D. O. Shevaga; O. K. Gorodetska, Cand. Sc. (Eng.), Assistant Professor;
L. M. Dobrovska, Cand. Sc. (Pedag.), Assistant Professor

PROGRAM APPLICATION FOR STRESS LEVEL MONITORING, BASED ON CLASSIFICATION MODELS

The paper considers studies, regarding the stress level. After COVID-19 pandemia and constant stay in the forced isolation, the level of stress became higher due to the increase of anxiety. That is why, the study of the basic mechanisms of the stress and monitoring of different biophysiological and biochemical reactions of the organism on the stress is of great interest for the researches. Reliable biomarker or stress indicator could provide accurate monitoring of stress, potentially enabling to avoid pathological states at the early stages. Long lasting stress may have negative health outcomes. Thus, the ability to determine when a person is in the state of stress, may be very useful to avoid health problems, especially for patients with suicidal thoughts.

The given study contains the results of the stress level monitoring by means of using the classification models as the forecasting and as a bio signal – heart rate variability (HRV) from the sensors of electrocardiography. Correlation of all variables was carried out , so that only those variables which have high correlation with stress could participate in the models learning. For achieving the set task the following methods have been used: artificial neural network, k-nearest neighbors (KNN), random forest, decision tree. Classification model random forest obtained the highest index of the forecasting accuracy of the presence or absence of the stress – 98 %. On the base of this model program application in programming language R with users interface was developed, it allows to load the data of the electrocardiogram and obtain the conclusion, regarding the level of the stress level. By means of the application the user can control the level of personal stress and lead a healthy life.

Key words: *stress monitoring, electrocardiography, machine learning, heart rate variability, classification algorithms, bio signals, decision tree, random forest.*

Problem statement

Studies in the sphere of stress show that stress can cause the disturbance of the heart rate that leads to the arrhythmia and damages both physical and mental health [1]. Actually, chronic stress can induce the development of depression, stroke emergence, diabetes and even cancer. Stress influences all the systems of the organism, including musculoskeletal, respiratory, cardiovascular, endocrine, gastrointestinal, nervous and reproductive systems. Besides, stress was determined as one of the main reasons of automobile accidents, causing high mortality rate [2]. Proceeding from these reasons, it is important to develop the application that can reliably detect and measure human stress. As the stress in every person is manifested differently, it is expedient to develop the model and program application which will personalize the evaluation of the individual level of stress, taking into account the input indices. As the parameter for the stress level assessment were chosen the indices of heart rate variability, heart rate, and RR intervals modifications from the sensors of electrocardiography, as they are proved by the numerous studies [3 – 5]. These matrices were selected for the data base development.

Analysis of the recent studies and publications

Overstress may cause serious problems and increase the risk of the diseases. Nowadays there exist studies devoted to stress identification. In the study [6] the authors presented the data base DRIVEDB and investigated several methods of clusterization, including the classification SVM-RBF, KNN and RBF. The best results were obtained with the accuracy of 83 %, using the classifier SVM-RBF [6]. Authors [7] determined the level of stress by means of biological signals on the base of sensors, such as heartbeats rate, breathing parameters. KNN and SVM methods were used

for the classification, high accuracy indices were obtained (Table 1). Other researchers analyzed the stress of the busy persons by means of different body sensors, Kinect 3D sensors, video observation, individual levels of stress are classified on the base of SVM, KNN algorithm. This study was carried out on the base of SWELL-KW data, the accuracy of the constructed model is 92.75 %. Table 1 contains the generalized data, regarding stress classification.

Table 1

Generalized table of the recent research regarding stress classification

Authors	Dataset Used	Classifier	Classification	Accuracy
Adnan Ghaderi et.al.	PHYSIONET	KNN, SVM	Low, Medium, High	100 sec & 200 sec- 98.41% 300 sec- 90%
Nermine Munla et.al.	DRIVEDB	SVM- RBF, KNN, RBF	Normal, Stressed	Highest accuracy SVM- RBF-83%
Sriramprakash. S et.al.	SWELL-KW	SVM	Stressed, Normal	92.75%
Ravinder Ahuja et.al.	206 students of JIIT Noida	Random forest, Naïve Bayes, SVM, KNN	Highly stressed, Stressed, Normal	Highest accuracy SVM- 85.71%
Md Fahim Rizwan et.al.	PHYSIONET	SVM	Stressed, Non- Stressed	98.6%
Cheng-Ping Hsieh	WESAD	XGBoost	Stressed, Non-Stressed	92.38%
S. M. Chaware et.al.	Information extraction from Facebook attributes	TVSM	Positive, Negative	84.2%

