DOI: https://doi.org/10.15276/aait.06.2023.19
UDC 004.414.32

Methodology for illness detection by data analysis techniques

Vira V. Liubchenko[2)]
ORCID: https://orcid.org/0000-0002-4611-7832; lvv@op.edu.ua. Scopus Author ID: 56667638800
Nataliia O. Komleva[1)]
ORCID: https://orcid.org/0000-0001-9627-8530; komleva@op.edu.ua. Scopus Author ID: 57191858904
Svitlana L. Zinovatna[1)]
ORCID: https://orcid.org/0000-0002-9190-6486; zinovatnaya.svetlana@op.edu.ua. Scopus Author ID: 57219779480
Jim Briggs[3)]
ORCID: https://orcid.org/0000-0003-1559-4879; jim.briggs@port.ac.uk. Scopus Author ID: 7201863846
[1)] Odessa Polytechnic National University, 1, Shevchenko Ave. Odessa, 65044, Ukraine
[2)] Hochschule für Angewandte Wissenschaften Hamburg, Fakultät Life Sciences, Ulmenliet 20
Hamburg, 21033, Germany
[3)] Centre for Healthcare Modelling and Informatics, University of Portsmouth, Portsmouth, PO1 3HE, United Kingdom

ABSTRACT

The research aims to develop information technology for identifying problematic health conditions by analyzing measurement data. The literature review highlights various approaches to medical diagnostics, including statistical and machine-learning models that predict the risk of adverse outcomes based on patient data. Developed information technology focuses on data classification and sufficiency, ensuring objective and relevant data is collected. The technology involves expert-defined rules for analysis, aiding in generating patient diagnosis candidates. The proposed information system comprises four components: data source, data storage, diagnosis module, and data sink. A comprehensive data storage structure is designed to store and manage data related to diagnoses and parameters efficiently. The rule set generation block prototype includes obtaining parameters and transforming algorithms into programming functions. A case study focuses on a diagnostic tool for assessing PTSD using an internationally recognized questionnaire. Telegram bot is selected as the data source due to its anonymity, flexibility, and automated data collection capabilities. The database structure is designed to accommodate questionnaire modifications and continue data collection. The implemented analytical system effectively categorizes individuals' states based on their responses. Overall, the research demonstrates the potential of information technology and the proposed information system to provide effective and user-friendly health diagnostics, aiding in timely medical interventions and improving population well-being.

Keywords: Health Monitoring; data analysis; diagnostics; information technology; analytical system; Telegram bot

***For citation*: Liubchenko V. V., Komleva N. O., Zinovatna S. L., Briggs Jim. "Methodology for illness detection by data analysis techniques". *Applied Aspects of Information Technology.* 2023; Vol.6 No.3: 273–285. DOI: https://doi.org/10.15276/aait.06.2023.19**

INTRODUCTION

Recent medical statistics indicate a rising occurrence of diseases at younger ages and among different populations. Furthermore, new factors such as the COVID-19 pandemic and the persistent stress of war, particularly affecting Ukrainians, contribute to deteriorating health. This situation gives rise to two significant challenges. Firstly, the existing medical system often struggles to cope with the mounting pressure. Secondly, new demographic groups, primarily comprised of young, active, and capable individuals, tend to refrain from seeking medical assistance, resulting in delayed access to timely healthcare. These circumstances significantly threaten the population's well-being and affect the country's economic development.

A potential solution is developing a user-friendly mobile application that serves as prompt, raising concerns and suggesting contacting a doctor. With the advent of various devices like fitness trackers, oximeters, glucometers and more, which gather data about an individual's health; this information can be utilized to create such an application. This application can operate discreetly in the background, minimizing any individual inconvenience while ensuring continuous monitoring and prompt response to changes in the individual's health condition.

This research aims to create an information technology that can detect an individual's problematic health state by analyzing measurements of various state parameters.

To accomplish this goal, the following steps need to be taken:

– conduct a comprehensive analysis of existing research using data analysis technologies in medical diagnostics;

– develop an information technology to assess an individual's condition by generalizing the results of measuring specific state parameters; create a prototype information system to implement the developed information technology;

– apply the developed information technology to identify problematic situations within the context of a specific disease and evaluate the outcomes of its application.

LITERATURE REVIEW

Utilizing digital technologies in medical diagnostics is not novel, and numerous researchers have achieved promising outcomes in this field.

Sheikh et al [1] describe the problems and challenges of using information technologies in the healthcare system: mobile applications, data compatibility between different systems, protection of confidential data, and expansion of data analysis capabilities. The principles of introducing health education systems in the example of the UK are considered. At the same time, the authors conclude that the available data contribute to the development of artificial intelligence; obviously only once ethical issues have been addressed.

Kumar et al [2] describe the use of artificial intelligence for diagnostics in healthcare and the benefits of finding the most effective methods for establishing a diagnosis; however, individual diseases are considered.

In Mangus et al [3], the issue of patient participation in the diagnosis is considered, for which information technologies can be used, from a test messenger to a virtual environment, which can help in obtaining information from the patient; however, the authors describe the features of using various applications specifically for emergency departments.

Developers from different countries, including Ukraine, offer their systems for work in healthcare. This is, for example, information technology to improve the efficiency of rehabilitation processes in sanatorium-resort institutions, in particular, with the help of mineral water "Naftusya" from the famous resort of Truskavets [4] or a separate mobile application for monitoring health data based on the parameters specified by the user [5].

There are now so many health information systems that the World Health Organization has released a report [6] proposing a classification of health information technology elements to promote a common taxonomy for software developers. The report includes categories for a large number of areas: for clients, for health care providers, for managers of the health system or resources, and others.

One approach to developing information technology to advice clinicians and/or patients is to collect data about the patient's condition and relate it to previous patients with similar data.

An example is the National Early Warning Score (NEWS) [7]. This is intended to assist nurses and doctors in identifying and responding to patients (normally hospital in-patients) whose condition may be deteriorating. NEWS uses the patient's vital signs (pulse rate, blood pressure, temperature, etc.) to assess their risk of an adverse outcome (death, cardiac arrest, or unanticipated admission to an intensive care unit) within the 24 hours following the observation [8]. Its use was exemplified in implementing the Vitalpac software and was shown to reduce mortality [9].

By collecting data on thousands of patients for whom their outcome was known, statistical and machine learning approaches were used to develop a computational model that could be applied to calculate the risk for any other patient.

The same team has more recently produced models that predict the risk of adverse outcomes for a patient based on blood test results [10, 11] and by combining vital sign observations, blood test results, and patient history [12].

