$. If there is a *rebate payment*, i.e. the holder is compensated with a fixed amount if one of the barriers L or U is touched before maturity, then we have additional conditions at the monitoring dates t_i, $i=1,\ldots,F$: $V(L,t_i) = R_L$ and $V(U,t_i) = R_U$.

The Black-Scholes equation is usually transformed to a *diffusion* equation or a heat equation with constant coefficients:

$$\frac{\partial v}{\partial \tau} = \frac{\partial^2 V}{\partial u^2}, \tau \geq 0,$$

with the following transformations:

$$S = Ke^u, t = T - \frac{2\tau}{\sigma^2}, \rho = \frac{2(r-q)}{\sigma^2}$$

$$V(S,t) = V(Ke^{\,u}, T - \frac{2\tau}{\sigma^2}) =: K\tilde{V}(u,\tau)$$

$$\tilde{V}(u,\tau) = e^{-\frac{(\rho-1)}{2}u - \frac{(\rho+1)^2}{4}\tau - \frac{2q}{\sigma^2}\tau} v(u,\tau)$$

One important problem when transforming the Black-Scholes equation into a heat equation is what happens with the payoff condition *max(S-K,0)* applying the upper transformations? The payoff condition becomes an initial condition *v(u,0)* for the heat equation

$$V(S,t) = \max(S(T)-K,0) = Kmax(e^{\,u}-1,0)$$

And having in mind (10) we obtain *v(u,0)*, i.e.

$$v(u,0) = \max\,(e^{\frac{(\rho+1)}{2}u} - e^{\frac{(\rho-1)}{2}u}, 0).$$

In the simplest case, if a constant barrier value *B* is *continuously* applied, R_b is the rebate received if the barrier is ever breached, the boundary conditions are

$$V(B,t) = R_b,$$

$$v(u_b,\tau) = \frac{R_b}{K} e^{-\frac{(\rho-1)}{2}u - \frac{(\rho+1)^2}{4}\tau - \frac{2q}{\sigma^2}\tau},$$

$$u_b = \ln\,(B/K)$$

And according to the used boundary and payoff (initial) conditions for the different option contracts we could either solve directly the heat equation *analytically* or approximate its solution *numerically* using for example finite difference methods [6].

In absence of explicit formulas numerical methods such as the binomial trees, Monte Carlo simulations and finite difference schemes have been frequently used in option pricing [3].

However, in the last twenty year the application of the Fourier method and Fast-Fourier Transform (FFT) is preferred for the valuation of option price that could be defined as a solution of the Black-Scholes partial differential equation (PDE). The considerable computational power of the FFT and even the most fundamental advances in scientific computing are inefficient when additional financial features such as barriers, dividends and rebate payments are incorporated in option pricing. Furthermore, though the decomposition of an option price into probability elements is theoretically attractive as explained in but it is numerically undesirable due to *discontinuity of the payoffs* [2], [5]. And it comes from the definition for a double barrier knock-out option that is a plain vanilla option but it is knocked out the moment the asset reaches either the upper barrier or the lower barrier. The complicating factor is the *presence of two barriers* and *reflection technique* no longer works: we can use it to satisfy the knock-out condition on one barrier or the other, but not on both simultaneously, as the reflection principle states that the probability of hitting the barrier is exactly twice the probability of ending up below the barrier at expiration. Fortunately, using definition for barrier options that a standard vanilla contract with payoff *max(S-K,0)* is obtained if none of the barriers is touched, the solution can be

found in *two different forms*, each written as an infinite series as it can be seen in Section 2.

2. Option pricing and Fourier solutions of the Black-Scholes PDE

The first form solution for pricing a barrier call option is essentially an *infinite series of reflections*, alternately in the lower and upper barriers. In option pricing most payoffs can be represented as a sum of plain vanilla payoffs [1], [3];

The second form is that of a *Fourier series*, *an infinite sum of special solutions* which satisfy the boundary conditions, and the payoff condition is satisfied by taking the correct combination. Both solutions give the same answer when all the terms have been added up, but as it is shown by Milev in [4] that the *rate of convergence of* the sum to the solution can be quite different, depending on the *time to expiry*. Each of the above-described solutions, i.e. *infinite series of reflections* and *Fourier series one*, has the following main distinctive features, respectively:

As the *number of reflections increases*, the *'images'* get further and further away from the original strip, and their contribution to the option value becomes correspondingly *smaller*, see the analysis below. Even in terms of the heat equation the coefficients are not easy to evaluate, and it may be better to calculate them numerically using for example Simpson's rule. But in contrast to the reflection solution, the Fourier series solution converges very badly when *t* is close to maturity *T*. It is easiest to explain these solutions in the heat equation framework. For the first, we use that a reflection for Black-Scholes equation correspond to translation by a constant for the heat equation. For the Fourier series it is much easier to calculate the special solutions for the simpler heat equation.

The series solution is clearly preferable to the Fourier series version in case of *small* τ , i.e. near the expiry *T,* where the direction of the time variable t is usually changed by $t = T - \tau, \tau \in [0, T]$. But one serious drawback of the solution is that as *the number of reflections increases*, the *'images'* get further and further away from the original strip, and their contribution to the option value becomes correspondingly *smaller*. This effect is particularly observed for *larger* τ , i.e. smaller values of time *t*. In such cases the Fourier series converges much more rapidly since its latter terms

$$e^{-\frac{\pi^2\sigma^2}{2\ln^2(U/L)}n^2(T-t)}$$

decay rapidly with increasing the time to expiry in the following option pricing formula *V(S,t)*

$$V(S,t) = S^k e^{\frac{1}{2}\sigma^2\lambda(T-t)} \sum_{n=1}^{\infty} \alpha_n \sin\left(\frac{n\pi(\ln(S/L)}{\ln(U/L)}\right) e^{\left(-\frac{\pi^2\sigma^2n^2(T-t)}{2\ln^2(U/L)}\right)} \quad (1)$$

where:

$$k = -\frac{1}{2}(\rho - 1), \lambda = -\frac{1}{4}(\rho + 1)^2, \rho = \frac{2r}{\sigma^2}$$

and the Fourier coefficients α_n are:

$$\alpha_n = \frac{2}{\ln(U/L)} \int_L^U S^{-k-1} \Lambda(S) \sin\left(\frac{n\pi(\ln(S/L)}{\ln(U/L)}\right) \quad (2)$$

where $\Lambda(S)$ is the payoff condition, i.e.

$$\Lambda(S) = \max(S - K, 0) \text{ and } \Lambda(S) = \max(K - S, 0)$$

for call and put option, respectively.

3. Numerical Discussion and Significance of the Maturity Length

It is well-known that *the reflection solution* should always be used close to expiry *T*, because its terms are customized to the discontinuities of the payoff [1], [4]. Conversely, the Fourier series converges well for times *t long before expiry*, because of the rapid time-decay of the exponential coefficients. It also converges better for *larger volatility parameter* σ. Unfortunately, the Fourier series solution converges very badly when *t* is close to *T* and when σ takes low values. i.e. well-known problem in Quantitative Finance known as *low volatility problem* [3], [6], [9]. The Fourier coefficients α_n in (2) should be calculated numerically.

But the restrictions of the described above analytical solutions consist not only of the *convergence rate dependence on time to maturity* but also of the difficult generalizations for financial contracts in real life. For example, we have, $\rho = \frac{2r}{\sigma^2}$ and if there are *continuous yield dividend D* then

$$\rho = \frac{2(r - D)}{\sigma^2}.$$

But the trading in real life is done in *discrete moments*, i.e. monthly, weekly or daily, and image and Fourier series solutions should be adapted for *discrete proportional dividends* and *discrete barriers*. This problem is easier to be resolved by *semi-analytical numerical methods* [5], [7], [10] or *traditional Quantitative methods* such as the well-known Monte Carlo simulation [2] and classical finite difference schemes [2], [6].

In case of *Lévy processes* where the *size of the maturity* parameter is important some existing algorithms such as this of Milev-Tagliani [5] are extended [10]. And it is necessary more sophisticated application of the classical Fourier method such as a Fast Fourier transform [2]. As shown in Chap. 3, any application of Fourier methods leads to the evaluation of a discrete Fourier transform of length N (DFT\((N)\)). Thus, the efficient computation of DFT\((N)\) is very important. A FFT is an algorithm for computing the DFT\((N)\) which needs only *a relatively low number of arithmetic operations (N),* [8]. In case of different than the lognormal processes such as these of Merton and Kou, Phelana et. al. in [7] present the characteristic exponent of some Lévy processes and discrete Fourier and Hilbert transforms for pricing barrier options.

When other path-dependent contracts are valuated most of the existing methods for pricing Asian options are less efficient in the limit of small maturities and small volatilities [9].

4. Conclusions

The application of the analytical Fourier method is often preferred in option valuation because finding the solution to the Black-Scholes equation is a trivial mathematical task, but the numerical calculation is inefficient, which is a disadvantage in contracts with small maturities/small volatilities or multi-asset pricing. The flexibility of any analytical or numerical method in terms of achieving both efficient computational/real times makes it very competitive with the most frequently used numerical methods in Quantitative Finance that have a heuristic origin, i.e. the binomial trees and the Monte Carlo simulations. But the classic parametrical models often suffer from several limitations in term of computational power required for parametric calibration and unrealistic economical and statistical assumptions [12].

A key advantage of recent computational methods is also simplicity of their implementation that includes many financial parameters such as discrete barriers, dividends and rebate payments as well as early exercise possibilities that could be directly incorporated at any moment during the life of the option and for every value of underlying asset price. Such features are indispensable for option contracts traded in real life but including all of them together is difficult for most numerical methods.

Probably the future advances in option pricing are artificial neural networks (ANN) where unlike the traditional methods, the neural network does not care about the specific meaning of the input parameters, but the degree of its influence on the model [11]. ANN and its derivatives are just all black boxes.

Acknowledgements

This study is financed by the European Union-NextGenerationEU, through the National Recovery and Resilience Plan of the Republic of Bulgaria, project SUMMIT BG-RRP-2.004-0008-C01".

References

[1] P. Buchen, O. Konstandatos, "A New Approach to Pricing Double Barrier Options with Arbitrary Payoffs and Curved Boundaries", Appl. Math. Finance, vol. 16(6), pp. 497-515, 2009. https://doi.org/10.1080/13504860903075480

[2] P. Carr, D. Madan, "Option pricing Using the Fast Fourier Transform", Journal of Computational Finance, vol. 2(4), pp. 61-73, 2001, DOI: 10.21314/JCF.1999.043. https://doi.org/10.21314/JCF.1999.043

[3] Y. K. Kwok, Mathematical Models of Financial Derivatives, Springer-Verlag, Heidelberg, 1998.

[4] M. Milev, "Application of Fourier Series in Finance", Fundamental Sciences and Applications, vol. 16 (2), pp. 175-181, 2011, ISSN 1310-8271.

[5] M. Milev, A. Tagliani, "Numerical valuation of discrete double barrier options", Journal of Computational and Applied Mathematics, vol. 233, pp. 2468 - 2480, 2010. https://doi.org/10.1016/j.cam.2009.10.029

[6] M. Milev, A. Tagliani, "Efficient Implicit Scheme with Positivity-Preserving and Smoothing Properties", Journal of Computational and Applied Mathematics, vol. 243, pp. 1-9, 2013, ISSN: 0377-0427. https://doi.org/10.1016/j.cam.2012.09.039

[7] C. E. Phelana, D. Marazzina, G. Fusai, G. Germano, "Fluctuation identities with continuous monitoring and their application to the pricing of barrier options", European Journal of Operational Research, vol. 271(1), 16, pp. 210-223, 2018. https://doi.org/10.1016/j.ejor.2018.04.016

[8] G. Plonka, D. Potts, G. Steifdl, M. Tasche, Num. l Fourier Analysis, Birkhauser 2ed., 2023, ISBN: 978-3-031-35005-4.

