PRECISION DIAGNOSIS OF DEPRESSION LEVELS VIA DISTRIBUTED CLASSIFICATION AND SHAP ANALYSIS

Vladislav Kaverinskiy, Kyrylo Malakhov

Abstract: *This article presents a novel classification service designed for medical datasets with limited sample sizes, specifically focusing on depression assessment based on blood pressure oscillograms. The service employs multiple binary classification components, each trained to differentiate specific class separations. A key feature of the method is the use of both individual input features and correlated feature products, enhancing classification accuracy. The system achieved notable performance on a small dataset with accuracy, recall, and precision values of 0.9172, 0.9341, and 0.9811, respectively. While precision is high, recall is relatively lower, indicating a slight tendency to underestimate depression levels. Class 4, representing the highest depression level, demonstrated the most classification mismatches due to the small sample size. The SHAP analysis identified five key features—ULF_70-100, ULF_per_100-70, O_L2_pos, ULF_100-70, and ULF_per_total—as the most influential, most of which are associated with ultra-low-frequency (ULF) oscillations. The developed classification service proves its applicability in healthcare, offering potential for further adaptation to broader medical domains, enabling personalized medical diagnostics and remote healthcare solutions.*

Keywords: *G.3 Probability and Statistics, J.3 Life and Medical Sciences, I.2.1 Applications and Expert Systems, I.2.8 Problem Solving, Control Methods, and Search.*

DOI: https://doi.org/10.54521/ijita32-02-p06

Introduction

Classification AI systems have become integral to fields like medical diagnostics, finance, and natural language processing. These systems categorize data into predefined classes, enabling automated decision-making. The success of such systems heavily relies on the choice of algorithms, with prominent examples being support vector machines (SVMs), neural networks, and decision trees. One of the most widely used algorithms is the Random Forest, introduced by Leo Breiman in 2001 [Breiman, 2001]. Random Forest is an ensemble learning method that builds multiple decision trees and outputs the mode for classification tasks or the mean for regression. It combines decision trees' simplicity with ensemble learning's power, offering robustness and versatility.

The Random Forest method builds a "forest" of decision trees, each trained on randomly selected subsets of data (bootstrapping). Randomness in feature selection reduces overfitting and improves accuracy and generalization. In various domains, Random Forests have shown strong performance. In medicine, they help predict disease outcomes based on patient data [Chen, Guestrin, 2016], while in finance, they are used for credit scoring and fraud detection [Liaw, Wiener, 2002]. Environmental sciences benefit from Random Forests' capacity to model complex ecological and climate systems [Prasad, Iverson, Liaw, 2006].

Recent enhancements, such as extremely randomized trees [Geurts, Ernst, Wehenkel, 2006] and deep learning integration, have expanded Random Forest capabilities, allowing them to handle increasingly complex datasets. Although the AI classification system developed in this study was trained on medical data, particularly blood pressure oscillograms, it is applicable beyond medical tasks. It is especially suitable for datasets with numerous features and relatively limited data rows.

Blood pressure measurement is critical in healthcare [Caro et al., 2011], [Vakulenko et al., 2022], and electronic blood pressure meters now offer more detailed data. Some models register arterial pulsations, allowing for precise blood pressure calculations and data transmission [Vakulenko et al., 2023], [Martsenyuk et al., 2022]. These readings can reveal deeper insights into a patient's health [Caro et al., 2011], [Pokrovsky, 1979], [Warner et al., 1953], [Langewouters et al., 1984]. This study focuses on using these oscillograms to create a classifier that can estimate depression levels, providing an objective alternative to potentially biased patient surveys.

The primary aim of this study is to develop and test an AI classification system that efficiently handles complex datasets, identifies significant features, and fine-tunes classifier hyperparameters to maximize accuracy.

1. Initial Dataset Characterization and Numerical Research Procedure

Patients aged 32–65 with mental disorders at Ternopil Regional Clinical Psychoneurological Hospital were assessed using the Hospital Depression Rating Scale (HDRS) and the DASS-21 scale. The diagnoses included bipolar affective disorder with a depressive episode and depressive disorders without psychotic features. Given the high number of input parameters (1030 factors) and a relatively small sample size (181 patients), the data was analyzed carefully to prevent distortion. Patients were divided into five groups based on their depression levels, with groups 1 to 4 determined by questionnaire results and group 5 consisting of hospitalized patients.