Objective of the research

Objective of the research is to improve the accuracy of the prediction of the stress presence by constructing the classifiers on the base of realization of the artificial neural network, usage of the method of k-nearest neighbors, random forest and development of program application, enabling to assess the possible level of stress, where for the prediction of the classification model, the results of ECG bio signals processing are used, namely values of the heart rate variability .

Materials and methods of study

The set of data, the given study is based on, contains the heart rate variability indices, calculated on the base of the multimodal set of data SWELL (SWELL-KW) for the stress study. SWELL was developed by the researchers from the Institute of Computer Engineering and Information Science of Rabuds University . It is the result of the experiments, performed on 25 subjects, who perform typical office work. Totally 369288 cases were studied for the stress state. Trial subject passed typical work stresses, such as interruption for the arrived e-mail letters and pressure to finish the work in time. Each participant passed three different conditions of the work:

1. No stress: participants are allowed to work over the tasks as much as is needed, maximum 45 min, but they do not known the maximum duration of their tasks.
2. Time pressure: during this period, the time of the task execution has been reduced to 2/3 of the time the participant spent in neutral state.
3. Interruptions: participants obtained eight e-mails during the execution of the sceduled tasks. Some e-mails were related to their tasks – and the participants were asked to perform some actions, whereas other e-mails were not related to their tasks.

Data contain different attributes, obtained from the signals, measured by means of ECG, recorded

for different persons, who have different pulse at the moment of measuring. These different peculiarities influence the heart rate at a certain moment of time for a patient.

In general 36 different features for the classification of the stress levels have been determined, among them heart rate, RR intervals and their modifications (median, standard deviation, mean square root, percent of serial intervals etc.).

In the data set the amount of cases when a person was not at rest dominated (Fig. 1).

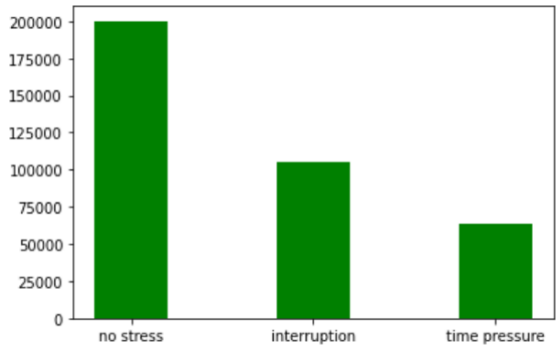

Fig. 1. Distribution according to the stress

Four classification algorithms have been investigated: method of k-nearest neighbors, neural network, random forest, decision tree.

Neural networks, also known as artificial neural networks (ANN) or simulated neural networks (SNN) are the subset of the machine learning and the base of the deep learning algorithms. Artificial neural networks (ANN) consist of the node layers, containing the input level, one or more hidden layers and the output level. Each node of artificial neuron is connected with another and has the corresponding weight and threshold [8].

Method of k-nearest neighbors – is a technique of the machine learning, that falls under the controlled learning. Algorithm assumes that the similar things exist very close to each other. The choice of the correct value of K is the most important task. For the set data, unknown for the machine, we can verify with different values of K, to obtain the reduced error for better prediction. First the number of K neighbors is selected, then the Euclidean distance is calculated and according to the calculation the selection of k-nearest neighbors is performed. Number of the data points is calculated in each category. For the category with maximum neighbor new points of data are assigned, after that the model is ready [9].

Decision tree – is the classifier, organized in the form of the tree, where each internal node corresponds to one of the input variables, branches are referred to the instructions regarding decision making, each terminal node is addressed to the result [10]. There are two nodes in the decision tree: decision making node and leaf node. Decision making nodes are used for the approval of any decision and have different branches, whereas leaf nodes are the result of these decisions and do not contain further branches.