The processing of diagnostic data in different fields of medicine can be complicated for a doctor due to the need to analyze many indicators, the relationships between which can be complex, and the degree of influence on the diagnostic result can be different. It was shown by Komlevoi et al [13] that this problem is significant in pulmonologist even in the presence of objective general clinical, biochemical, and questionnaire methods used to confirm the diagnosis and allow description of the state of the bronchopulmonary system by various indicators. Therefore, the authors investigated the possibility of classifying the values of 32 biophysical indicators obtained by laser correlation spectroscopy. The value of the work lies in expanding the set of diagnostic indicators due to high-precision results of the classification of biophysical indicators, which increases the objectivity of diagnosing pulmonological diseases.

In a recent study by Pathak et al. [14], the authors explore the utilization of computers and methodologies for sentiment analysis as a practical approach for diagnosing and monitoring mental illnesses, particularly depression. The system incorporates sentiment analysis techniques and affective computing approaches. The authors employed word classifications and specific characteristics of human facial expressions as training data for the system.

Another relevant research by Jain et al. [15] focuses on analyzing diverse social media data to develop an approach for detecting depression and impaired mental health on social media platforms. The study combines machine learning algorithms for depression detection with recommendation systems to establish correlations between specific features and individuals with depression, aiming to identify continuous patterns indicative of mental health issues.

Additionally, Sankaranarayanan et al. [16] and Vardhan et al. [17] conducted studies on the application of machine learning methods for early diagnosis of diabetic patients and the development of new drugs to ensure patient safety. These studies employed data mining techniques such as FP-Growth and Apriori to uncover hidden patterns and structures within the data. The k-nearest neighbors algorithm (KNN) algorithm methodology was employed to assess the health condition and performance and create new medications using medical data.

In this context, the work of Van den Berg et al. [18] is interesting. The authors studied the detection of late-onset sepsis. They found that none of the machine learning techniques was good enough. So, they combined three techniques to get the appropriate result.

Combining different techniques looks like a promising approach. Ahmed et al. [19] compared various data mining techniques, knowledge extraction tools and software platforms for usage in a decision support system (DSS) for analysis of medical measures.

In conclusion, using digital technologies in healthcare is a rapidly evolving field with significant potential. The relevant research explored various aspects, including challenges, artificial intelligence in diagnostics, patient involvement, and international developments in health technology. It also highlighted the importance of data-driven approaches, machine learning, and interdisciplinary methods in advancing healthcare.

INFORMATION TECHNOLOGY FOR DIGITAL ILLNESS DETECTION

By examining existing studies, we have devised an information technology to identify problematic health conditions by analyzing measurement data.

As the proposed information technology focuses on data, the initial step involves defining the requirements for data. This step is critical for successfully implementing the technology since the correct requirements definition allows efficient data processing and analysis. Various medical measurements can be categorized based on distinct characteristics. The authors suggest utilizing the classification presented in Fig. 1 for this purpose.

This classification provides a systematic approach to organizing and structuring medical data, facilitating their analysis and use. The category of measured data greatly influences the data processing technology, the type of result obtained, and the level of confidence in the outcome. In the specific tasks being addressed, it can be determined that the measurements will be readily attainable and easily defined. This implies that the data processing technology can be streamlined, and results can be obtained promptly.

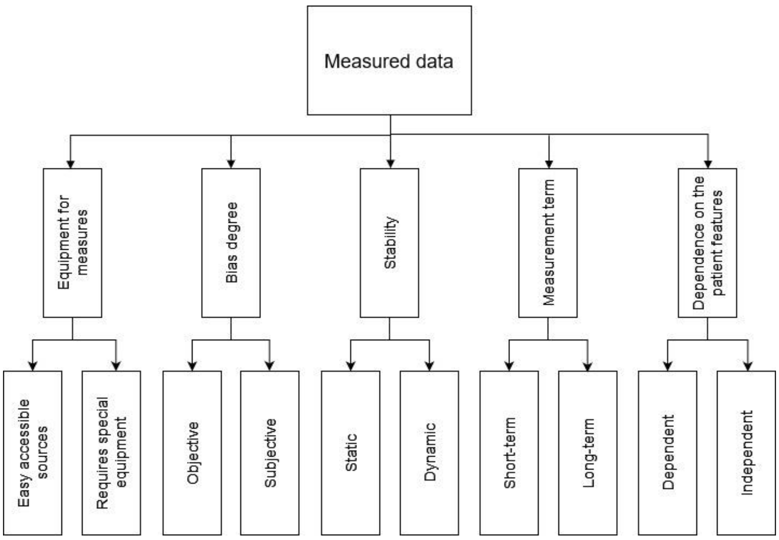

Fig. 1. **Categorization system for medical measurement**
***Source:* compiled by the authors**

We should note that some medical data has multiple timestamps: for example, for a blood test, there is the time a specimen was taken from the patient, the time the test was processed in the lab (and that could be two times: start and finish), and the time the result was reported (i.e., available to the requesting clinicians). So, a single timestamp model is simplistic. But such simplicity is appropriate for our purposes.

The input data should preferably fulfill the criteria of objectivity and sufficiency.

Objectivity is the absence of deliberate manipulation or distortion in the initial data. Ideally, the data should originate from sensors. In situations where sensors are not accessible, it is crucial to collect raw data under circumstances that discourage any distortion.

Sufficiency, on the other hand, means that the data should contain measurements on a range of attributes that enable the determination of the desired diagnosis. When studying changes in a patient's state over time, an additional requirement is that data acquisition occurs at a frequency sufficient to capture relevant information [20].

The outcome of applying information technology yields a presumed diagnosis. However, it is essential to acknowledge that this result does not constitute a conclusive medical diagnosis. Its purpose is solely to assist in making decisions regarding subsequent actions by the patient or the doctor.

Since information technology is involved in diagnosing individuals' conditions, it is crucial to process pertinent raw data accurately. As such, the involvement of experts is essential in developing information technology. These experts are responsible for determining the sufficiency of attributes required for diagnosis and establishing decisive rules for different diagnoses, along with potential recommendations.

When discussing information technology, the initial step is to establish its scope. Its purpose is to diagnose a particular disease by analyzing pertinent measurements. However, it is essential to note that the technology is not intended to diagnose all possible diseases solely based on the available measurements.

Therefore, it becomes necessary for the expert to determine the specific set of relevant parameters P that an individual can measure:

$$P = f(D, E),$$

where D is the set of diagnoses, and E denotes the expert.

In addition, the expert is responsible for defining a set of rules R, which prescribe interpreting the measurement data. These rules serve as guidelines for understanding and making sense of the data collected. By establishing these rules, the expert enables information technology to analyze the measurements and effectively provide meaningful insights and interpretations.