The study builds upon previous research at the V. M. Glushkov Institute of Cybernetics of NAS of Ukraine, which developed a cloud telerehabilitation platform [Palagin et al., 2023], [Palagin et al., 2023, 2], [Kaverinskiy, Malakhov, 2023], [Palagin, Petrenko, 2018]. The current study aims to enhance this platform by integrating a classification service for medical purposes. Initial analysis of the dataset was conducted using the "Lazy Predict" technique from Scikit Learn, which ranks classifiers by accuracy. With over 1000 parameters but only 181 data rows, feature selection was critical. Correlated features within each class were identified to improve classification accuracy.

Raw data alone did not produce satisfactory results, so key features were identified based on their correlations. The classifier was retrained by adding new features from correlated pairs, enhancing accuracy. Various classifiers were tested, even those not ranked highest by "Lazy Predict," as tuning their hyperparameters could improve performance. A truth table was created to map classifier results to patient classes, formalized in JSON for integration into an automatic classification service.

2. Results and Discussion

From the large dataset, ultra-low-frequency (ULF) parameters were selected as the foundation for classification. ULF represents oscillations below 0.003 Hz and plays a crucial role in coordinating the body's functional state with external factors. UMAP analysis showed that ULF factors effectively divided patients into two groups: Group 1 (classes 1 and 2, with milder depression) and Group 2 (classes 3, 4, and 5, with more severe depression). These groups were used for initial classification using the "Lazy Predict" technique.

UMAP also identified 11 patients who formed a distinct cluster across all depression levels. These patients exhibited unique feature values, including zero ULF_100-70 and ULF_per_100-70, along with high ULF_70-100 and ULF_70-end values. As these patients did not align with the primary classification groups, they were excluded from the dataset to improve accuracy.

While using ULF values directly did not yield high accuracy, the study introduced a novel approach by identifying feature pairs that were strongly correlated in one class but weakly correlated in another. Products of these correlated features were added to the model, significantly improving classification accuracy. Figure 1 illustrates the feature selection process, where features that enhanced accuracy were incorporated into the model until no further improvements were observed.