[9] H. Shoshi, I. Sengupta, "Some asymptotic for short maturity Asian options", February 2023. https://doi.org/10.1080/15326349.2024.2394818

[10] X. Shuang, M. Shihu, "Pricing discrete double barrier options under Lévy processes: An extension of the method by Milev and Tagliani", Finance Research Letters, vol. 19, pp. 67-74, 2016. https://doi.org/10.1016/j.frl.2016.06.004

[11] Li. Wenda, Application of Machine Learning in Option Pricing: A Review, Advances in Economics, Business and Management Research, vol. 215, 2022, ICSSED 2022. https://doi.org/10.2991/aebmr.k.220405.035

[12] Codruț-Florin Ivașcu, Option pricing using Machine Learning, Expert Systems with Applications, vol. 163, January 2021, 113799. https://doi.org/10.1016/j.eswa.2020.113799

Computer Science and Interdisciplinary Research Journal
Volume 1, Issue 1, ISSN: 3033-1218, DOI: 10.70862/CSIR.2024.0101-03, 2024

Analysis of Suicide Risk in European Countries Using Artificial Intelligence Methods

Valentin Georgiev[1], Kostadin Yotov[2]

University of Plovdiv Paisii Hilendarski, 236, Bulgaria Blvd, 4027 Plovdiv, Bulgaria

[1] valentin_georgiev02@abv.bg, [2] kostadin_yotov@uni-plovdiv.bg

Abstract – Suicides are a phenomenon that negatively impacts society. The issues include the loss of human life, negative psychological effects on close ones, increased healthcare costs, lost productivity, and more. Suicides also create social and emotional tensions that can deteriorate the quality of life in affected communities. This is why it is necessary to study and research trends to enable prevention.

In the present study, an analysis of suicide risk in some European countries is conducted using an artificial intelligence system implemented with Neural Networks, Random Forests, Support Vector Machines, and more. The study seeks answers to the following questions: Are suicide rates increasing or decreasing? Are there differences in the number of suicides based on age and gender? Do economic factors affect the number and rate of suicides? Can we predict the number of suicides based on the given data? The machine learning methods used for this purpose are created and modeled in the Orange Data Mining environment, which also offers numerous tools for analyzing and visualizing the obtained results. It is important to note that the study of suicide risk is not equivalent to the diagnosis of suicidal disorders; however, the answers to the questions posed in the study are crucial for building and analyzing the overall picture of suicidal intentions.

Keywords – suicides, suicide risk, Europe, ANN.

DOI: https://doi.org/10.70862/CSIR.2024.0101-03

Corresponding author: Kostadin Yotov,
University of Plovdiv Paisii Hilendarski
Email: kostadin_yotov@uni-plovdiv.bg

Received: 12 July 2024.
Revised: 15 August 2024.
Accepted: 16 August 2024.
Published: 17 August 2024.

© 2024 Valentin Georgiev, Kostadin Yotov. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

The article is published with Open Access at http://csirjournal.com/

1. Significance of issues related to suicidal intentions

Issues related to suicidal intentions are of tremendous significance, as suicide is a serious social and health challenge. It is not only a matter of individual human suffering but also of our societal responsibility towards those battling severe mental disorders often leading to life-threatening actions. Suicidal intentions are frequently the result of prolonged suffering, depression, anxiety, and other mental disorders [1-3]. They have a significant impact not only on the individual but also on their immediate surroundings and the community as a whole [4]. The loss of human life due to suicide represents a grave loss for society from both a human and socio-economic perspective [5]. Therefore, combating suicidal intentions requires a comprehensive approach that includes analysis, prevention, early detection, and effective treatment of mental illnesses.

Preventing suicidal thoughts is essential not only in a medical context but also in educational and societal contexts. Individuals struggling with suicidal intentions often experience intense psychological pain and a sense of hopelessness. Depressive states are among the common symptoms, often accompanied by real physical pain [6, 7]. Hence, support and professional help are crucial for suicide prevention, with the first step being the analysis of these phenomena. Society must engage in various initiatives to raise awareness, reduce the stigma surrounding mental illnesses, and provide support to those in crisis situations. Suicidal intentions should not go unnoticed or ignored, and the issues should be analyzed with understanding. The challenges associated with suicidal intentions require joint efforts from the healthcare system, educational institutions, government bodies, and society as a whole to create better understanding and response to this serious problem.

The application of new technologies and methods such as artificial intelligence can play a crucial role in

the prevention and management of suicidal intentions through better understanding and risk analysis [8-10]. Building effective systems for analyzing suicidal intentions can provide us with a real picture of the situation and help in making the right decisions to address these problems. *The aim of the presented research is to create machine learning models to analyze data on suicide rates in European countries accumulated during the period 1992-2022.* Throughout the research, various approaches were used, including Support Vector Machines (SVM), Linear Regression (LR), Decision Trees (DT), k-Nearest Neighbors (kNN), Random Forest (RF), and Artificial Neural Networks (ANNs). These were implemented using data analysis software with extensive capabilities for graphical and tabular visualization of the analysis – Orange Data Mining. This article presents the results of two methods that interpolate the data with the least error and the highest reliability in the analysis. Predictions of suicidal intentions in the studied regions were made using RF and ANNs.

2. Machine learning methods used in the study

Several different machine learning methods were used for the purposes of the study to assess suicide risk in European countries. All the models used were created in the Orange Data Mining application, which offers excellent capabilities for analyzing and visualizing the results.

2.1. Support Vector Machine

SVM is a machine learning algorithm used for both classification and regression tasks. The main idea of SVM is to find a hyperplane in a multidimensional space that separates different classes of data. However, in many cases, the classes are not linearly separable, making it difficult to construct and use such a hyperplane initially.

In SVM, so-called "induced feature spaces" are used through of kernels, which transform the data into higher-dimensional feature spaces where they can become more easily separable. The selection of kernels in Orange Data Mining can be easily done through settings window (fig. 1). The kernel can be:

- *Linear:* Used for linearly separable data.
- *Polynomial:* Suitable for more complex, nonlinear separations.
- *Radial Basis Function (RBF)*: Preferred when the data are not linearly separable.
- *Sigmoid:* Used less frequently but is an option for certain specific tasks.

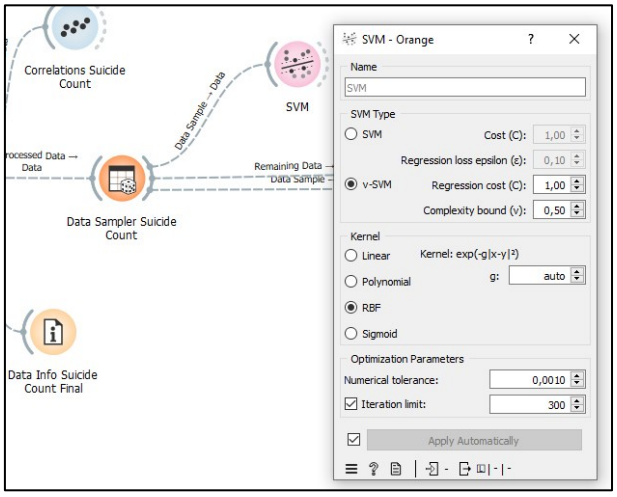


Figure 1. Setting the parameters of SVM in Orange

2.2. Random Forest

RF is a machine learning method that can be used for various tasks related to data analysis, predictions, classification, regression, and more. By combining decisions from multiple trees, RF provide high accuracy and robustness while reducing the risk of overfitting.

Each tree is trained on a different subset of data, which helps increase diversity and reduce correlation between individual trees. In the Orange, various parameters such as the number of trees, the number of attributes in each partition, the maximum depth of individual trees, etc. can be adjusted to optimize RF performance (fig. 2)

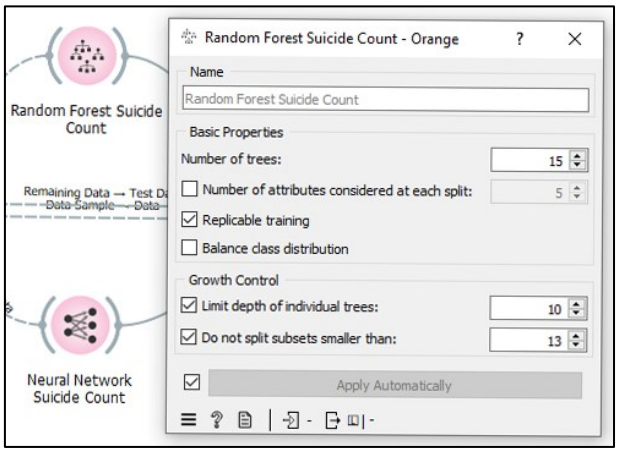


Figure 2. Setting the parameters of RF

2.3. Artificial Neural Networks

ANNs find applications in many areas such as computer vision, speech recognition, natural language processing, and more.

The Orange Data Mining software offers capabilities for constructing feedforward ANNs, with options for defining multiple hidden layers, selecting the number of neurons in each layer, choosing the

training method, the number of epochs, and more (fig. 3). The following transfer functions are available:

- Linear: $f(x) = x$;
- Logistic: $f(x) = \frac{1}{1+e^{-x}}$;
- Hyperbolic tangent: $f(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$, and
- Rectified Linear Unit (RELU): $f(x) = \max(0, x)$.

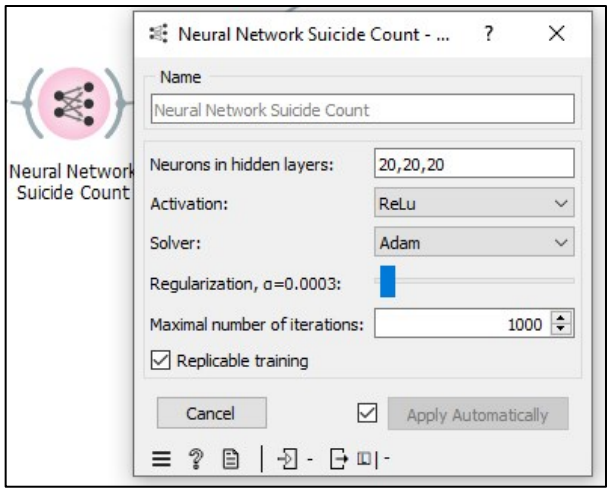


Figure 3. Setting the parameters of an ANN in Orange

The supported training methods for ANNs are:

- Limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS-B) algorithm;
- Stochastic Gradient Descent (SGD);
- Adam optimization algorithm, an extension of Stochastic Gradient Descent [11].

3. Criteria for the effectiveness of the methods used

To determine the effectiveness of the methods used for analyzing and interpolating the levels of suicide risk, the following errors were used:

- Mean Squared Error

$$(1) \quad MSE = \frac{1}{n}\sum_{i=1}^{n}(y_t - \hat{y}_t)^2$$

- Root Mean Squared Error ($RMSE$),

$$(2) \quad RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_t - \hat{y}_t)^2}$$

- Mean Absolute Error (MAE)

$$(3) \quad MAE = \frac{\sum_{t=1}^{n}|y_t - \hat{y}_t|}{n}$$

Additionally, we use the coefficient of determination (R^2) related to the correlation between the model responses and the actual data.

$$(4) \quad R^2 = 1 - \frac{\sum_{i=1}^{n}(y_t - \hat{y}_t)^2}{\sum_{i=1}^{n}(y_t - \bar{y}_t)^2}$$

In the given formulas, n is the sample size available (the number of input-output training samples), $(y_t - \hat{y}_t)^2$ are the squares of the differences during the interpolation of the respective nodes, and $(y_t - \bar{y}_t)^2$ are the squares of the residuals in the calculation of the R^2 coefficient.

4. Analysis and assessment of suicide intent levels in European countries

For the purposes of the study on suicide risk analysis, several different methods were used: LR, SVM, kNN, DT, RF, and ANN. The approaches with the best performance in data interpolation and result analysis were found to be RF and ANN.

The dataset used [12] contains nearly 120,000 records collected from the World Health Organization and the World Bank, featuring various social and economic characteristics. It includes the following columns: *RegionCode, RegionName, CountryCode, CountryName, Year, Sex, AgeGroup, Generation, SuicideCount, CauseSpecificDeathRate, DeathRatePer100K, Population, GDP, GDPPerCapita, GrossNationalIncome, GNIPerCapita, InflationRate, EmploymentPopulationRatio*. The columns *RegionCode* and *CountryCode* were removed at the beginning of the study due to duplication with other characteristics. The columns *GrossNationalIncome* and *GNIPerCapita* were removed later on as they had high correlation with GDP and GDPPerCapita.

The dataset contains three features related to the number of suicides: *SuicideCount* (the number of suicides in one year), *DeathRatePer100K* (the number of suicides per 100,000 people in one year), and *CauseSpecificDeathPercentage* (the percentage of deaths in one year due to suicide). All three features have different values depending on gender, age, country, and other factors for a given year.