The individual provides the most recent data on parameter measurements P. Based on the proposed classification of measures (Fig. 1), it can be inferred that, in our case, these measurements are obtained through easily accessible sources. They are short-term and can be objective (captured by sensors) and subjective (collected through surveys). Additionally, they can be either dependent or independent.

When the data is static, a single snapshot of parameters PS is sufficient to determine the individual's state. However, if the data is dynamic and the analysis involves tracking changes over time, storing the set of parameters snapshots $\{PS_1, \ldots, PS_N\}$ measured at consecutive time moments t=1, …, N becomes necessary. This allows for a comprehensive analysis of the evolving state over time.

Generating the personal state information PS_i for a specific individual on a given timestamp, denoted as rd, involves several steps (as shown in Fig. 2).

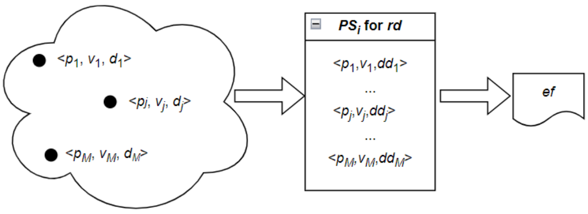

Fig. 2. **Schema of *PS***

1. Retrieval from the data warehouse: A set of indicator values $\{<p_j, v_j, d_j>\}$ is fetched from the data warehouse. Here, p_j represents the parameter, v_j corresponds to the parameter's value for the individual, and d_j signifies the date when v_j was measured. The d_j is determined as d_j=max($d_t \le rd$), where d_t represents the measurement dates for the parameter p_j.

2. Evaluation of data completeness: The obtained data is assessed for its completeness, indicated as ef. This evaluation helps determine the extent to which the data is sufficient and reliable for further processing. In this step, two assessments are conducted: evaluating the completeness of the set of indicators $|PS_i|=|P|$ and determining the relevance of the indicator values based on the correspondence between the time interval of their acquisition and the measurement categories $CT(p_j) \rightarrow dd_j = rd - d_j$, where $CT(p_j)$ is the set of categories to which indicator pj belongs.

 ISSN 2617-4316 (Print)
ISSN 2663-7723 (Online)

3. Applicability of rules: Based on the obtained data, a conclusion is drawn regarding the applicability of rules. This analysis determines whether the collected data is suitable for applying predefined rules or if further data gathering is necessary.

The measurement data provided is analyzed using the predefined rules established by the expert. The outcome of this analysis is a candidate diagnosis, which can take the form of a record indicating a problematic situation or a description of changes that have occurred since the last measurement. This diagnosis is an initial assessment based on the interpreted data, highlighting potential issues or alterations in the individual's condition. It could serve as a trigger for the individual to decide on the visit to the doctor.

Fig. 3 presents the proposed information technology as the activity diagram.

The selection of appropriate technology to select candidate diagnoses is crucial in determining the results' quality. This choice depends on the type of parameters and rules defined by the expert. Rule sets, or classification algorithms can be employed when dealing with parameters with nominal or ordinal values. Various machine learning (ML) techniques can be utilized when the parameters have quantitative values.

We want to highlight that information technology does not limit the choice of algorithms, software, and hardware. In fact, it is an abstract framework that needs to be adapted to the specific task at hand.

INFORMATION SYSTEM DESIGN AND IMPLEMENTATION

For information technology to be effectively utilized in real-life scenarios, it must be implemented as an information system. This section outlines the design of such a system, detailing its structure and functionality. The information system is comprised of four key components:

1. Data Source: This component handles the acquisition of data, which can be either objective or subjective. Suppose the data source is objective, such as from sensors like fitness trackers. In that case, the component provides an interface to access and gather information from these sensors. If the data source is subjective, the component offers a user-friendly interface to collect data through interactions with the user.

2. Data Storage: The data storage serves as a storage facility for various reasons, including monitoring changes in an individual's states over time and supporting model calibration. It ensures that relevant data is stored and readily available for analysis and processing.

3. Diagnosis Module: This component generates diagnoses based on data analysis. Data can be analyzed using rules defined by experts or data analysis models, including machine learning techniques. It is essential to consider that machine learning models require a training dataset for model identification. The selection of the appropriate technology can be influenced by the availability of a sufficiently large sample size, especially for neural network models, which demand ample data for successful identification.

4. Data Sink: The data sink represents the user interface where the results of analyzing measurement data are displayed. This component allows users to view the outcomes and insights derived from the data analysis process in the current development state.

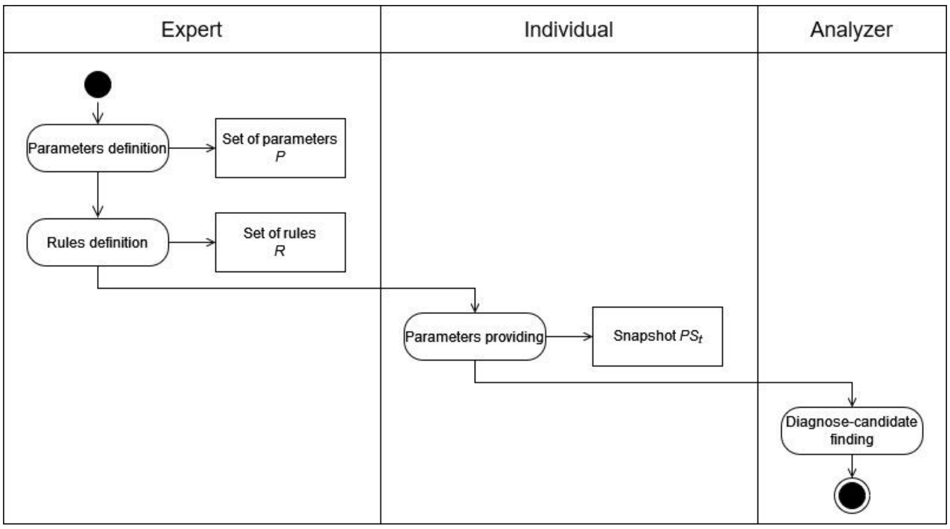

Fig. 3. **Activity diagram for information technology**
Source: **compiled by the authors**

A data storage structure (as shown in Fig. 4) is proposed to store the necessary information efficiently. This structure captures the relationships between entities within the Diagnosis-Parameter domain.