Random forest is a popular assembler tool of the machine learning on the base of the decision trees, that has high adaptivity to data. This method reduces the degree of relearning by combining several evaluators of the relearning (that is decision trees) for the formation of the ensemble algorithm of learning [9]. Each decision tree can obtain corresponding result of the classification decision. Using the results of voting of each decision tree in the forest, the category of the sample, to be the subject of verification, is determined according to the principle of the majority rule. Category with higher votes in all the decision trees is defined as the final result.

Results of the research

For further study the correlation of all variables was performed to involve in the models training only the variables, which have high correlation with the stress. Having evaluated the correlation matrix,

the following variables were selected for the construction of the classification models: average value of RR intervals, median of RR intervals, ratio of the standard deviation of RR intervals to mean-square root of the serial differences of RR intervals, median of the relative RR intervals, ratio of the standard deviation RR intervals to mean square root of the serial differences of RR intervals for the relative differences of RR intervals, absolute power of the range of very low frequencies (VLF), transformation of the main component of VLF.

Normalization was performed, applying min-max method, as it is one of the most widely used methods of data normalization. For each function minimal value of this function is transformed into 0, maximum value is transformed into 1, and each other value is transformed into decimal value between 0 and 1.

For obtaining the objective model each sample was divided into learning (70 %) and testing (30 %) samples. Matrices of the classified models assessment are: accuracy sensitivity targeting, metrics F-1.

The constructed artificial neural network consists of six layers with the direct connection, where the data are transferred from the input layer to the output layer without the feedback (Fig. 2). Truncated Linear Node ReLU performs the function of activation in the network, however for obtaining the result of classification on the output layer the activation function softmax is used. Function of losses is categorial cross-entropy, optimization method – algorithm Adam.

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 7)	56
dense_1 (Dense)	(None, 35)	280
activation (Activation)	(None, 35)	0
dense_2 (Dense)	(None, 20)	720
activation_1 (Activation)	(None, 20)	0
dense_3 (Dense)	(None, 9)	189
activation_2 (Activation)	(None, 9)	0
dense_4 (Dense)	(None, 4)	40
activation_3 (Activation)	(None, 4)	0
dense_5 (Dense)	(None, 1)	5

Total params: 1,290
Trainable params: 1,290

Fig. 2. Architecture of the constructed artificial neural network

For the classification model KNN, the value 10 (k = 10) is selected as the basic parameter, namely a number of neighbors. For the classification model “Decision tree” the depth parameter equals 13 and the criterion “entropy” has been chosen as the function of the division quality measurement. Depth parameter for the random forest model equals 12 and the quantity of the trees in the model equals 21. Schematic representation of the architecture of the decision tree and random forest methods is shown in Fig. 3.

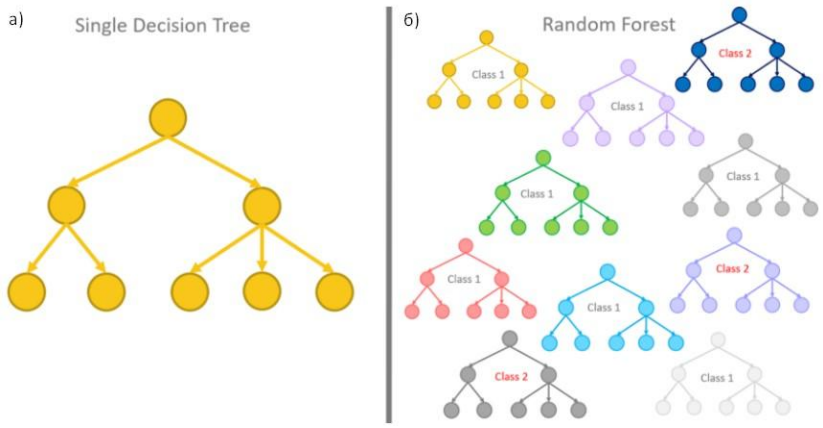

Fig. 3. Architecture: a) decision tree; b) random forest

Accuracy results for each constructed model are demonstrated (Fig. 4).