Although the dataset contains data for the entire world, we have conducted this study focusing on Europe, which reduces the data volume to approximately 45,000 records. The main questions we sought to answer are:

1. *What is the trend for suicides - are they increasing or decreasing?* Figures 4 and 5 respectively show the number and rate of suicides. In these figures, the size of the symbols corresponds to the magnitude of the characteristic being examined, the color represents gender, and the type of symbol represents the age group. We can see that both the number of suicides and the suicide rate have been decreasing over the years.

2. *Is there a correlation between age and gender on one hand, and the number and rate of suicides on the other?* Tables 1 and 2 provide descriptive statistics for the data from 2015 to 2022. Both the figures and Tables 1 and 2 show clear answers to these questions. Men tend to commit suicide more frequently, both in terms of quantity and rate. Regarding age, the highest number of suicides is observed among people aged 25 to 54, but this age group has the lowest rate. For the population over 55 years old, we observe a lower number of suicides but the highest rate. The youngest age group has the lowest number of suicides but ranks second in terms of the suicide rate.

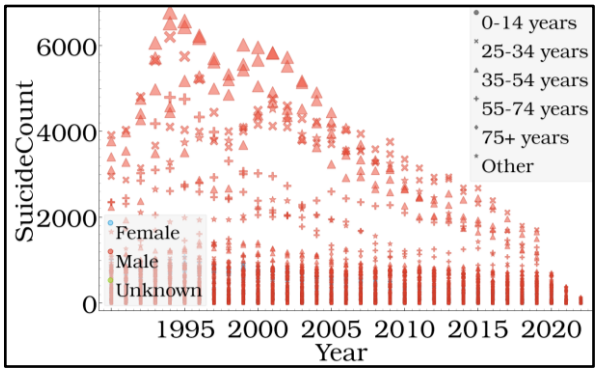


Figure 4. Number of suicides in Europe

Table 1. Statistics on the number of suicides in Europe

Number of suicides for:	Mean	Median
Men	84.34	25
Women	23.38	7
Ages 15 to 24	36.26	11
Ages 25 to 54	85.88	26
Ages over 55	69.28	26

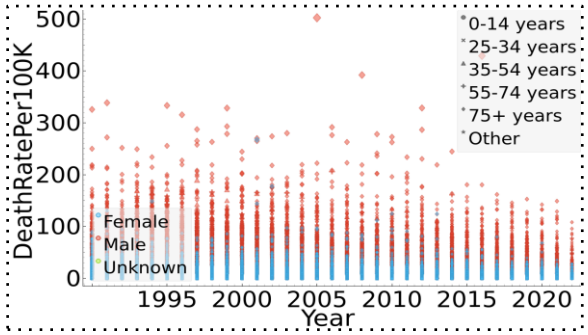


Figure 5. Suicide rate in Europe per 100,000 People

Table 2. Statistics on the suicide rate in Europe per 100,000 People

Number of suicides for:	Mean	Median
Men	20.58	16.95
Women	5.17	4.10
Ages 15 to 24	12.82	9.17
Ages 25 to 54	7.33	5.19
Ages over 55	21.74	14.00

3. *Do economic factors affect the number and rate of suicides?* Table 3 presents correlation coefficients and the ranking of characteristics (with highly correlated characteristics removed) according to RReliefF. From the data, we can conclude that there is no linear or monotonic relationship between economic factors and suicide metrics. According to their ranks, we see that the inflation rate does not have a significant impact, GDP and GDP per capita have an influence, although not the most significant, while the employment rate has a noticeable significance regarding all three suicide metrics. It should be noted that due to the low correlation values, this significance is nonlinear.

4. Can we predict the three suicide-related characteristics using machine learning methods?

To answer the last question, we selected the three characteristics *SuicideCount, DeathRatePer100K,* and *CauseSpecificDeathPercentage,* and built the models three times, each time choosing one of the characteristics as the target and excluding the others.

In preparing the data, we first excluded all characteristics that duplicated information in any way or had a high correlation with other characteristics, as previously mentioned. We removed all records with missing values due to the large volume of data and normalized the predictor data to the range [-1,1]. Finally, we split the remaining data into training and testing sets in a 70/30 ratio.

For the analysis of suicide risk, we used several different methods: LR, SVM, RF, Gradient Boosting, and ANN. For training and parameter tuning, we used cross-validation, splitting the test data into 5 parts.

The approaches with the best performance in data interpolation and result analysis were found to be RF and ANNs. The parameters of individual machine learning models are determined using the features provided by the software, which are specific to each widget. In order to achieve the necessary accuracy when using neural networks, experiments were conducted by changing and adjusting the following parameters offered by the Orange Data Mining software:

- Neurons per hidden layer**:** defined as the i-th element represents the number of neurons in the i-th hidden layer;
- Activation function for the hidden layer;
- Solver for weight optimization;
- L2 penalty (regularization term) parameter;
- Maximum number of iterations;

The main parameters of the RF used for model tuning are as follows:

- Number of trees;
- Number of trees considered at each split;
- Replicable training;

- Balance class distribution;
- Growth control: Limit depth of individual trees;

In our case, the most suitable model turned out to be an artificial neural network with three hidden layers, each containing 20 neurons. All neurons in the hidden layers are equipped with the ReLU transfer function, and the network is trained using the Adam method. The optimal random forest contains 15 trees, and the limit depth of individual trees is 10.

The results of these models are shown in Table 3. As we can see, the results are satisfactory for the models predicting the number and rate of suicides. The models for the third characteristic, *CauseSpecificDeathPercentage*, perform noticeably worse. This is understandable, as this metric is significantly more complex in nature than the others. It turns out that with the information available to us, we cannot predict what percentage of deaths in a given year are due to suicide when limited to a specific country, age group, and gender.

Table 3. Correlation and rank according to RReliefF. The first value is Pearson correlation, the second is Spearman correlation, and the third is the rank according to RReliefF

	SuicideCount			**DeathRatePer100K**			**CauseSpecific DeathPercentage**		
GDP	+0.189	+0.367	5	-0.079	+0.008	6	+0.006	+0.106	5
GDPPerCapita	-0.096	-0.043	4	-0.124	-0.065	5	+0.164	+0.094	3
InflationRate	+0.049	+0.059	8	+0.063	+0.094	8	-0.022	-0.040	7
Employment PopulationRatio	+0.017	-0.011	3	-0.015	+0.028	4	+0.124	+0.069	2

Table 4. Results of the best models on the test set
Abbreviations used: SC – the target characteristic is SuicideCount; DR – the target characteristic is DeathRatePer100K; CSDP – the target characteristic is CauseSpecificDeathPercentage

№	**Model**	**RMSE**	**MAE**	R^2
1	SC ANN	79.579	20.737	0.931
2	SC RF	86.422	17.842	0.919
3	DR ANN	9.191	4.450	0.855
4	DR RF	9.554	4.594	0.843
5	CSDP ANN	4.585	2.213	0.686
6	CSDP RF	4.742	2.297	0.664

5. Conclusion

In this study, we examined the socially significant issue of suicides. We used social and economic data for Europe, covering the period from 1992 to 2022. We found that the number and rate of suicides are decreasing. Men are more prone to suicide compared to women. The highest number of suicides is observed in the middle-aged population, but the rate is the lowest. The highest rate is observed in the elderly population. For all three suicide metrics, there is no linear or monotonic relationship. The most significant characteristic is the employment rate, followed by GDP and GDP per capita. The inflation rate is relatively insignificant.

Using machine learning methods, we built various models. The most successful ones, Artificial neural networks and Random Forests, performed satisfactorily in predicting the number and rate of suicides. Predicting the percentage of deaths due to suicide proved to be too complex to forecast with the available information.

Acknowledgements

This work was partly funded by the MUPD23-FMI-021 project of the Research Fund of the University of Plovdiv "Paisii Hilendarski".

Data Source

The original data used in the study are openly available at the websites of the World Health Organization and the World Bank, as follows:
https://www.who.int/health-topics/suicide
https://data.worldbank.org

References

[1] P. Dome, Z. Rihmer and X. Gonda, "Suicide Risk in Bipolar Disorder: A Brief Review", Medicina, vol. 55(8), 403, 2019. https://doi.org/10.3390/medicina55080403

[2] K. Wetherall, K. Robb and R. O'Connor, "Social rank theory of depression: A systematic review of self-perceptions of social rank and their relationship with depressive symptoms and suicide risk", Journal of Affective Disorders, vol. 246, pp. 300-319, 2019. https://doi.org/10.1016/j.jad.2018.12.045

[3] V. Fox, Ch. Dalman, H. Dal, A.-C. Hollander, J. Kirkbride and A. Pitman, "Suicide risk in people with post-traumatic stress disorder: A cohort study of 3.1 million people in Sweden", Journal of Affective Disorders, vol. 279, pp. 609-616, 2021. https://doi.org/10.1016/j.jad.2020.10.009

[4] M. Sajan, K. Kakar and U. Majid, "Social Interactions While Grieving Suicide Loss: A Qualitative Review of 58 Studies", The Family Journal, vol. 30 (3), 401-410, 2021. https://doi.org/10.1177/10664807211052492

[5] T. Bak, K. Kardis and D. Trong, "Suicide and Society: The Sociological Approach", Clinical Social Work and Health Intervention, vol. 10 (3), pp. 63-69, 2019. https://doi.org/10.22359/cswhi_10_3_02

[6] M. Trivedi, "The Link Between Depression and Physical Symptoms", The Primary Care Companion to The Journal of Clinical Psychiatry, vol. 6, pp. 12-16, 2004. PMID: 16001092.

[7] C. Arango-Dávilaa, H. Rincón-Hoyos, "Depressive disorder, anxiety disorder and chronic pain: Multiple manifestations of a common clinical and pathophysiological core", Revista Colombiana de Psiquiatría, vol. 47(1), pp. 46-55, 2018. https://doi.org/10.1016/j.rcpeng.2017.12.003

[8] V. Kumar, K. Sznajder and S. Kumara, "Machine learning based suicide prediction and development of suicide vulnerability index for US counties", Npj Mental Health Research, vol. 1, 2022. https://doi.org/10.1038/s44184-022-00002-x

[9] M. Ansari, S. Tiwari, S. Gupta, H. Vardhan, V. Shrivastava, "Suicide Rate Prediction Using Machine Learning", International Journal of Advances in Engineering and Management, vol. 4(5), pp. 2731-2737, 2022.

[10] N. Ismail, F. Razak and S. Masrom, "Suicide Prediction Using Machine Learning Techniques and Interactive Visualization of Suicide Information", International Journal of Academic Research in Business and Social Sciences, vol. 13(5), pp. 1905 - 1917, 2023.

[11] D. Kingma, J. Ba, "Adam: A Method for Stochastic Optimization", The 3rd International Conference for Learning Representations, San Diego, May 7-9, 2015.

[12] R. Onyango, Suicide Rates & Socioeconomic Factors (1990 - 22) [Data set]. Kaggle, 2024. https://doi.org/10.34740/KAGGLE/DS/4597596

Computer Science and Interdisciplinary Research Journal
Volume 1, Issue 1, ISSN: 3033-1218, DOI: 10.70862/CSIR.2024.0101-04, 2024

Impact of Global Country Indicators on Life Expectancy

Valentin Georgiev[1], Stanka Hadzhikoleva[2], Emil Hadzhikolev[3]

University of Plovdiv Paisii Hilendarski, 236, Bulgaria Blvd, 4027 Plovdiv, Bulgaria

[1] valentin_georgiev02@abv.bg, [2] stankah@uni-plovdiv.bg, [3] hadjikolev@uni-plovdiv.bg

Abstract – Investigating the factors that influence life expectancy enables government institutions to identify measures to improve public health, pension planning, health services, etc. Understanding these factors would enable the development and implementation of policies to improve the quality of life. This article presents a study that analyzes data from 195 countries covering demographic, economic, health and educational factors to identify those that influence life expectancy. Various models were experimented with, including linear regression, random forest, gradient boosting, SVM, and artificial neural networks. Different machine learning algorithms, feature combinations, and strategies for handling missing data set values were tested in building the models. The best model uses linear regression. Reproductive characteristics, the number of individuals enrolled in higher education, and the number of general practitioners were found to have the highest correlations with life expectancy. An interesting finding is that out-of-pocket health costs were not the leading factor in determining life expectancy.

Keywords – life expectancy, linear regression, random forest, gradient boosting, artificial neural networks, SVM.