The structure includes the following relationships:

1) Diagnosis-Parameter (many to many): This relationship captures the associations between diagnoses and parameters;

2) Parameter-Class (many to many): This relationship represents the connections between parameters and their corresponding classes;

3) Parameter-Possible_answer_option (one to many): This relationship captures the possible answer options for each parameter;

4) Parameter-Individual_Features (many to many): This relationship captures normative values of the parameter based on the individual's features;

5) Individual-Feature (many to many): This relationship captures the individual's features related to a specific point in time;

6) Individual-Parameter (many to many): This relationship captures the individual's measurement results related to a specific point in time.

With this data storage structure in place, retrieving the required data slices for an individual (multiple slices) within a specific period is possible. These obtained slices can be further processed using predefined rules *R*.

Regarding the proposed structure, we would like to highlight that we have considered several crucial factors, including:

1. Multiple Diagnoses: We acknowledge that a patient may have multiple diagnoses, and the structure accommodates the possibility of capturing and managing multiple diagnostic outcomes for a single individual.

2. Dynamic Diagnoses: The system accounts for the dynamic nature of diagnoses. It allows for adding new diagnoses at any point in time and removing or ruling out existing diagnoses. This flexibility ensures that the diagnosis aspect of the system remains adaptable and up-to-date.

3. Levels of Certainty: We recognize that diagnoses can vary in certainty levels, ranging from certain to provisional or probable. The structure considers this by providing mechanisms to record and manage different levels of certainty for each diagnosis. This information can be valuable for monitoring the success of the system's diagnoses and adapting its rules (algorithms) as needed in the future.

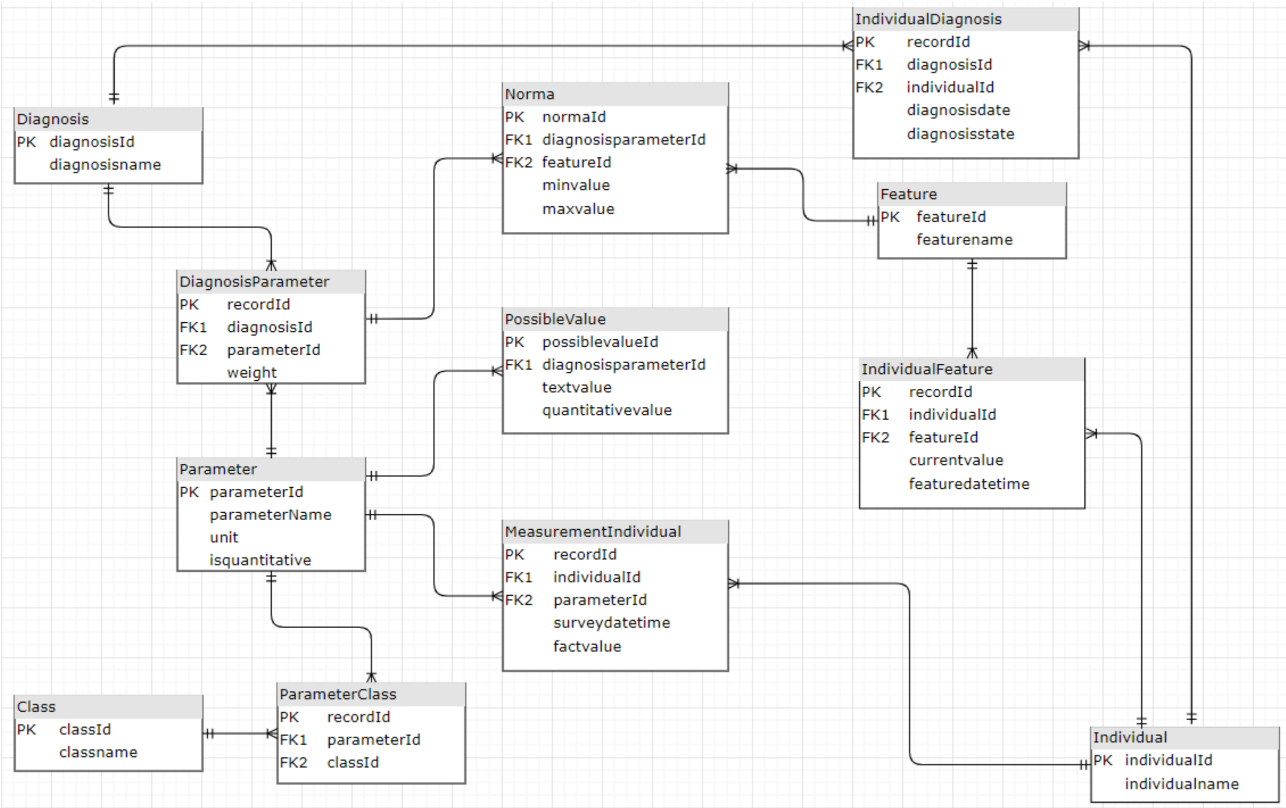

Fig. 4. **General data storage structure**
***Source:* compiled by the authors**

By incorporating these considerations into the proposed structure, we aim to build a robust and comprehensive system that can effectively handle the intricacies of diagnosis management and support continuous improvement in its diagnostic capabilities.

Based on the proposed structure, an additional structure can be created, which includes the following:

1) a subset of data specifically tailored for a particular diagnosis or a set of diagnoses;

2) additional tables that enhance the convenience of obtaining and processing specific data.

This overall database structure allows for efficient data management and facilitates the retrieval and processing of relevant information for analysis and diagnosis purposes.

The prototype of the rule set *R* generation block comprises several components (Fig. 5):

1. Acquisition of the parameter set: This component involves obtaining the parameters required for rule generation.

2. Algorithm constructor display: This component presents the algorithm constructor, which includes a list of parameters and a set of blocks for forming algorithm components. These components encompass logical constructions, mathematical operations, and comparison operations. Interacting with each constructor component creates a text fragment in the algorithm field.

3. Algorithm transformation into a programming function: This component involves converting the algorithm into a function written in a specific programming language. This process includes creating a header with a list of parameters used, an individual identifier, and a designated period. Additionally, a block is designed to initialize parameter values using methods from classes that access the database. The algorithm logic is displayed as part of this transformation process.

CASE STUDY: DIAGNOSTIC TOOL TO ASSESS POST-TRAUMATIC STRESS DISORDER

The war in Ukraine has resulted in significant losses for the nation and its people, creating many challenges that must be addressed in the present and post-war periods. One pressing issue is individuals' psychological impact from living under constant stress, threats to life, and experiencing loss. Consequently, many people are affected by post-traumatic stress disorder (PTSD), both identified cases and those that remain undiagnosed.