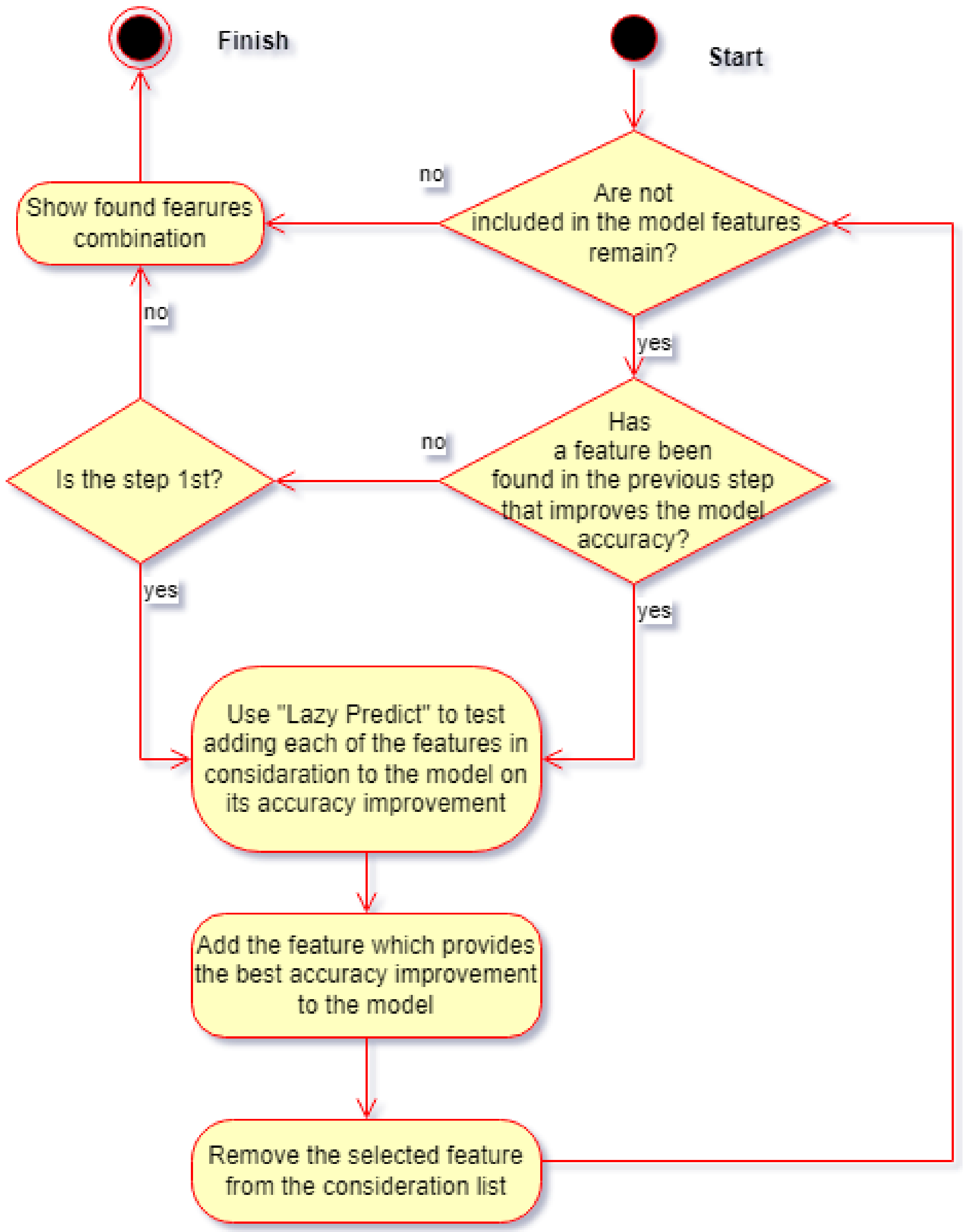

Figure 1. Model's **features selection scheme**

Initially, 2 ULF features and 12 of their products increased the model’s accuracy to 0.78. By adding 11 more feature products, accuracy improved to 0.95. KNeighborsClassifier performed best using the "Lazy Predict" technique, though RandomForestClassifier also showed good results. While the current classifier distinguishes between two depression groups, the aim is to eventually create binary classifiers to differentiate all five classes for practical applications.

Although KNeighborsClassifier initially showed the highest accuracy in grouping classes (classes 1 and 2 as group 1, and classes 3, 4, and 5 as group 2) using the "Lazy Predict" technique, further hyperparameter tuning revealed that Random Forest could achieve even better accuracy—up to 0.97. This classifier also performed well for other class groupings. The optimized hyperparameters for distinguishing class groups are shown in Table 1. The Powell method, implemented in SciPy, was used to fine-tune N estimators, Max depth, and Min samples split.

For distinguishing Class 5 (clinical cases) from others, the Decision Tree classifier was preferred over Random Forest due to its simplicity and high accuracy. Each classifier provides binary output based on its training. The results were compiled into a truth table (Table 2) and organized as a JSON structure for the classification service. Controversial cases were numerically adjusted (+0.25 for burdened, -0.25 for palliated).

Table 1. Classifiers hyper parameters value and achieved accuracy

Classifier No. and Classes groups	Classifier type	Hyper parameters			Achieved accuracy
		N estimators	Max depth	Min samples split	
Classifier 1 1 – Level 1 2 – Levels 2, 3, 4, 5	Random Forest	116	40	12	0.95
Classifier 2 1 – Levels 1 and 2 2 – Levels 3, 4, 5	Random Forest	47	7	7	0.97
Classifier 3 1 – Levels 1, 2, 3 2 – Levels 4, 5	Random Forest	47	11	5	0.93
Classifier 4 1 – Levels 1, 2, 3, 4 2 – Level 5	Decision Tree	—	—	2	0.99