DOI: https://doi.org/10.70862/CSIR.2024.0101-04

Corresponding author: Emil Hadzhikolev,
University of Plovdiv Paisii Hilendarski
Email: hadjikolev@uni-plovdiv.bg

Received: 4 August 2024.
Revised: 24 August 2024.
Accepted: 25 August 2024.
Published: 27 August 2024.

© 2024 Valentin Georgiev, Stanka Hadzhikoleva & Emil Hadzhikolev. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

The article is published with Open Access at http://csirjournal.com/

1. Introduction

Life expectancy plays an essential role in the social and economic development of the countries. Life expectancy statistics assist health institutions in creating policies and programs for disease prevention and treatment. This information is needed by governments for demographic planning, including forecasting future needs for health and social services, formulating social security policies, and managing pension systems. High life expectancy is also an indicator of high economic well-being, which provides better living conditions and a functioning health system with access to quality health services. To improve the quality of life, it is necessary to study the factors that influence life expectancy and to create effective policies and laws.

Some of the most significant global organizations conducting research and projects in the field of health are the ***WHO (World Health Organization)*** and the ***IHME (Institute for Health Metrics and Evaluation)***. The WHO leads efforts to improve health globally. It collects and analyzes health data, and develops policies and recommendations for better health [1]. The IHME analyze different health indicators, including the life expectancy [2]. It operates the ***Global Burden of Disease (GBD)*** project, whose goal is to study the health of populations worldwide. This includes quantifying health losses across different locations and over time, to improve health systems and eliminate inequalities. The project is significant and representative, it involves over 12,000 researchers from more than 160 countries [3].

The ***World Bank***, in addition to economic financing, also finances projects and initiatives for improving health. It periodically releases studies and findings on the health status and economic development of the countries [4]. Another global organization concerned with the health and improvement of the quality of life of children worldwide is ***UNICEF***. Its efforts to reduce child

mortality include numerous activities like vaccination, nutrition, and more [5].

The work of ***The Global Fund*** also influences higher life expectancy by financing programs to combat AIDS, tuberculosis, and malaria, having saved 59 million lives over the past 20 years [6]. ***The United Nations Development Programme (UNDP)*** operates a wide range of projects focused on socio-economic development, including health projects that directly impact life expectancy [7]. Current research, statistics, and analyses concerning the health of people in the European Union are published by ***Eurostat*** [8].

Life expectancy is a complex indicator that is influenced by different groups of factors, such as:

- ***healthcare***, which includes access to quality medical services, diagnostics, prevention, vaccination programs, and hereditary diseases;
- ***socio-economic conditions***, such as income levels, education, social protection, working conditions, and workplace safety;
- ***lifestyle factors***, involving dietary habits, physical activity, smoking, alcohol and drug use, and stress levels;
- ***environment***, which covers soil, air, and water pollution, the availability of green spaces, housing conditions, and more.

Numerous studies in the scientific literature explore the relationships between these specific factors and life expectancy across various countries.

Lelieveld et al. found that ***air pollution*** is a factor that leads to lower life expectancy, notably in cases of existing cardiovascular diseases. They conclude that air pollution is comparable to ***smoking*** when it comes to their impact on life expectancy. The global average loss of life expectancy due to air pollution significantly exceeds that from ***all forms of violence*** combined, with losses of 2.9 years and 0.3 years, respectively. The share of preventable loss of life expectancy from anthropogenic air pollution, often attributed to the use of fossil fuels, is nearly two-thirds globally and up to around 80% in high-income countries [9].

The ***Energy Policy Institute at the University of Chicago***, in their research found that air pollution is the greatest external threat to human life expectancy. According to their findings, consistently reducing global ***PM2.5*** particulate matter pollution to WHO-recommended levels could add 2.3 years to the average human lifespan. The impact of PM2.5 on global life expectancy is more than three times greater than ***alcohol consumption and unsafe water***, more than five times greater than ***transport-related injuries***, and more than seven times greater than ***HIV*** [10].

In this direction is the research of Richard Burnetta and his colleagues. They created a PM2.5-mortality hazard ratio function based only on cohort studies of outdoor air pollution, covering the global exposure range. Their study indicates that PM2.5 exposure may be linked to more causes of death than the five considered by the GBD and that incorporation of risk information from other, ***non-outdoor particle sources*** leads to the underestimation of disease burden, especially at higher concentrations [11].

Jackowska et al. investigated the impact of ***lifestyle*** on life expectancy [12]. They developed a multivariate logistic model to comprehensively assess the influence of lifestyle on mortality among young and middle-aged individuals. The model used predictors such as smoking, diet, physical activity, and body mass index, while also considering the presence of chronic diseases like diabetes, arrhythmia, coronary heart disease, myocardial infarction, stroke, and malignant neoplasms. The study used data from 38,481 participants over the age of 20. The findings highlight that lifestyle is a key factor in determining life expectancy. Premature mortality can be effectively prevented through changes in daily habits, even for healthy, young individuals. The conclusion that lifestyle is both the most controllable and influential factor affecting overall health and disease resilience is supported by other studies [13]. A healthy diet, moderate and regular physical exercise, and sufficient high-quality sleep form the foundation of a healthy lifestyle and have a positive impact on life expectancy.

Fadnes et al. have conducted various studies on how ***diet*** affects life expectancy. They researched how the consistent consumption of different types of food affects longevity. The findings show that maintaining dietary changes can bring significant health benefits to people of all ages, with greater benefits observed when changes are implemented earlier in life. The researchers developed an online calculator, Food4HealthyLife, allowing individuals to explore the effects of their dietary choices on health [14]. In another study, they found that a long-term shift towards healthier eating habits could increase life expectancy by up to 10 years in the United Kingdom. The greatest benefits were achieved by consuming more whole grains, nuts, and fruits, and fewer sugary drinks and processed meats [15]. Another of their study, involving participants from China, France, Germany, Iran, Norway, the United Kingdom, and the United States aimed to evaluate how life expectancy could be extended by switching from typical national dietary patterns to diets optimized for longevity. This research considered changes in dietary habits along with height, weight, and physical activity. Different potential life expectancy values were identified for different countries, with the most significant benefits from dietary changes found in the United States [16].

Roesch et al. conducted a study on ***the differences in life expectancy between black and white populations*** in three American cities. They examined 26 categories of causes of death and found that the primary factors contributing to mortality were circulatory diseases, cancer, injuries, and diabetes, with varying significance and impact across different cities [17]. Roberts et al. also researched this topic, focusing on residents of Washington, D.C. They found that heart disease, homicides, and cancer were the most significant contributors to the life expectancy gap between black and white men. For women, the most critical factors were heart disease, cancer, and unintentional injuries [18].

The GBD study conducted in 2021 found that due to the ***COVID-19 pandemic***, life expectancy decreased by an average of 1.6 years between 2019 and 2021. The negative effects on health were influenced by accompanying non-communicable diseases, high blood sugar, drug use, and obesity.

Despite the challenges posed by current global threats, the authors project that global life expectancy will increase by 4.6 years over the next three decades [19].

Studies like those presented focus on factors directly influencing life expectancy, relying on data collected for numerous individuals. These studies provide insights into average life expectancy based on lifestyle, genetic factors, socio-economic conditions, environmental factors, and more. The primary objective of our research is to analyze the factors affecting life expectancy. Models were created using linear regression, random forest, gradient boosting, SVM, and artificial neural networks [20]. The Orange Data Mining software application was utilized [21], allowing users to create data analysis workflows by connecting various components, called "widgets", in a graphical user interface. The widgets enable visualization of data, statistical analyses, and model building. The Orange application supports multiple machine learning algorithms suitable for classification, regression, clustering, and association tasks.

2. Materials and methods

In the current study, a dataset containing 195 records with a total of 35 features for all countries in the world for the year 2023 was used [22]. It includes demographics and economic indicators, environmental factors, health and education statistics, and more. The data is collected from multiple sources, including the WHO and the World Bank. The features are presented in *Appendix A. Dataset characteristics*.

Some of the features in the dataset were not used in the models, as they were assessed to have no significant impact on life expectancy or were highly correlated with each other, as determined during the study.

In the initial phase of the study, the correlation between life expectancy and other features was examined. We used both Pearson's and Spearman's correlation coefficients to capture different aspects of the relationships between factors and *Life Expectancy*. While Pearson's correlation coefficient is suitable for detecting only linear relationships, Spearman's correlation coefficient is used to detect relationships that are linear or non-linear.

The ten features with the highest Pearson and Spearman correlations with life expectancy are presented in Tables 1 and 2, respectively. A high negative correlation was identified for factors related to reproduction rates, such as Birth Rate, Fertility Rate, Maternal Mortality Ratio, and Infant Mortality. A relatively high positive correlation was observed for Gross Tertiary Education Enrollment (%) and Physicians per Thousand. This is likely due to the trend in developed countries, where declining birth rates coincide with greater access to education and higher education opportunities. Additionally, a higher standard of living in these countries is associated with a greater availability of doctors, contributing to better overall health. In developing countries, the opposite trends are observed.

Table 1. Features with the highest Pearson correlation with Life Expectancy

Feature	Pearson Correlation coefficient
Infant Mortality	-0.915
Birth Rate	-0.866
Fertility Rate	-0.846
Maternal Mortality Ratio	-0.817
Gross Tertiary Education Enrollment (%)	+0.714
Physicians per Thousand	+0.676
Minimum Wage	+0.496
Tax Revenue (%)	+0.340
Out of Pocket Health Expenditure	-0.315
Agricultural Land (%)	-0.241

The correlation for *Minimum Wage, Tax Revenue (%), GDP, CO2 Emissions*, and *CPI* is relatively weak, which can again be explained by the differences between developed and developing countries. There is minimal correlation for *Out of Pocket Health Expenditure* (Spearman's correlation is -0.316), leading to the conclusion that higher out-of-pocket payments by patients for healthcare do not necessarily impact life expectancy.

Table 2. Features with the highest Spearman Correlation with Life Expectancy

Feature	Spearman Correlation coefficient
Infant Mortality	-0.904
Maternal Mortality Ratio	-0.857
Birth Rate	-0.837
Fertility Rate	-0.837
Physicians per Thousand	+0.784
Gross Tertiary Education Enrollment (%)	+0.767
Minimum Wage	+0.583
GDP	+0.493
Co2-Emissions	+0.442
CPI	-0.417

For conducting experiments with artificial intelligence methods, *Life Expectancy* was selected as the target feature. The data preparation process involved several steps:

- Excluding irrelevant features from the initial dataset;
- Addressing the issue of missing values;
- Scaling the data.

Initially, certain features that were not directly related to *Life Expectancy* were removed from the original dataset: *Country, Abbreviation, Armed Forces Size, Calling Code, Capital/Major City, Currency Code, Gasoline Price, Largest City*, and *Official Language*. Geographic location, described by *Latitude* and *Longitude*, was also excluded to prevent the models from overfitting to specific locations. The selection of features in this study was guided by a focus on variables that directly impact life expectancy and are widely recognized in existing literature. Geographic location, while potentially influential due to its connection to climatic conditions, was excluded to prevent the models from overfitting to specific regional characteristics. Including geographic location could have introduced multicollinearity issues, particularly since many of the socioeconomic and healthcare-related features in the dataset are themselves influenced by geographic factors. Furthermore, the inclusion of latitude and longitude might have led to the models capturing specific location-based effects rather than generalizable patterns applicable across diverse contexts.

In the initial stage, high correlations were identified between the following features:

- *Fertility Rate* and *Infant Mortality*, *Birth Rate, Maternal Mortality Ratio*
- *Population* and *Urban Population (%), CO2* Emissions
- *Consumer Price Index (CPI)* and *CPI Change (%)*

Therefore, the final dataset included *Fertility Rate, Population*, and *CPI*.

As a result, in the initial stage, the dataset was organized such that the characteristics could be grouped as follows:

- ***Demographic and Geographic Data***: *Land Area (Km2); Density (P/Km2); Population; Agricultural Land (%)*.
- ***Economic Data***: *GDP; CPI (Consumer Price Index); Minimum Wage; Tax Revenue (%); Total Tax Rate (%); Unemployment Rate (%); Labor Force Participation (%)*.
- ***Healthcare***: *Fertility Rate; Out of Pocket Health Expenditure (%); Physicians per Thousand*.
- ***Education***: *Gross Primary Education Enrollment (%); Gross Tertiary Education Enrollment (%)*.