In Ukraine, seeking help from psychotherapists and psychoanalysts is not traditionally embraced. Therefore, developing a diagnostic tool in a simple and accessible format, allowing individuals to assess their mental state, can prove invaluable. Such a tool would provide a familiar and user-friendly means for individuals to evaluate their psychological well-being, helping to bridge the gap in seeking appropriate support and assistance [21].

An internationally recognized questionnaire [22] is utilized to gather and analyze data about the individual's condition. This questionnaire, developed by experts and extensively studied, has been chosen for this purpose. The 18-question version of the questionnaire is employed, with each question requiring a response on a Likert scale.

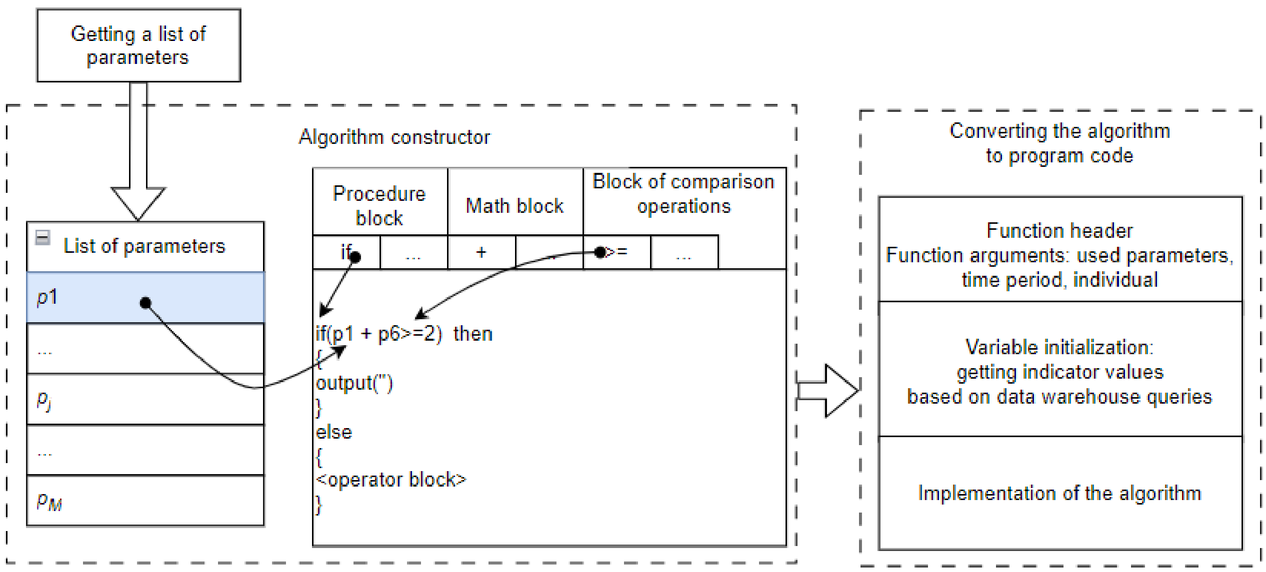

Fig. 5. **Logical schema of the rule set *R* forming block**
Source: **compiled by the authors**

Additionally, rules for processing the answers have been established and thoroughly studied. These rules support two types of results. The first type is binary and follows Boolean logic. Each response, denoted as Q_i, is labeled as True or False based on the condition $Q_i \geq 2$. Disjunctions of the corresponding question labels determine symptom labels, while conjunctions of the relevant symptom labels provide the labels for PTSD and Disturbances in Self-Organization (DSO) diagnoses.

The second type of result is numerical and employs dimensional scoring. Scores are computed for each symptom and aggregated in clusters to generate PTSD and DSO scores. This dimensional scoring approach provides a quantitative assessment of symptom severity, enabling a more comprehensive understanding of an individual's condition related to PTSD and DSO.

Supplementary descriptive-type questions were included in the questionnaire. These questions aimed to investigate the relationships between the questionnaire's environmental factors and latent factors.

In this case, the data source is the individual, and it is crucial to provide an interface that is familiar and comfortable for answering the questionnaire. To address this, a Telegram bot has been chosen as the solution. This approach offers several advantages.

1. Anonymity and data protection: Using a Telegram bot ensures the anonymity of participants and safeguards their private data. Respondents can trust that their responses and information will remain confidential, which promotes open and honest answers, especially when addressing sensitive topics.

2. Flexibility and accessibility: Telegram bots provide a flexible and accessible survey platform. Participants can complete the questionnaire conveniently, without schedule limitations or geographic location. The bot is accessible through various devices, including smartphones, tablets, and computers, making it convenient for many people to participate.

3. Ability to pause the survey: Telegram bots allow participants to interrupt the survey anytime. This feature is essential in studies involving traumatic events or psychological symptoms where participants may feel uncomfortable or stressed. The option to pause the survey empowers participants to control their involvement and protects them from any potential negative emotional consequences.

4. Automated data collection and processing: A Telegram bot enables data collection and processing automation. Participant responses are directly recorded and stored electronically, simplifying the analysis and interpretation of results while reducing the time spent on manual data entry.

By leveraging a Telegram bot, these benefits contribute to a user-friendly and efficient data collection process while prioritizing participant privacy and convenience.

Questionnaire development activity is closely connected with the development of database structure. Because we planned the questionnaire modification based on the result of the research, we should have provided flexibility in the data storage structure. Fig. 6 shows the actual database structure, which was realized on the base of the general database structure.

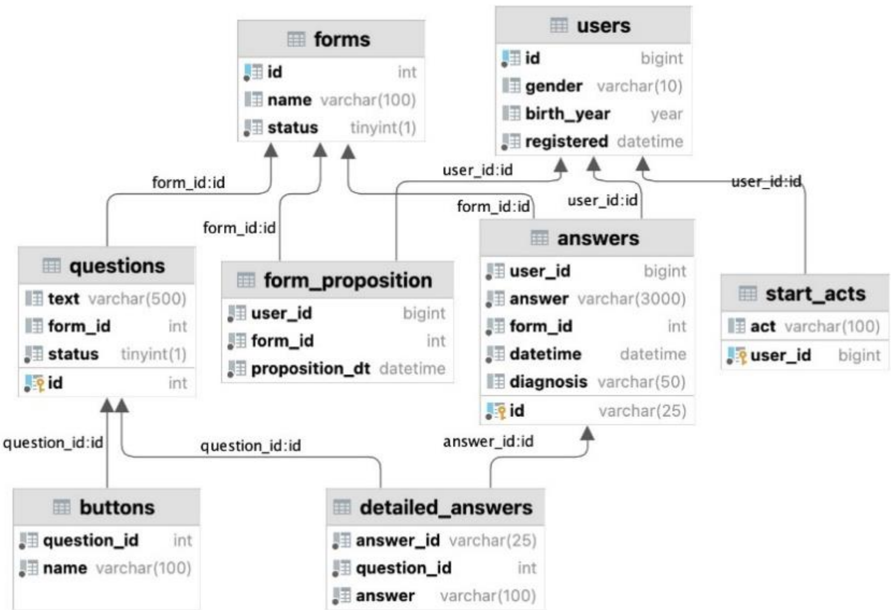

Fig. 6. **Database structure for Telegram bot**
Source: **compiled by the authors**

During the design of the Telegram bot, various correspondences were utilized to identify PTSD, including the following:

– the table “Users” represents the result of the de-normalization process involving tables “Individual”, “Feature," and “IndividualFeature”

– the table “Questions” uses data from the table “Parameter”;

– the table “Buttons” uses data from the table “PossibleValue”;

– the tables “Answers” and “Detail_answers” enable the generation of data for the table “MeasurementIndividual”.