Table 2. A truth table for the classifiers possible results combinations

Classifier results combinations				Result interpretation
Classifier 1	Classifier 2	Classifier 3	Classifier 4	
Class 1	Class 1	Class 1	Class 1	Class 1 (light depression)
Class 2	Class 1	Class 1	Class 1	Class 2 (higher depression level)
Class 1	Class 2	Class 1	Class 1	not reliable result, probably Class 3 (middle higher depression level)
Class 2	Class 2	Class 1	Class 1	Class 3 (serious depression level)
Class 1	Class 1	Class 2	Class 1	not reliable result, probably Class 2 (higher depression level)
Class 2	Class 1	Class 2	Class 1	not reliable result, probably Class 2 (higher depression level, may be burdened)
Class 1	Class 2	Class 2	Class 1	not reliable result, probably Class 4 (serious or severe depression level)

Class 2	Class 2	Class 2	Class 1	Class 4 (severe depression level)
Class 1	Class 1	Class 1	Class 2	Class 5 (clinical cases)
Class 2	Class 1	Class 1	Class 2	
Class 1	Class 2	Class 1	Class 2	
Class 2	Class 2	Class 1	Class 2	
Class 1	Class 1	Class 2	Class 2	
Class 2	Class 1	Class 2	Class 2	
Class 1	Class 2	Class 2	Class 2	
Class 2	Class 2	Class 2	Class 2	

The developed distributed classification system achieved an accuracy of approximately 0.92 when distinguishing between all five depression classes in the final test. However, the performance varied among different classes. Since the system is not binary, standard true positive and negative classifications do not fully apply. We can consider true positives when the estimated class matches the actual class, false positives when the system overestimates the depression level, and false negatives when it underestimates. The confusion matrix (Table 3) shows 156 true positives, 3 false positives, and 11 false negatives, with no true negatives (TN = 0).

Table 3. The confusion matrix

	Positive	Negative
True	156	0*
False	3	11

* - we unable to consider any result as a true negative for the given case

The calculated formal criteria are as follows:

Accuracy: 0.9172

Recall: 0.9341

Precision: 0.9811

These results suggest that the classifier has high precision but slightly lower recall, indicating it tends to overestimate rather than underestimate depression severity. The system struggles most with identifying patients in Class 4, where 37.5% of the classifications were incorrect. This could be due to the small sample size (only 10 patients) or potential overestimation during data collection.

A SHAP analysis on a 20% test sample revealed the five most significant features for classification:ULF_70-100, ULF_per_100-70, O_L2_pos, ULF_100-70, and ULF_per_total.

ULF_70-100 was the most influential feature, confirming its importance as it appears as both a direct parameter and in product form within the model. In contrast, features like HF_70-100_int_p and Hurst_20-70 were found to have minimal impact. The main test results are shown in a diagram in Figure 2.