In the second stage, a significant amount of missing data was identified for *Minimum Wage* (45) and *Tax Revenue* (%) (23). The importance of these features was assessed using Univariate Regression and RReliefF, both of which yielded relatively high values for these characteristics. To determine whether the absence of these data points would affect the models, experiments were conducted with input datasets that both included and excluded these two characteristics.

Considering the diversity of characteristics and their significance, the following approach was adopted to handle missing data: two separate models were created. In one model, records with missing values were removed, while in the other, the missing values were imputed with the mean value of the corresponding characteristic.

In the third stage of preparing the experimental dataset, the values were scaled to the range of [-1, 1] using Min-Max normalization. The main advantages of scaling include improving model performance, enhancing the efficiency of algorithms with similar scales, preventing any single characteristic from dominating, and facilitating the interpretation of results [23].

A part of the data preprocessing is illustrated in Figure 1.

To test the SVM method, an additional categorical feature – *Life Expectancy (categorical)* – was created by dividing *Life Expectancy* into four categories. Given that *Life Expectancy* ranges from $[52.8, 85.4]$, the following categories were defined:

- short: $x \in [52.8, 60)$;
- medium_l: $x \in [60, 70)$;
- medium_h: $x \in [70, 80)$,
- long: $x \in [80, 85.4]$.

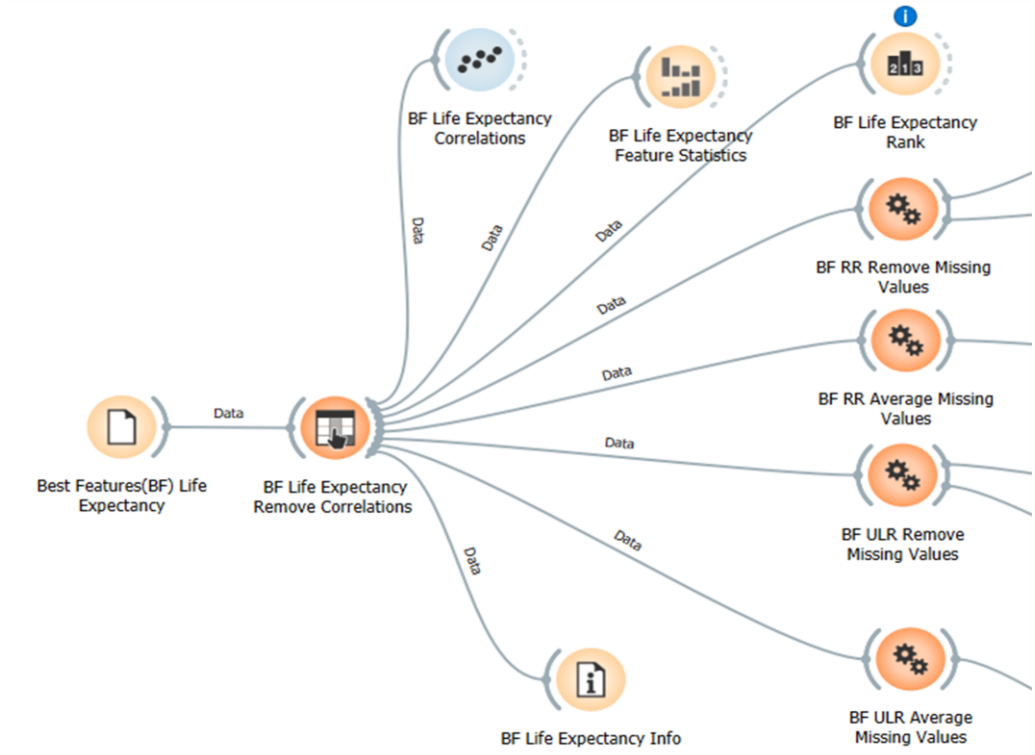


Figure 1. Data preprocessing of selected features in Orange

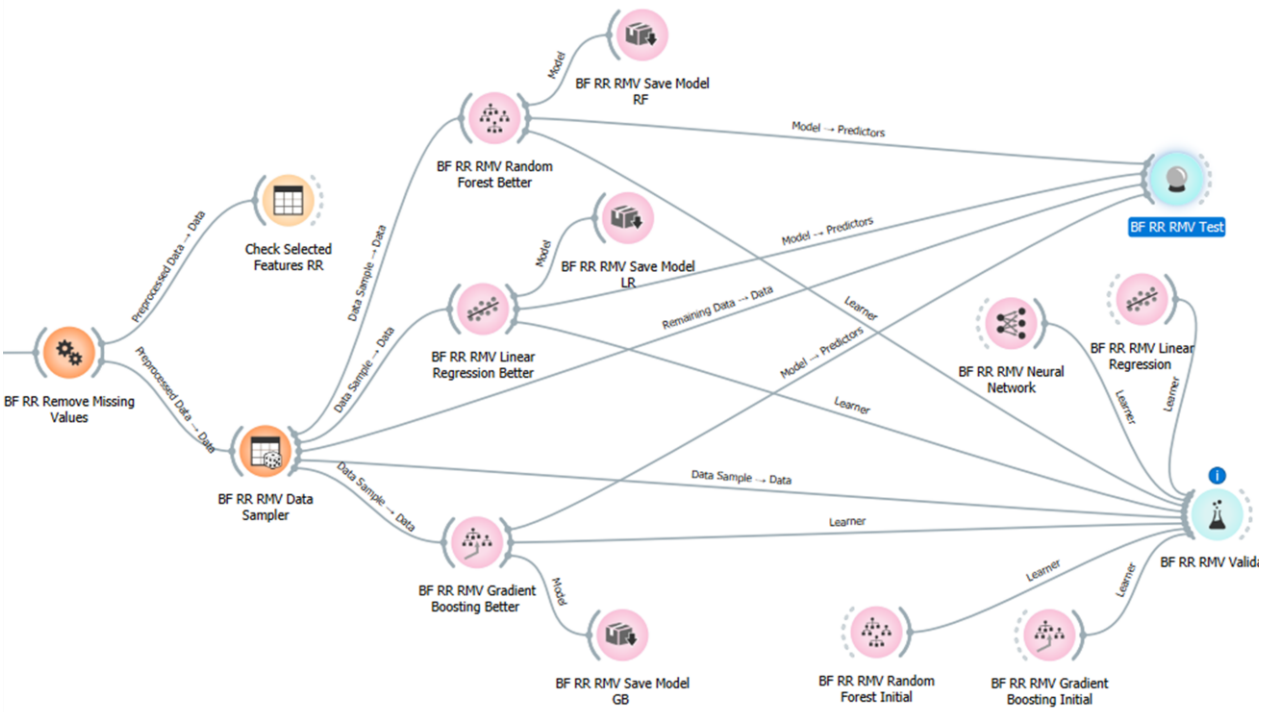


Figure 2. Diagram of the forecasting process with parts of the created models

To train the models, the data was split into two parts: training and validation (80%) and testing data (20%). We used cross-validation to select appropriate models, dividing the validation data into five non-overlapping subsets. The models were trained on all combinations of four subsets, and the predictions were then compared with the fifth subset. The accuracy of the created models was evaluated using the following metrics:

- MSE $= \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ (Mean Squared Error);
- RMSE $= \sqrt{MSE}$ (Root Mean Squared Error);
- MAE $= \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$ (Mean Absolute Error);
- MAPE $= \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|$ (Mean Absolute Percentage Error);
- $R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$ (Coefficient of Determination),

where n is the total number of samples, y_i is the actual value, $\hat{y}_i$ is the predicted value, and $\bar{y}$ is the mean of the target variable.

3. Experiments

Initial tests showed that Linear Regression, Random Forest, and Gradient Boosting outperformed SVM and Artificial Neural Networks (ANNs), which may be due to the small number of records in the dataset. Although SVM and ANN are powerful tools, they often require larger datasets to function effectively [24]. On the other hand, Linear Regression, Random Forest, and Gradient Boosting tend to perform better with smaller datasets. Due to the poor performance of SVM and ANNs, they were excluded from further experiments.

Using the three selected methods – Linear Regression, Random Forest, and Gradient Boosting – 24 models were developed with various features and approaches to handling missing data. Depending on the chosen features, the models were as follows:

- ***Models using the most relevant features according to RReliefF***: Minimum Wage, Labor Force Participation, Unemployment Rate (%), Agricultural Land (%), Tax Revenue (%), Gross Tertiary Education Enrollment (%), Fertility Rate, Forested Area (%), Out of Pocket Health Expenditure;
- ***Models using the most relevant features according to Univariate Linear Regression***: Fertility Rate, Gross Tertiary Education Enrollment (%), Physicians per Thousand, Minimum Wage, Tax Revenue (%), Out of Pocket Health Expenditure, Agricultural Land (%), Total Tax Rate (%), CPI;
- ***Models using all features***;
- ***Models using all features except Minimum Wage and Tax Revenue (%)***, which had many missing values in the original dataset.

To address the issue of missing data, two approaches were chosen: removing records or substituting with the mean value of the respective feature.

A part of the experimental model for "BF RR remove missing values" is shown in Figure 2.

Figure 3 presents the results of the evaluation of three models – *Linear Regression*, *Random Forest*, and *Gradient Boosting* – after removing the features *Minimum Wage* and *Tax Revenue* (%), as well as records with missing values. The x-axis shows the actual *Life Expectancy*, and the y-axis represents the error of each model in predicting this value. It is observed that Linear Regression performs the best in this case, with the smallest extreme values of error.

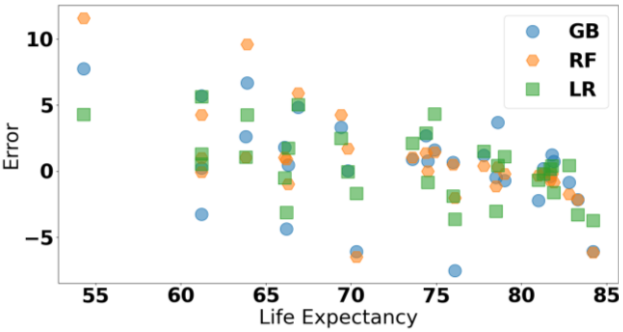


Figure 3. Results of tests on Gradient Boosting (GB), Random Forest (RF), and Linear Regression (LR)

Table 3 presents the results from ten of the most effective models.

Table 3. Selected effective models for predicting Life Expectancy

№	Model	Features	Missing values	MSE	RMSE	MAE	MAPE	R^2
1	LR	R2F	RMV	6.718	2.592	2.061	0.029	0.895
2	RF	RR	RMV	8.242	2.871	2,283	0.033	0.877
3	LR	RR	RMV	8.764	2.960	2.358	0.034	0.870
4	GB	RR	RMV	9.446	3.073	2.403	0.036	0.859
5	LR	ULR	RMV	9.827	3.135	2.505	0.037	0.858
6	RF	ULR	RMV	10.018	3.165	2.435	0.037	0.855
7	RF	ULR	AMV	8.284	2.878	2.289	0.032	0.780
8	LR	ALL	RMV	15.981	3.998	2.991	0.045	0.724
9	RF	RR	AMV	12.216	3.495	2.799	0.039	0.676
10	GB	RR	AMV	12.436	3.526	2.718	0.037	0.670

The following abbreviations are used

- ***Model*:** LR – Linear Regression; RF – Random Forest; GB – Gradient Boosting.
- ***Features***: RR – Top 10 features selected with RReliefF; ULR – Top 10 features selected with Univariate Linear Regression; ALL – All features used; R2F – All features used except Minimum Wage and Tax Revenue (%);
- ***Missing Values***: RMV – Records with missing values removed; AMV – Missing values filled with mean values of the respective feature.

The parameters of the created models are presented in Tables 4, 5, and 6.

Table 4. Parameters of the Linear Regression Models

Model №	L1 (Lasso) α
1	0.03
3	0.3
5	0.3
8	0.1

Table 5. Parameters of the Random Forest Models

Model №	Number of trees	Number of attributes at each split	Maximum depth of the trees	Minimum number of instances in leaves
2	500	5	7	5
6	425	3	10	5
7	225	3	11	2
9	150	5	9	5

Table 6. Parameters of the Gradient Boosting Models

Model №	Number of trees	Learning rate	Maximum depth of the trees	Minimum number of samples per leaf	Fraction of training instances
4	450	0.075	2	3	0.75
10	450	0.075	2	4	0.75

Models 1 through 6 are the best according to the coefficient of determination R^2, with three of them being linear regression models. It is notable that the most effective models work with datasets from which records with missing values for *Minimum Wage* and *Tax Revenue (%)* have been removed. Clearly, substituting missing values with mean values is not a good idea. This can be explained by the inherent complexity of the features, which makes the mean values unreliable substitutes. For Model 8, where records with missing values have been removed, the poor performance is likely due to the loss of data (almost 25% of the original dataset). Its alternative version (Model 1), which includes all records except for the two problematic features, is the best-performing model.