Additionally, the following tables were added:

– the table “Forms” stores a list of bot forms;

– the table “Start_acts” is used to determine user actions when reusing the bot.

The proposed database structure made it possible to reconfigure the questionnaire “on the fly” and continue collecting data. For example, the fourth supplementary question in the first version of the questionnaire was phrased as “Do you hide during a missile attack?” Then it was decided to change the question's wording and offer five possible answers.

The implemented analytical system was used to study the ITQ properties.

The study, conducted on a non-clinical sample of 286 people aged 16 to 60, provided valuable results using the Telegram bot as an anonymous online survey.

We examined the responses provided to the bot. The findings demonstrated that the responses effectively categorize the state of individuals, as depicted in Fig. 7.

DISCUSSION AND CONCLUSIONS

In conclusion, the increasing occurrence of diseases at younger ages and among diverse populations, coupled with new challenges posed by events like the COVID-19 pandemic and ongoing conflicts in regions like Ukraine, highlights the urgent need for innovative healthcare solutions. Developing a user-friendly application that utilizes health monitoring devices and data analysis technologies presents a potential solution to address these challenges.

The proposed information technology aims to detect problematic health states by analyzing measurements of various parameters. The comprehensive literature review indicates that data analysis technologies, including machine learning, sentiment analysis, and data mining, have shown promise in various medical applications, such as early diagnosis, drug development, and mental health monitoring.

Since information technology is a framework, it is impossible to give specific measurements of these characteristics. They depend on the specific implementation, particularly the choice of algorithms used, software implementation tools and hardware on which the information system is deployed.

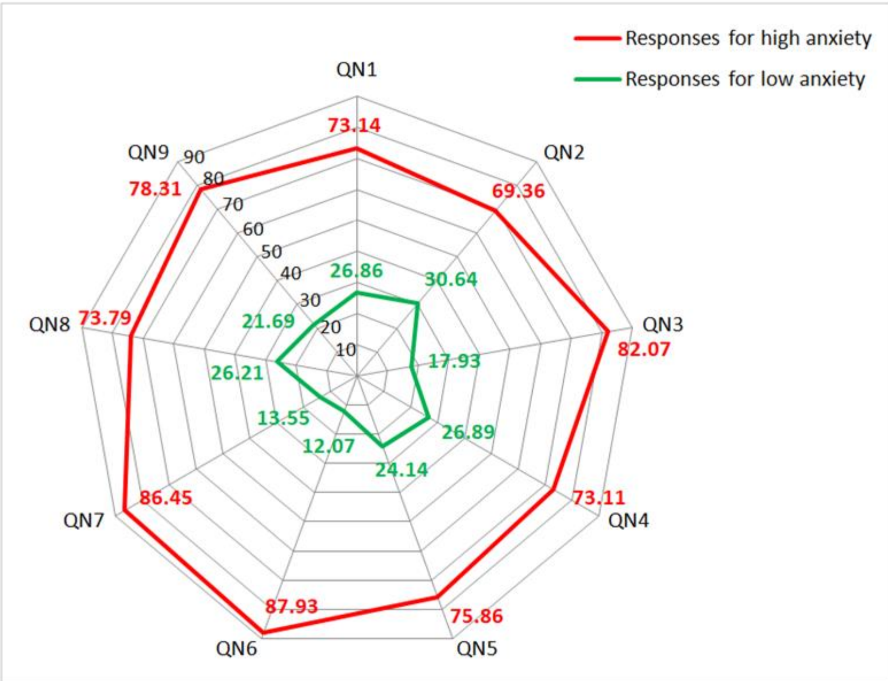

Fig. 7. **Profiles of groups with and without post-traumatic stress disorder**
Source: **compiled by the authors**

The proposed information system, composed of data source, data storage, diagnosis module, and data sink components, provides a robust framework for effectively implementing the information technology in real-life scenarios. The database structure and the prototype rule set generation block offer flexibility and efficiency in managing data and generating diagnoses.

The case study on developing a diagnostic tool to assess post-traumatic stress disorder (PTSD) using a Telegram bot showcases the practical application of the proposed approach. By leveraging a Telegram bot for data collection, the case study demonstrates the benefits of anonymity, flexibility, accessibility, and automated data processing in conducting surveys related to sensitive psychological issues. The database structure's adaptability allows for the reconfiguration of the questionnaire, enabling efficient data collection and analysis.

Overall, the research provides insights into developing and implementing information technology for digital illness detection, potentially improving healthcare access, diagnosis, and monitoring for individuals and populations. However, it is essential to note that this is a theoretical proposal. Practical implementation would require thorough testing, validation, and consideration of ethical implications to ensure the accuracy, reliability, and privacy of the healthcare data being processed.

ACKNOWLEDGMENT

This project was made possible through the UK-Ukraine twinning grants scheme, funded by Research England with the support of Universities UK International and UK Research and Innovation.

REFERENCES

1. Sheikh, A., Anderson, M., Albala, S., et al. “Health information technology and digital innovation for national learning health and care systems”. *Lancet Digit Health*. 2021; 3: e383–96, https://www.scopus.com/authid/detail.uri?authorId=7202522962. DOI: https://doi.org/10.1016/S2589-7500(21)00005-4.

2. Kumar, Y., Koul, A., Singla, R. & Ijaz, M. F. “Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda”. *J Ambient Intell Humaniz Comput*. 2023; 14(7):8459-8486, https://www.scopus.com/authid/detail.uri?authorId=57225085312. DOI: https://doi.org/10.1007/s12652-021-03612-z.