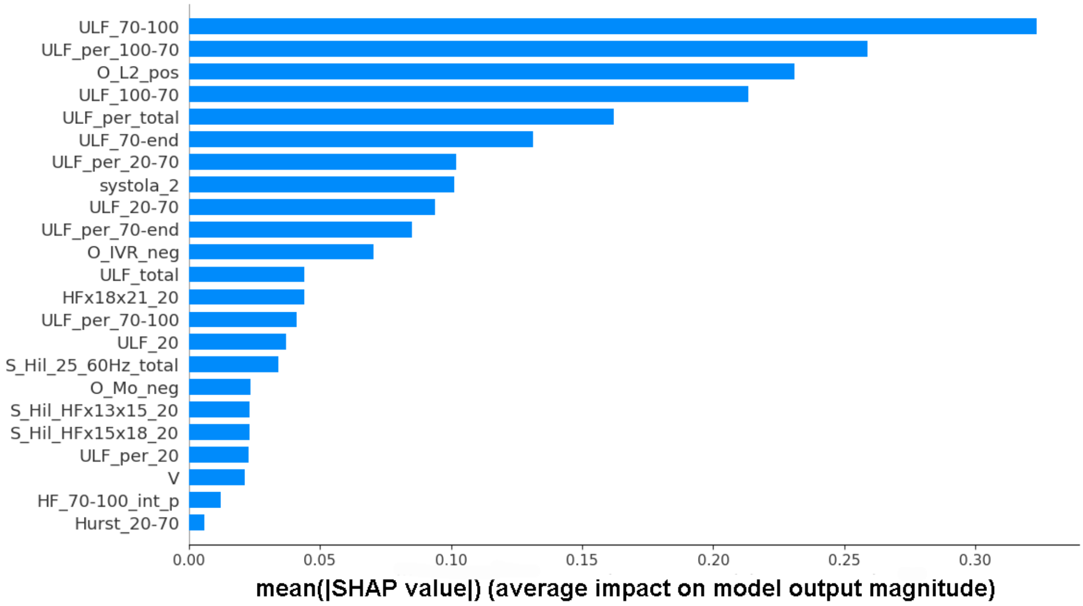

Figure 2 SHAP test results on the features impact to the classification

The classification system demonstrates strong scalability and adaptability. Its distributed architecture enables efficient processing of large datasets and complex tasks through parallel processing and load balancing. The modular design allows for easy integration of new classifiers and incremental training, making it suitable for evolving medical environments.

This classifier can enhance healthcare systems by automating data analysis, providing real-time assessments, and integrating with Electronic Health Record (EHR) systems. Such integration can reduce healthcare professionals' diagnostic workload and improve patient assessment accuracy. While initially designed for depression classification, the system is adaptable to other medical domains, such as cardiovascular, respiratory, and neurological conditions, making it a promising tool for personalized medicine.

Conclusion

A classification service was developed that utilizes several binary classification components, each trained to distinguish specific class separations. This method allows for accurate classification using a relatively small dataset by incorporating both individual feature values and the products of correlated features, significant only within particular data groups. The system was tested on a medical dataset of blood pressure oscillograms, which were linked to depression levels.

Despite the small dataset, the system achieved strong results: Accuracy = 0.9172, Recall = 0.9341, and Precision = 0.9811. While the classifier demonstrates high precision, recall is slightly lower, indicating a tendency to underestimate depression levels rather than overestimate them. The greatest classification mismatch occurred in Class 4, likely due to a small sample size.

According to the SHAP test, the most important features for classification were ULF_70-100, ULF_per_100-70, O_L2_pos, ULF_100-70, and ULF_per_total, with most of these belonging to the ULF group, which reflects the power of ultra-low-frequency oscillations (below 0.003 Hz).

Acknowledgements

This study would not have been possible without the financial support of the National Research Foundation of Ukraine (Open Funder Registry: 10.13039/100018227). Our work was funded by Grant contract: Development of the cloud-based platform for patient-centered telerehabilitation of oncology patients with mathematical-related modeling [Malakhov, 2023], [Stetsyuk et al., 2023], [Sergienko et al., 2023] application ID: 2021.01/0136.

Bibliography

[Breiman, 2001] Random Forests. Machine Learning 45, 5-32, 2001. doi:10.1023/A:1010933404324.

[Chen, Guestrin, 2016] XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785-794, 2016. doi:10.1145/2939672.2939785.

[Liaw, Wiener, 2002] Classification and Regression by Random Forest. R News 2, 18-22, 2002. http://CRAN.R-project.org/doc/Rnews/.

[Prasad, Iverson, Liaw, 2006] Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. Ecosystems 9, 181-199, 2006. doi:10.1890/07-0539.1.

[Geurts, Ernst, Wehenkel, 2006] Extremely Randomized Trees. Machine Learning 63, 3-42, 2006. doi:10.1007/s10994-006-6226-1.

[Caro et al., 2011] The Mechanics of the Circulation. 2nd ed. Cambridge University Press, 2011. doi:10.1017/CBO9781139013406.

[Vakulenko et al., 2022] Application of Arterial Oscillography to Study the Adaptive Capacity of Subjects with COVID-19 in Primary Care. In: Primary Health Care, A. Emel Önal, Ed., IntechOpen, 2022. doi:10.5772/intechopen.98570.