Selection of the best features using RReliefF leads to reliable predictive models that rank from 2nd to 4th place. Selection using Univariate Linear Regression yields slightly weaker results but still has high accuracy.

It should be noted that even the least effective models have an average error of 5%, which is interpreted as less than 3 years, while the best models have an average error of approximately 3.5%, which is 2.2 years.

4. Conclusion

The conducted study aimed to identify the factors affecting life expectancy. It was found that reproductive characteristics, the number of people with higher education, and the number of general practitioners have the highest correlation with life expectancy. A surprising discovery was that out-of-pocket health expenses do not significantly impact life expectancy.

Various models were experimented with, including linear regression, random forest, gradient boosting, SVM, and ANNs. Better results were achieved with linear regression, random forest, and gradient boosting. Due to the relatively small dataset, SVM and ANNs did not achieve sufficiently good accuracy. Different machine learning algorithms, feature combinations, and strategies for handling missing data were tested when building the models. The best model used linear regression on a dataset where records with missing values were removed. Selection of the best features, determined by RReliefF and Univariate Linear Regression, also led to satisfactory results. Substituting missing values with mean values resulted in less effective models. The most effective models have an average error of approximately 3.5%, which corresponds to 2.2 years in life expectancy.

Appendix A. Dataset characteristics

- ***Country***: Name of the country.
- ***Density (P/Km2)***: Population density measured in persons per square kilometer.
- ***Abbreviation***: Abbreviation or code representing the country.
- ***Agricultural Land (%)***: Percentage of land area used for agricultural purposes.
- ***Land Area (Km2)***: Total land area of the country in square kilometers.
- ***Armed Forces Size***: Size of the armed forces in the country.
- ***Birth Rate***: Number of births per 1,000 population per year.
- ***Calling Code***: International calling code for the country.
- ***Capital/Major City***: Name of the capital or major city.
- ***CO2 Emissions***: Carbon dioxide emissions in tons.
- ***CPI***: Consumer Price Index, a measure of inflation and purchasing power.
- ***CPI Change (%)***: Percentage change in the Consumer Price Index compared to the previous year.
- ***Currency Code***: Currency code used in the country.
- ***Fertility Rate***: Average number of children born to a woman during her lifetime.
- ***Forested Area (%)***: Percentage of land area covered by forests.
- ***Gasoline Price***: Price of gasoline per liter in local currency.

- ***GDP***: Gross Domestic Product, the total value of goods and services produced in the country.
- ***Gross Primary Education Enrollment (%)***: Gross enrollment ratio for primary education.
- ***Gross Tertiary Education Enrollment (%)***: Gross enrollment ratio for tertiary education.
- ***Infant Mortality***: Number of deaths per 1,000 live births before reaching one year of age.
- ***Largest City***: Name of the country's largest city.
- ***Life Expectancy***: Average number of years a newborn is expected to live.
- ***Maternal Mortality Ratio***: Number of maternal deaths per 100,000 live births.
- ***Minimum Wage***: Minimum wage level in local currency.
- ***Official Language***: Official language(s) spoken in the country.
- ***Out of Pocket Health Expenditure (%)***: Percentage of total health expenditure paid out-of-pocket by individuals.
- ***Physicians per Thousand***: Number of physicians per thousand people.
- ***Population***: Total population of the country.
- ***Population: Labor Force Participation (%)***: Percentage of the population that is part of the labor force.
- ***Tax Revenue (%)***: Tax revenue as a percentage of GDP.
- ***Total Tax Rate (%)***: Overall tax burden as a percentage of commercial profits.
- ***Unemployment Rate (%)***: Percentage of the labor force that is unemployed.
- ***Urban Population (%)***: Percentage of the population living in urban areas.
- ***Latitude***: Latitude coordinate of the country's location.
- ***Longitude***: Longitude coordinate of the country's location.

Acknowledgements

The work is funded by the SP23-FMI-008 and MUPD23-FMI-021 projects at the Research Fund of the University of Plovdiv "Paisii Hilendarski".

References

[1] World Health Organization (WHO), https://www.who.int/.

[2] Institute for Health Metrics and Evaluation (IHME). "Global Burden of Disease 2021: Findings from the GBD 2021 Study". Seattle, WA: IHME, 2024. https://www.healthdata.org/research-analysis/library/global-burden-disease-2021-findings-gbd-2021-study.

[3] Global Burden of Disease, https://www.healthdata.org/research-analysis/gbd.

[4] World Bank Group, https://www.worldbank.org/.

[5] UNICEF, https://www.unicef.org/.

[6] The Global Fund, https://www.theglobalfund.org/.

[7] United Nations Development Programme, https://www.undp.org/.

[8] Eurostat, https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Category:Health.

[9] Lelieveld, J., A. Pozzer, U. Pöschl, M. Fnais, A. Haines and T. Münzel, "Loss of life expectancy from air pollution compared to other risk factors: a worldwide perspective", Cardiovascular research, vol. 116(11), pp. 1910-1917, 2020. https://doi.org/10.1093/cvr/cvaa025

[10] Energy Policy Institute at the University of Chicago, https://global.uchicago.edu/news/air-quality-life-index-2023-update-finds-air-pollution-reduces-average-persons-life-expectancy-23.

[11] R. Burnett, H. Chen, M. Szyszkowicz, N. Fann, et al. "Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter". Proceedings of the National Academy of Sciences, vol. 115(38), pp. 9592-9597, 2018.

[12] B. Jackowska, P. Wiśniewski, T. Noiński and P. Bandosz, "Effects of lifestyle-related risk factors on life expectancy: A comprehensive model for use in early prevention of premature mortality from noncommunicable diseases", Plos one, vol. 19(3), 2024. https://doi.org/10.1371/journal.pone.0298696

[13] H. Oster and I. Chaves, "Effects of Healthy Lifestyles on Chronic Diseases: Diet, Sleep and Exercise", Nutrients, vol. 15(21), 4627, 2023. https://doi.org/10.3390/nu15214627

[14] L. T. Fadnes, J. M. Økland, Ø. A. Haaland and K. A. Johansson, "Estimating impact of food choices on life expectancy: A modeling study". PLoS medicine, vol. 19(2), 2022. https://doi.org/10.1371/journal.pmed.1003889

[15] L. T. Fadnes, C. Celis-Morales, J. M. Økland, S. Parra-Soto, et al. "Life expectancy can increase by up to 10 years following sustained shifts towards healthier diets in the United Kingdom", Nature Food, vol. 4(11), pp. 961-965, 2023. https://doi.org/10.1038/s43016-023-00868-w

[16] L. T. Fadnes, E. J. Arjmand, J. M. Økland, Celis-Morales, et al. "Life expectancy gains from dietary modifications: a comparative modeling study in 7 countries", The American Journal of Clinical Nutrition, vol. 150(1), pp. 170-177, 2024. https://doi.org/10.1016/j.ajcnut.2024.04.028

[17] P. T. Roesch, N. S. Saiyed, E. Laflamme, F. G. De Maio and M. R. Benjamins, "Life Expectancy Gaps Among Black and White Persons and Contributing Causes of Death in 3 Large US Cities, 2018-2019". JAMA Network Open, vol. 6(3), 2023, e233146. https://doi.org/10.1001/jamanetworkopen.2023.3146

[18] M. Roberts, E. N. Reither and S. Lim, "Contributors to the black-white life expectancy gap in Washington

DC", Scientific reports, vol. 10(1), 13416, 2020. https://doi.org/10.1038/s41598-020-70046-6

[19] Global Burden of Disease 2021: Findings from the GBD 2021 Study, https://www.healthdata.org/research-analysis/library/global-burden-disease-2021-findings-gbd-2021-study.

[20] J. Wang, Encyclopedia of Data Science and Machine Learning, 2022. https://doi.org/10.4018/978-1-7998-9220-5

[21] Orange Datamining, https://orangedatamining.com/.

[22] N. Elgiriyewithana, Countries of the World 2023 [Dataset]. Kaggle. https://www.kaggle.com/datasets/nelgiriyewithana/countries-of-the-world-2023/data.

[23] K. Cabello-Solorzano, I. Ortigosa de Araujo, M. Peña, L. Correia, L and A. Tallón-Ballesteros, "The impact of data normalization on the accuracy of machine learning algorithms: a comparative analysis". In International Conference on Soft Computing Models in Industrial and Environmental Applications. Cham: Springer Nature Switzerland, pp. 344-353, 2023. https://doi.org/10.1007/978-3-031-42536-3

[24] A. R. Ajiboye, R. Abdullah-Arshah, H. Qin and H. Isah-Kebbe, "Evaluating the effect of dataset size on predictive model using supervised learning technique", International Journal of Software Engineering & Computer Sciences, vol. 1(1), pp. 75-84, 2015. https://doi.org/10.15282/ijsecs.1.2015.6.0006

Computer Science and Interdisciplinary Research Journal
Volume 1, Issue 1, ISSN: 3033-1218, DOI: 10.70862/CSIR.2024.0101-05, 2024

Using Ontologies in e-Learning in Software Engineering Course

Asya Stoyanova-Doycheva[1], Georgi Cholakov[2], Todorka Glushkova[3]

University of Plovdiv Paisii Hilendarski, 236, Bulgaria Blvd, 4027 Plovdiv, Bulgaria

[1] astoyanova@uni-plovdiv.bg, [2] gcholakov@uni-plovdiv.bg, [3] glushkova@uni-plovdiv.bg

Abstract Ontologies enhance educational processes by structuring knowledge and supporting personalized learning. This study investigates their use in e-learning for generating educational content and test questions within software engineering. We discuss two ontologies: one for educational content creation and the other for test question generation. By integrating these ontologies, we improve the adaptability and effectiveness of both educational resources and assessments, offering a more comprehensive learning experience.

Keywords – ontology, software engineering, integration ontologies.

1. Introduction

Ontologies are formal, explicit specifications of a shared conceptualization [1]. They provide a structured framework for representing knowledge about a particular domain by defining a set of concepts and the relationships between them. Ontologies play a pivotal role in enhancing learning processes by offering structured, semantic frameworks for organizing and integrating knowledge. This approach enables systems to understand and interpret information in a consistent and meaningful way, allowing for better data integration, sharing, and reasoning. They are crucial for semantic web technologies, where they help in structuring and connecting data to improve its usability and accessibility.

In educational contexts, ontologies enable precise management of educational content, improving the accessibility and relevance of resources through well-defined relationships between concepts. This structured approach supports personalized learning by tailoring educational materials to individual needs and preferences. Ontologies facilitate semantic search, which enhances the accuracy of content retrieval and recommendations. Furthermore, they underpin adaptive learning systems by allowing real-time adjustments based on learner progress and feedback. Ontologies can also enhance test systems by enabling the automatic generation of test questions and assessments, leveraging structured knowledge to create contextually relevant and personalized evaluations.

The integration of ontologies into educational systems has become increasingly significant, particularly in enhancing the adaptability and personalization of learning processes. Stancin et al. [2] provide a comprehensive overview of how ontologies have been applied in education, emphasizing their role in structuring educational content and supporting personalized learning paths. Similarly, Tapia-Leon et al. [3] explore the systematic application of ontologies in higher education, demonstrating their impact on curriculum development and content delivery. In the context of software engineering education, Wille et al. [4] discuss the practical application of ontologies in creating e-learning infrastructures, which have been pivotal in the development of adaptive and context-aware educational resources.

The emergence of Large Language Models (LLMs) has introduced new dimensions to ontology research and application. Mukanova et al. [5] and Joachimiak et al. [6] highlight how LLMs can be leveraged to enrich and refine ontologies, enhancing

DOI: https://doi.org/10.70862/CSIR.2024.0101-05

Corresponding author: Asya Stoyanova-Doycheva, University of Plovdiv Paisii Hilendarski
Email: astoyanova@uni-plovdiv.bg

Received: 14 August 2024.
Revised: 29 August 2024.
Accepted: 30 August 2024.
Published: 1 September 2024.