3. Mangus, C. W., Singh, H. & Mahajan, P. “Health Information technology for engaging patients in diagnostic decision making in emergency departments”. *Agency for Healthcare Research and Quality*. 2021. p. 1–15. – Available from: https://www.ahrq.gov/sites/default/files/wysiwyg/patient-safety/reports/issue-briefs/healthited-issuebrief.pdf. – [Accessed: 20 July 2022].

4. Odrekhivskyy, M., Pasichnyk, V., Kunanets, N. et al. “The use of modern information technology in medical and health institutions of truskavets resort”. *Proceedings of the 2nd International Workshop on Modern Machine Learning Technologies and Data Science*. Lviv-Shatsk: Ukraine. 2020; 1: 184–197. – Available from: https://ceur-ws.org/Vol-2631/paper14.pdf. – [Accessed: 29 July 2022], https://www.scopus.com/authid/detail.uri?authorId=57216702915.

5. Oheruk, B. & Peleshchak, R. “Information system for monitoring personal health status based on big data analysis”. *Proceedings of the 3rd International Workshop on Modern Machine Learning Technologies and Data Science*. Lviv-Shatsk: Ukraine. 2021; 1: 749–782. – Available from: https://ceur-ws.org/Vol-2917/paper44.pdf. – [Accessed: 29 July 2022]. https://www.scopus.com/authid/detail.uri?authorId=57226523591.

6. “Classification of Digital Health Interventions v 1.0”. *World Health Organization*. 2018. – Available from: https://apps.who.int/iris/bitstream/handle/10665/260480/WHO-RHR-18.06-eng.pdf. – [Accessed: 29 July, 2022].

7. “National Early Warning Score (NEWS) 2”. – Available from: https://www.rcplondon.ac.uk/projects/outputs/national-early-warning-score-news-2. – [Accessed: 28 July, 2022].

8. Prytherch, D. R., Smith, G. B., Schmidt, P. E. & Featherstone, P. I. “ViEWS – Towards a national early warning score for detecting adult in-patient deterioration”. *Resuscitation*. 2010: 81 (8): 932–937, https://www.scopus.com/authid/detail.uri?authorId=6603937995.
DOI: https://doi.org/10.1016/j.resuscitation.2010.04.014.

9. Schmidt, P. E., Meredith, P., Prytherch, D. R. et al. "Impact of introducing an electronic physiological surveillance system on hospital mortality". *BMJ Quality & Safety*. 2015; 24: 10–20, https://www.scopus.com/authid/detail.uri?authorId=14061165000. DOI: http://dx.doi.org/10.1136/bmjqs-2014-003073.

10. Jarvis, S. W., Kovacs, C., Badriyah, T., Briggs, J., Mohammed, M. A., Meredith, P. et al. "Development and validation of a decision tree early warning score based on routine laboratory test results for the discrimination of hospital mortality in emergency medical admissions". *Resuscitation*. 2013: 84 (11): 1494–1499, https://www.scopus.com/authid/detail.uri?authorId=55909787700.
DOI: https://doi.org/10.1016/j.resuscitation.2013.05.018.

11. Redfern, O. C., Pimentel, M. A.F., Prytherch, D., Meredith, P., Clifton, D. A., Tarassenko, L., Smith, G. B. & Watkinson, P. J. "Predicting in-hospital mortality and unanticipated admissions to the intensive care unit using routinely collected blood tests and vital signs: Development and validation of a multivariable model". *Resuscitation*. 2018; 133: 75–81, https://www.scopus.com/authid/detail.uri?authorId=6505781601.
DOI: https://doi.org/10.1016/j.resuscitation.2018.09.021.

12. Pimentel, M. A. F., Redfern, O. C., Malycha, J., Meredith, P., Prytherch, D., Briggs, J. et al. "Detecting Deteriorating Patients in the Hospital: Development and Validation of a Novel Scoring System". *American Journal of Respiratory and Critical Care Medicine*. 2021; 204 (1): 44–52, https://www.scopus.com/authid/detail.uri?authorId=55246176600.
DOI: https://doi.org/10.1164/rccm.202007-2700OC.

13. Komlevoi, O., Komleva, N., Liubchenko, V. & Zinovatna, S. "Biological data mining and its applications in pulmonology". *Proceedings of the 4th International Conference on Informatics & Data-Driven Medicine*. Valencia: Spain. 2021; 3038: 44–53, https://www.scopus.com/authid/detail.uri?authorId=57194525992. – Available from: http://ceur-ws.org/Vol-3038/paper5.pdf. – [Accessed: 28 July, 2022].

14. Pathak, K., Gupta, P. & Nimala, K. "Diagnosis of Depressive Disorder Based on Facial and Text-Based Features Using Effective Techniques". *Proceedings of the International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF)*. Chennai: India. 2023. p. 1–7, https://www.scopus.com/authid/detail.uri?authorId=58191053200.
DOI: https://doi.org/ICECONF57129.2023.10083832.

15. Jain, V., Chandel, D., Garg, P. & Vishwakarma, D. K. "Depression and impaired mental health analysis from social media platforms using predictive modeling techniques". *Proceedings of the 4th International Conference on IoT in Social, Mobile, Analytics and Cloud*. 2020. p. 855–860, https://www.scopus.com/authid/detail.uri?authorId=57676114200.
DOI: https://doi.org/10.1109/I-SMAC49090.2020.9243334.

16. Sankaranarayanan, S., Perumal, P. T. "Diabetic prognosis through data mining methods and techniques". *2014 International Conference on Intelligent Computing Applications*. Coimbatore: India. 2014. p. 162–166, https://www.scopus.com/authid/detail.uri?authorId=57214522230.
DOI: https://doi.org/10.1109/ICICA.2014.43.

17. Vardhan, P. V., Reddy, C. V. A., Krishna, R., Reddy, N. Y., Kumar, K. S., Kumar, B. S. & Rao, T. V. V. L. N. "Machine learning applications for early-stage diagnosis of diabetic patients". *AIP Conference Proceedings*. 2023. https://www.scopus.com/authid/detail.uri?authorId=57739380300. DOI: https://doi.org/10.1063/5.0134168.

18. Van den Berg, M., Medina, O., Loohuis, I., van der Flier, M., Dudink, J., Benders, M., Bartels, R. & Vijlbrief, D. "Development and clinical impact assessment of a machine-learning model for early prediction of late-onset sepsis". *Computers in Biology and Medicine*. 2023; 163: 107156, https://www.scopus.com/authid/detail.uri?authorId=57960218300.
DOI: https://doi.org/10.1016/j.compbiomed.2023.107156.