[Vakulenko et al., 2023] Spectral Analysis of Arterial Pulsations Recorded During Blood Pressure Measurement by the Oranta-AO Information System. In: New Developments in Medical Research. Nova Science Publishers, Inc., 2023. https://novapublishers.com/shop/arterial-oscillography-new-capabilities-of-the-blood-pressure-monitor-with-the-oranta-ao-information-system/.

[Martsenyuk et al., 2022] On the Development of Directed Acyclic Graphs in Differential Diagnostics of Pulmonary Diseases with the Help of Arterial Oscillogram Assessment. In: Graph-Based Modelling in Science, Technology and Art, Mechanisms and Machine Science, vol. 107. Springer International Publishing, pp. 157–173, 2022. doi:10.1007/978-3-030-76787-7_8.

[Pokrovsky, 1979] Clinical Angiology. Medicine, 1979. Available: https://cdn.e-rehab.pp.ua/u/pokrovsky-clinical-angiology-1979.pdf.

[Warner et al., 1953] Quantitation of Beat-to-Beat Changes in Stroke Volume from the Aortic Pulse Contour in Man. J Appl Physiol. 5, 495–507, 1953. doi:10.1152/jappl.1953.5.9.495. Available: https://pubmed.ncbi.nlm.nih.gov/13034677/.

[Langewouters et al., 1984] The Static Elastic Properties of 45 Human Thoracic and 20 Abdominal Aortas In Vitro and the Parameters of a New Model. J Biomech 17, 425–435, 1984. doi:10.1016/0021-9290(84)90034-4.

[Palagin et al., 2023, 1] OntoChatGPT Information System: Ontology-Driven Structured Prompts for ChatGPT Meta-Learning. International Journal of Computing 22, 170–183, 2023. doi:10.47839/ijc.22.2.3086.

[Palagin et al., 2023, 2] Digital Health Systems: Ontology-Based Universal Dialog Service for Hybrid e-Rehabilitation Activities Support. In: 2023 IEEE 12th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), volume 1, pp. 84–89, 2023. doi:10.1109/IDAACS58523.2023.10348639.

[Kaverinskiy, Malakhov, 2023] Natural Language-Driven Dialogue Systems for Support in Physical Medicine and Rehabilitation. South African Computer Journal 35, 119–126, 2023. doi:10.18489/sacj.v35i2.17444.

[Palagin, Petrenko, 2018] Methodological Foundations for Development, Formation and IT-Support of Transdisciplinary Research. Journal of

Automation and Information Sciences 50, 1–17, 2018. doi:10.1615/JAutomatInfScien.v50.i10.10.

[Malakhov, 2023] Insight into the Digital Health System of Ukraine (eHealth): Trends, Definitions, Standards, and Legislative Revisions. International Journal of Telerehabilitation 15, 1–21, 2023. doi:10.5195/ijt.2023.6599.

[Stetsyuk et al., 2023] Unified Representation of the Classical Ellipsoid Method. Cybern Syst Anal 59.5, 784–793, 2023. doi:10.1007/s10559-023-00614-x.

[Sergienko et al., 2023] Algorithm Unions for Solving Discrete Optimization Problems. Cybern Syst Anal 59.5, 753–762, 2023. doi:10.1007/s10559-023-00611-0.

Authors' Information

Vladislav Kaverinskiy – *Glushkov Institute of Cybernetics of the NAS of Ukraine; Kyiv, 03187, Ukraine,*

Academician Glushkov Avenue, 40, Ukraine;

e-mail: insamhlaithe@gmail.com

Major Fields of Scientific Research: AI, Semantic Web; Ontology engineering; Transdisciplinary research.

Kyrylo Malakhov – *MSc, Researcher, Backend developer, DevOps in Microprocessor Technology Lab, Glushkov Institute of Cybernetics of the NAS of Ukraine, Kyiv, 03187,*

Academician Glushkov Avenue, 40, Ukraine;

e-mail: k.malakhov@outlook.com

Research interests: AI, Computational linguistics; Ontology engineering; Digital health (Hybrid e-rehabilitation).