© 2024 Asya Stoyanova-Doycheva, Georgi Cholakov & Todorka Glushkova. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

The article is published with Open Access at http://csirjournal.com/

their semantic depth and applicability in complex educational settings. These advancements underscore the evolving landscape of ontology-driven education, where LLMs play an increasingly critical role in the development and deployment of intelligent educational systems.

In education, ontologies support personalized learning by modeling educational content and learner profiles, enabling adaptive learning paths and automated content generation tailored to individual needs and preferences [7]. This personalization addresses diverse learning styles and paces, ultimately improving student engagement and outcomes [7]. Ontologies also streamline the creation of educational materials and assessments by providing a structured representation of knowledge, ensuring content relevance and alignment with learning objectives. This automation reduces the workload for educators and allows for consistent and scalable educational resources. Sapna and Balakrishnan [8] discuss the use of ontologies to support the teaching of software engineering, highlighting their role in improving educational outcomes in this field. Altan [9] further emphasizes the importance of ontologies in software engineering education programs, noting their contribution to structuring complex educational content and enhancing the learning experience.

This paper contributes to this growing body of research by focusing on the development and integration of ontologies within a Software Engineering course, specifically within the Virtual Education Space [10]. These ontologies are concentrated on two main areas: the automatic generation of test questions and the presentation of educational content within the "Software Engineering" (SE) discipline. In the final part of the article, we will demonstrate how the two ontologies, structured differently to meet the needs of two distinct environments – one for content generation in the SE discipline and the other developed for test question generation – can be integrated and used effectively in both contexts.

2. Ontology Development in Software Engineering Course.

This ontology describes the key concepts within in the lecture course on "Software Engineering" at the Faculty of Mathematics and Informatics of the University of Plovdiv [11]. The ontology was developed in Protégé-OWL. Concepts and notions are represented by means of classes. Also described are the relationships between the classes and their properties. Figure 1 shows a graph of some of the classes implemented in the ontology.

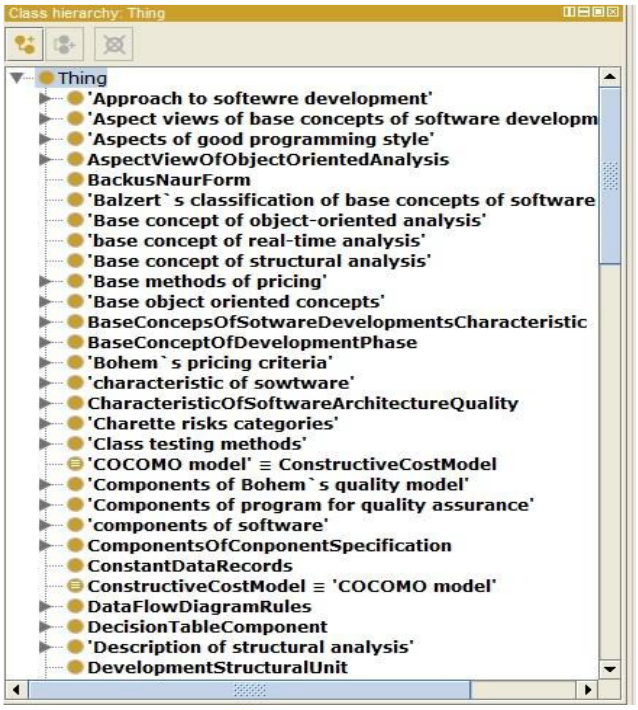


Figure 1. Hierarchy of classes in SE ontology

To be able to filter the data by certain criteria, additional information (meta-knowledge) have been added to the classes, entities, as well as constraints and properties. This have been implemented by creating two new annotations in Protégé-OWL: *importanceLevel* and *lectureNumber*. Figure 2 shows an example of the added annotation.

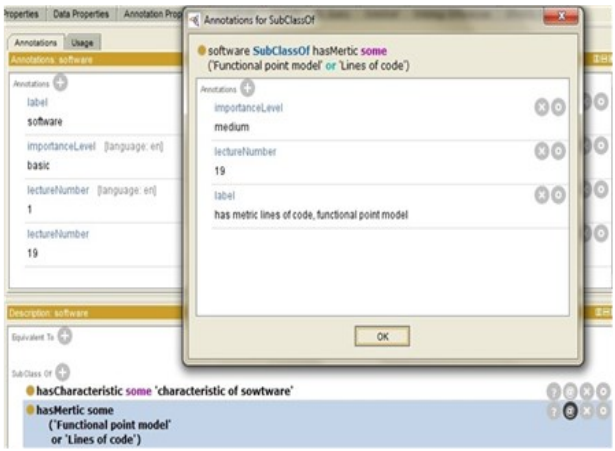


Figure 2. Annotation properties in SE ontology

The *importanceLevel* annotation takes values basic, medium and high. It predefines the importance level for each class, individuality and constraint to be displayed by the application. The second annotation added is *lectureNumber*. It represents information about the lecture number from which the concept is.

An intelligent agent processes the ontology, which aims to generate the content of the course material according to the level of the student and the relevant topics that the respective teacher would like to include in the course of study. Frequently the same subject is taught to students with different profiles. The environment for generating learning content from ontologies enables to select appropriate course topics for the corresponding level of students. Figure 3 shows a snapshot of the interface of the developed environment.

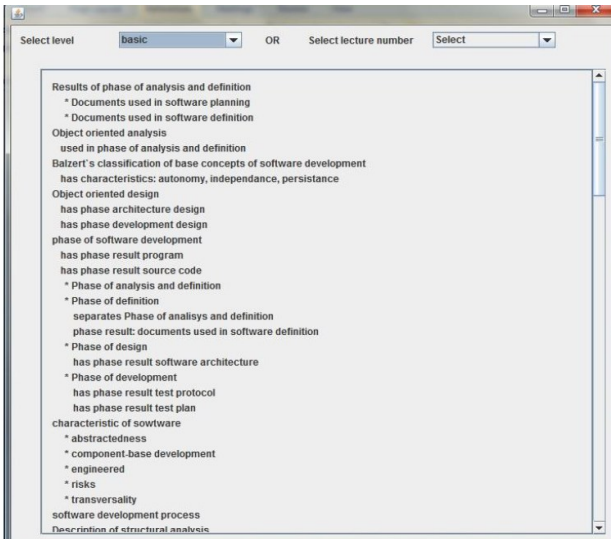


Figure 3. Generated educational content from SE ontology

The ontology's content can be updated - new classes and axioms can be added as the course is extended, or the weight of each of the topics and concepts in the ontology can be changed as decided by the instructor.

3. UML Ontology for Test Question Generation

This part of the article illustrates the structure of an ontology for test question generation. UML Ontology was developed for the syntax of Unified Modelling Language 2.4.1 using Protégé-OWL v4.3. The UML Ontology comprises basic classes that represent the fundamental concepts in UML (Fig. 4). These classes were selected to encompass the language specification and the topics covered in the Software Engineering course at the Faculty of Mathematics and Informatics. The concepts are organized in a class hierarchy, linked through object properties or data properties, forming axioms that represent the UML language syntax. This method stores information both symbolically and semantically, allowing it to be used not just as text for processing and rendering, but as interconnected concepts. This meaningful relationship between concepts enables the derivation of conclusions about the truth of statements in the field.

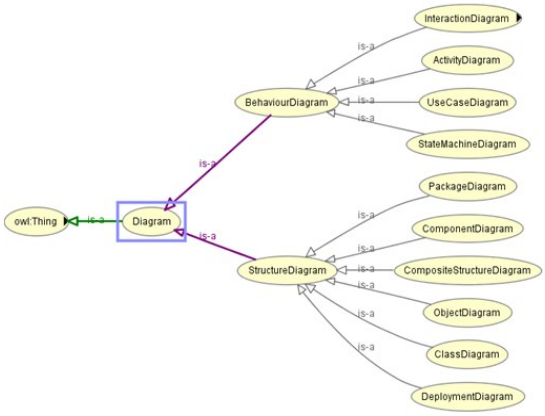


Figure 4. Hierarchy of classes in UML ontology

The class names match the construct names specified in UML 2.4.1. Some constructs are subclasses of multiple parent classes, as per the UML specification, certain elements belong to several packages, and this is accurately reflected in the ontology. The UML ontology includes two main types of properties to represent relationships between different classes: "object properties," which connect classes with other classes, and "data properties," which link classes with specific values. In Protégé-OWL, an "object property" is a property of one class that connects it to another class or links an instance of a class (individual) with another instance. Similar to classes, properties can be arranged in a hierarchy. All object properties in the Protégé-OWL hierarchy fall under topObjectProperty. Figure 5 illustrates a Protégé view of the object properties in the UML ontology.

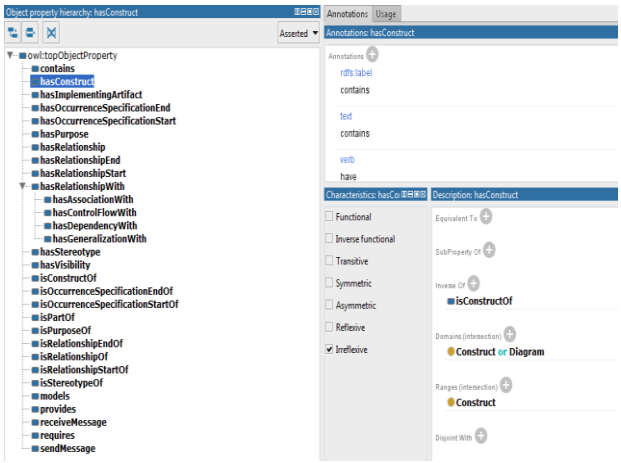


Figure 5. Object properties in UML ontology

In the ontology, classes are categorized into two types based on their conditions: primitive and defined.

Most classes in the ontology are primitive, meaning they only have necessary conditions. An individual of such a class must meet these conditions, but possessing these properties alone does not classify it as an individual of that class. Defined classes, on the other hand, have both necessary and sufficient conditions. If an individual meets these conditions, it can be conclusively classified as an individual of that specific type. In Protégé and OWL, necessary and sufficient conditions are referred to as equivalent. The defined classes in the UML ontology include Device, ExclusionEnvironment, BidirectionalAssociation, UnidirectionalAssociation, Deploy, Import, Extend, Include, Merge, Realization, Usage, CompleteMessage, FoundMessage, and LostMessage. Figure 6 illustrates the definition of the Deploy class. All other classes are primitive. To the right, the Usage view shows the constraints involving the class, with the class description below it. The first section, EquivalentTo, lists the necessary and sufficient conditions. The SubClassOf section contains the remaining necessary restrictions, and below that, SubClassOf (Anonymous Ancestor) displays the inherited restrictions from parent classes.

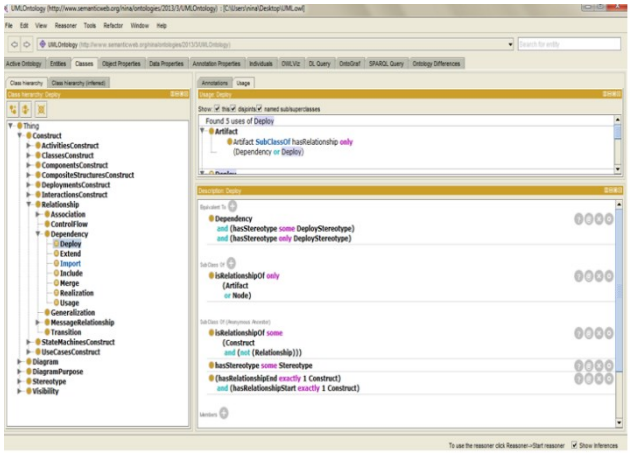


Figure 6. Constraints for class *Deploy*

Meta knowledge have been integrated into the UML ontology, significantly aiding in the generation of tests and test questions through the ontology. This type of knowledge, which provides additional information about individual elements, is known as knowledge about knowledge. In OWL, meta-knowledge is represented by annotations, which can be of various types as defined by the AnnotationProperty. The language includes several predefined annotation properties, but also allows for user-defined ones. Each ontology element can be annotated and further explained with meta-knowledge.

For the UML ontology, labels and comments are used extensively. An annotation (rdfs:comment) is added to each class, providing a textual description of the concept's essence according to the UML specification. The label (rdfs:label) provides information on how to present the ontology element to the user, eliminating the need to visualize class names directly or process them through the system. Instead, labels are used for this purpose. Each class and property in the ontology is assigned a label with a string value that meaningfully names the element in English. Since classes represent concepts, their labels are typically nouns, while properties represent relationships between classes and are best labeled with verbs (Fig. 7).