19. Ahmed, A. S. & Salah, H. A. "A comparative study of classification techniques in data mining algorithms used for medical diagnosis based on DSS". *Bulletin of Electrical Engineering and Informatics*. 2023; 12 (5): 2964–2977, https://www.scopus.com/authid/detail.uri?authorId=57204477939.
DOI: https://doi.org/10.11591/eei.v12i5.4804.

20. "Safer and more efficient vital signs monitoring to identify the deteriorating patient: An observational study towards deriving evidence-based protocols for patient surveillance on the general hospital ward". – Available from: https://www.fundingawards.nihr.ac.uk/award/17.05.03. – [Accessed: 28 July 2022].

21. Liubchenko, V., Komleva, N. & Zinovatna S. "A Bot-Based Self-Report Diagnostic Tool to Assess Post-Traumatic Stress Disorder". *Proceedings of the 18th International Conference on ICT in Education, Research and Industrial Applications*. Ivano-Frankivsk: Ukraine. 2023. CEUR Workshop Proceedings (in press).

22. Cloitre, M., Shevlin, M., Brewin, C. R., Bisson, J. I., Roberts, N. P., Maercker, A., Karatzias, T. & Hyland, P. "The International Trauma Questionnaire: development of a self-report measure of ICD-11 PTSD and complex PTSD". *Acta Psychiatrica Scandinavia.* 2018; 138: 536–546, https://www.scopus.com/authid/detail.uri?authorId=35420810100.

Conflicts of Interest: The authors declare no conflict of interest

Received 15.06.2023
Received after revision 02.09.2023
Accepted 18.09.2023

DOI: https://doi.org/10.15276/aait.06.2023.19
УДК 004.414.32

Методологія виявлення хворіб методами аналізу даних

Vira V. Liubchenko[2)]
ORCID: https://orcid.org/0000-0002-4611-7832; lvv@op.edu.ua. Scopus Author ID: 56667638800
Комлева Наталія Олегівна[1)]
ORCID: https://orcid.org/0000-0001-9627-8530; komleva@op.edu.ua. Scopus Author ID: 57191858904
Зіноватна Світлана Леонідівна[1)]
ORCID: https://orcid.org/0000-0002-9190-6486; zinovatnaya.svetlana@op.edu.ua. Scopus Author ID: 57219779480
Briggs Jim[3)]
ORCID: https://orcid.org/0000-0003-1559-4879; Jim.Briggs@port.ac.uk. Scopus Author ID: 7201863846
[1)] Національний університет «Одеська політехніка», пр. Шевченка 1, Одеса, 65044, Україна
[2)] Hochschule für Angewandte Wissenschaften Hamburg, Fakultät Life Sciences, Ulmenliet 20
[3)] Центр моделювання та інформатики охорони здоров'я, Університет Портсмута, Портсмут, PO1 3HE. Великобританія

АНОТАЦІЯ

Метою дослідження є розробка інформаційної технології для виявлення проблемних станів здоров'я шляхом аналізу даних вимірювань. Огляд літератури висвітлює різні підходи до медичної діагностики, включаючи статистичні моделі та моделі машинного навчання, які прогнозують ризик несприятливих наслідків на основі даних пацієнта. Розроблена інформаційна технологія фокусується на класифікації та достатності даних, забезпечуючи збір об'єктивних та релевантних даних. Технологія включає визначені експертами правила аналізу, які допомагають генерувати діагнози-кандидати для пацієнтів. Запропонована інформаційна система складається з чотирьох компонентів: джерело даних, сховище даних, модуль діагностики та поглинач даних. Комплексна структура зберігання даних призначена для ефективного зберігання та управління даними, пов'язаними з діагнозами та параметрами. Прототип блоку генерації наборів правил включає отримання параметрів і перетворення алгоритмів у програмні функції. Тематичне дослідження фокусується на діагностичному інструменті для оцінки ПТСР за допомогою міжнародно визнаного опитувальника. Джерелом даних обрано телеграм-бот через його анонімність, гнучкість та можливість автоматизованого збору даних. Структура бази даних розроблена таким чином, щоб пристосуватись до модифікацій опитувальника та продовжити збір даних. Впроваджена аналітична система ефективно класифікує стани респондентів на основі їхніх відповідей. В цілому, дослідження демонструє потенціал інформаційних технологій та запропонованої інформаційної системи для забезпечення ефективної та зручної діагностики стану здоров'я, що сприяє своєчасному медичному втручанню та покращенню добробуту населення.

Ключові слова: Моніторинг здоров'я; аналіз даних; діагностика; інформаційна технологія; аналітична система; Telegram-бот

ABOUT THE AUTHORS

Vira V. Liubchenko – Doctor of Engineering Sciences, Professor; Lecturer, Fakultät Life Sciences, Hochschule für Angewandte Wissenschaften Hamburg, Ulmenliet 20. Hamburg, 21033, Germany
ORCID: https://orcid.org/0000-0002-4611-7832; lvv@op.edu.ua. Scopus Author ID: 56667638800
Research field: Data Science; Software Engineering; Project Management

Nataliia O. Komleva – PhD, Head of the Department of Software Engineering. Odessa Polytechnic National University, 1, Shevchenko Ave. Odessa, 65044, Ukraine
ORCID: https://orcid.org/0000-0001-9627-8530; komleva@op.edu.ua. Scopus Author ID: 57191858904
Research field: Data analysis; software engineering; knowledge management

Наталія Олегівна Комлева – кандидат технічних наук, завідувач кафедри Інженерії програмного забезпечення Національного університету «Одеська політехніка», пр. Шевченка, 1. Одеса, 65044, Україна

Svitlana L. Zinovatna – PhD, Associate Professor of the Department of Software Engineering, Odessa Polytechnic National University, 1, Shevchenko Ave. Odessa, 65044, Ukraine
ORCID: https://orcid.org/0000-0002-9190-6486; zinovatnaya.svetlana@op.edu.ua. Scopus Author ID: 57219779480
Research field: Data analysis; information system productivity

Світлана Леонідівна Зіноватна – кандидат технічних наук, доцент кафедри Інженерії програмного забезпечення Національного університету «Одеська політехніка», пр. Шевченка, 1. Одеса, 65044, Україна

Jim Briggs – Professor of Informatics, Centre for Healthcare Modelling and Informatics. University of Portsmouth, Portsmouth, PO1 3HE. United Kingdom
ORCID: https://orcid.org/0000-0003-1559-4879; Jim.Briggs@port.ac.uk. Scopus Author ID: 7201863846
Research field: Health informatics; particularly clinical outcome modeling and digital well-being

Briggs Jim – професор інформатики, Центр моделювання та інформатики охорони здоров'я.
Університет Портсмута. Портсмут, PO1 3HE. Великобританія