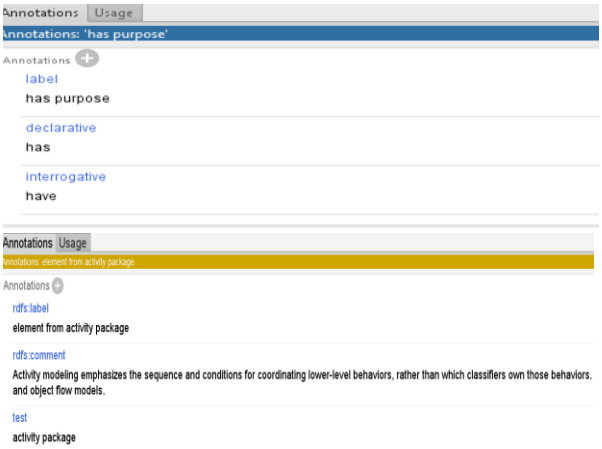


Figure 7. Annotation properties

For the current approach to automatically generate test questions, two additional annotation properties have been defined in the UML ontology: *declarative* and *interrogative* annotations. These annotations are related to the property and are intended to extend and complement their labels based on the use of the property. The values of these annotations provide the appropriate words needed for constructing different types of sentences. Essentially, these two annotations are specialized types of labels. In English, verbs change their form depending on the type of sentence being formed. Therefore, the declarative and interrogative annotations provide the label of a property for use in corresponding declarative and interrogative sentences.

The automatic question generation environment is a multi-agent system with two operative assistants and a graphical interface [12]. These assistants process the ontology to generate and validate questions. Each question is derived from an axiom, with extracted elements including a base class and its restrictions. If an axiom has multiple restrictions, one is chosen for simplicity. Properties and defining classes are also extracted. If a property is absent, the base class is a subclass of the defining class, using "is" in sentences. Otherwise, annotations are used to form declarative or

interrogative sentences. The process includes extracting annotated labels and placing them in templates for sentences. Different question types, such as open questions, follow this algorithm. The system generates questions by selecting axioms and using annotations of the base class, properties, and defining classes [13].

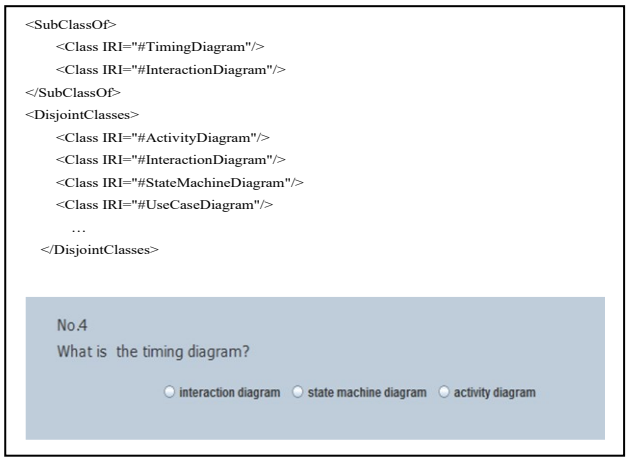


Figure 8. UML ontology axiom and generated test question from it

Example about generated test question and axiom from ontology can be seen on Figure 8.

4. Integration of Ontologies in Software Engineering.

One of the topics our team is working on is the integration of the ontologies we have already created to enrich the knowledge in software engineering and the possibility to have more capable environments for test question generation and learning content generation. To offer richer learning materials and a greater number of questions.

As presented earlier in the paper for the needs of the two environments (to generate learning content and to generate test questions), ontologies have specific elements. These specific elements are the specific annotation properties that are needed when generating e-learning content and test questions. When generating e-content from the Software Engineering Ontology, the corresponding annotation properties importanceLevel and lectureNumber need to be added to each class, and on the other hand, for the needs of the test question generation application, the corresponding annotation properties label, interrogative and declarative need to be added to the classes and properties.

Beyond altering the ontology structures to make them usable by both environments, their integration is important. For this purpose, a mapping was made between the corresponding classes and properties from the two ontologies. This may involve creating new classes and properties if necessary to reflect the combination of the two ontologies. The second important activity is the alignment of terminology. Merge differences in terminology by defining standardized terms and concepts. This is extremely important to avoid ambiguities and double interpretations. When examining the concepts in the ontologies in detail, it was concluded that the most appropriate concepts by which the two ontologies can be integrated are the different views by which an application can be represented - for this purpose, different basic concepts are used, where UML diagrams are also included. The concepts for the different core concepts exist in both ontologies (Fig. 9).

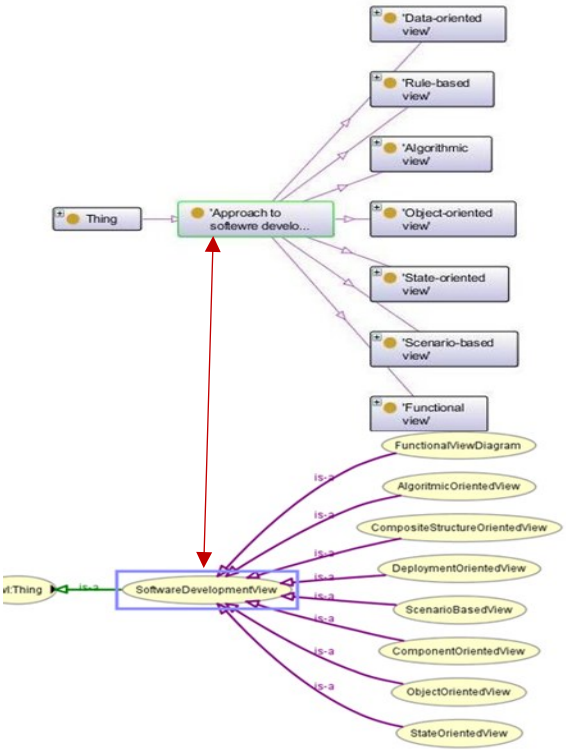


Figure 9. Common concepts in both ontologies

Each SoftwareDevelopmentView in turn contains the corresponding diagrams. For example, a class of diagrams, object diagrams, is used to model the object-oriented view in the UML ontology (Fig. 10). This knowledge is also represented in the Software Engineering Ontology. In the UML ontology these diagrams are described with the full syntax, which complements the knowledge of these diagrams with the basic notational elements in the UML syntax. On the other hand, to these two views in the Software Engineering ontology, there is knowledge about other basic concepts, but they are not part of the UML language.

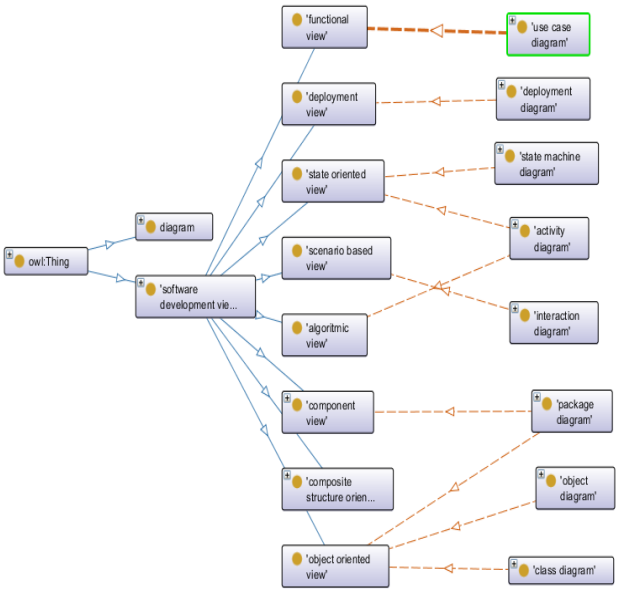


Figure 10. UML diagrams in different software views

As a final step, the classes had to be renamed to make them unique for the two ontologies. The newly created ontology basically adopted the class names from the UMLOntology because we decided to stick with the UML standard.

5. Conclusion

Integrating ontologies into an interconnected network can greatly enhance educational assessments. By combining ontologies from different disciplines such as history, science, literature, and mathematics, educators can create interdisciplinary test questions that encourage critical thinking and deeper understanding. Including ontologies related to real-world contexts, like cultural or historical events, enables the development of situational questions that challenge learners to apply their knowledge practically. Incorporating linguistic ontologies allows for assessments that test comprehension and translation across multiple languages, while integrating emotional intelligence and social behavior ontologies helps in evaluating learners' understanding of emotional responses and ethical decision-making. Expanding ontologies to cover STEM fields and technical skills can lead to complex problem-solving questions that simulate real-world challenges. Adding career-oriented ontologies can generate assessments that test both theoretical knowledge and practical skills relevant to specific professions. Incorporating environmental science and sustainability ontologies fosters questions that promote eco-conscious learning. Including cultural heritage and anthropology, ontologies enable assessments that explore cultural diversity and historical developments. Finally, integrating health, nutrition, and wellness ontologies supports the development of questions that enhance health literacy and understanding of wellness practices.

Acknowledgements

This study is financed by the project FP23-FMI-002 “Intelligent software tools and applications in research in mathematics, informatics, and teaching pedagogy” at the Plovdiv University “Paisii Hilendarski”.

References

[1] T. R. Gruber, "Toward Principles for the Design of Ontologies Used for Knowledge Sharing", International Journal of Human-Computer Studies, vol. 43, pp. 907-928, 1995. https://doi.org/10.1006/ijhc.1995.1081

[2] K. Stancin, P. Poscic and D. Jaksic, "Ontologies in education-state of the art", Education and Information Technologies, vol. 25(6), pp. 5301-5320, 2020. https://doi.org/10.1007/s10639-020-10226-z

[3] M. Tapia-Leon, A. C. Rivera, J. Chicaiza and S. Luján-Mora, "Application of ontologies in higher education: A systematic mapping study", In 2018 IEEE Global Engineering Education Conference (EDUCON), pp. 1344-1353, 2018. https://doi.org/10.1109/EDUCON.2018.8363385

[4] C. Wille, R. R. Dumke, A. Abran and J. M. Desharnais, "E-learning infrastructure for software engineering education: Steps in ontology modeling for SWEBOK", In IASTED Conf. on Software Engineering, pp. 520-525, 2004.

[5] A. Mukanova, M. Milosz, A. Dauletkaliyeva, A. Nazyrova, G. Yelibayeva, D. Kuzin and L. Kussepova, "LLM-Powered Natural Language Text Processing for Ontology Enrichment". Applied Sciences, vol. 14(13), 2024. https://doi.org/10.3390/app14135860

[6] M. P. Joachimiak, M. A. Miller, J. H. Caufield, R. Ly, N. L. Harris et al. "The Artificial Intelligence Ontology: LLM-assisted construction of AI concept hierarchies". arXiv preprint arXiv:2404.03044, 2024.

[7] Y. Wang and Y. Wang, "A Survey of Ontologies and Their Applications in e-Learning Environments". Journal of Web Engineering, vol. 20(6), pp. 1675-1720, 2021. https://doi.org/10.13052/jwe1540-9589.2061

[8] P. G. Sapna and A. Balakrishnan, "Using ontologies to aid the teaching of software engineering". In 2009 22nd Conference on Software Engineering Education and Training, pp. 224-231, 2009. https://doi.org/10.1109/CSEET.2009.32

[9] Z. Altan, "The importance of ontology in software engineering education programs". The Online Journal of Science and Technology, vol. 7(1), 119, 2017.

[10] D. Orozova, S. Stoyanov and I. Popchev, "Virtual Education Space", in International Conference, Free University of Burgas, Burgas, Bulgaria, 2013.

[11] S. Stoqyanov, A. Stoyanova - Doycheva, M. Trendafilova and E. Doychev, Software Engineering, University publishing house "Paisii Hilendarski" 2006.

[12] N. Stancheva, A. Stoyanova-Doycheva, S. Stoyanov and I. Popchev, "Automatic generation of test questions by software agents using ontologies". In: Proceedings of the IEEE 8th International Conference on Intelligent Systems, Sofia, Bulgaria, 4-6 September 2016, pp. 741-746. https://doi.org/10.1109/IS.2016.7737395

[13] N. Stancheva, A. Stoyanova-Doycheva, S. Stoyanov, I. Popchev and V. Ivanova, "A Model for Generation of Test Questions", Comptes Rendus de L'Academie Bulgare des Sciences, vol. 70(5), pp. 619-630, 2017.


https://csirjournal.com