	Model			
	Artificial Neural Network	KNN	Random Forest	Decision Tree
Interruption	0.85	0.7	1.00	0.98
No stress	0.91	0.81	0.98	0.96
Time pressure	0.8	0.69	0.97	0.97
Accuracy	0.85	0.74	0.98	0.96

Fig. 4. Accuracy of the classification models

As it is seen, the highest index of accuracy obtained the prediction model of the random forest – 0.98. For the descriptive reasons the results of the models accuracy on the test sample will be demonstrated in the form of bar diagram (Fig. 5).

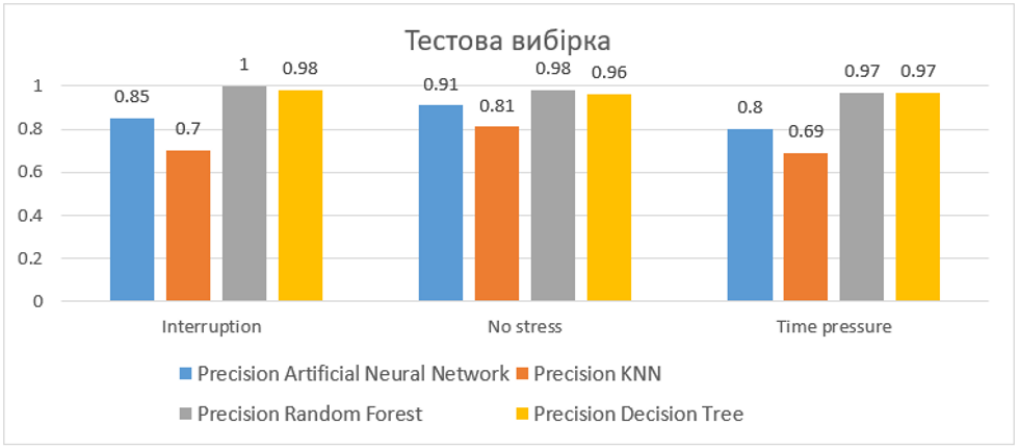

Fig. 5. Bar diagram of the precision results

For better understanding of the prediction models contradiction matrix NxN has been constructed, it is used for the assessment of the efficiency of the classification model, where N – is the number of the target classes. Contradiction matrix for the model of decision tree classifier is shown in Fig. 6.

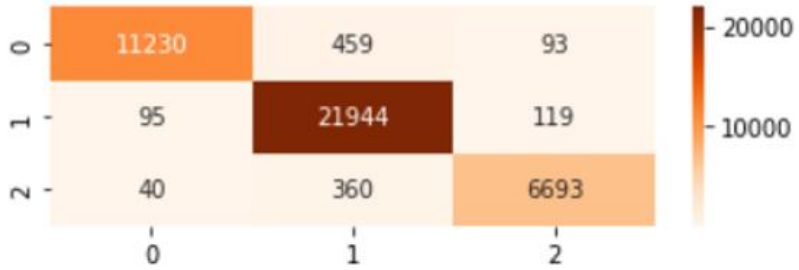

Fig. 6. Random forest contradiction matrix on the test sample

Fig. 6 shows that from 11230 cases in the state 0 (time pressure) 21945 are correctly classified in the state 1 (no stress), 6693 are correctly classified in the state 2 (time pressure). In general model made 1165 erroneous statements and correctly classified 39868 cases.

Random forest classification model obtained the highest precision indices, as compared with other constructed analogues, that is why, this model is used for the development of the software product, that will help the users control the level of the personal stress and lead a healthy life.

By means of programming language R program application for monitoring the stress level was developed. This application is cross-platform and can operate both on operation system Windows and operation system Linux. Data are stored in the built-in data base MySQL. Application was developed in accordance with modern requirements to UI/UX [11], maximally simple and understandable. For the operation with the program it is necessary to load the ECG data of the patient in the application. Input data must be written in CSV format. For the stress level monitoring it is sufficient to load the observations during one – two minutes, after that the software, using random forest classification model, outputs the result and the accuracy of the prediction made.

The program contains 4 pages, the first page – «Insert data» is responsible for patients data loading to detect the presence or absence of the stress. After dataset loading the user obtains the result of the program operation, namely, the conclusion regarding the belonging to one of the classes: no stress, interruption, time pressure and circular diagram of his states during the time period, loaded in the dataset (Fig. 7).

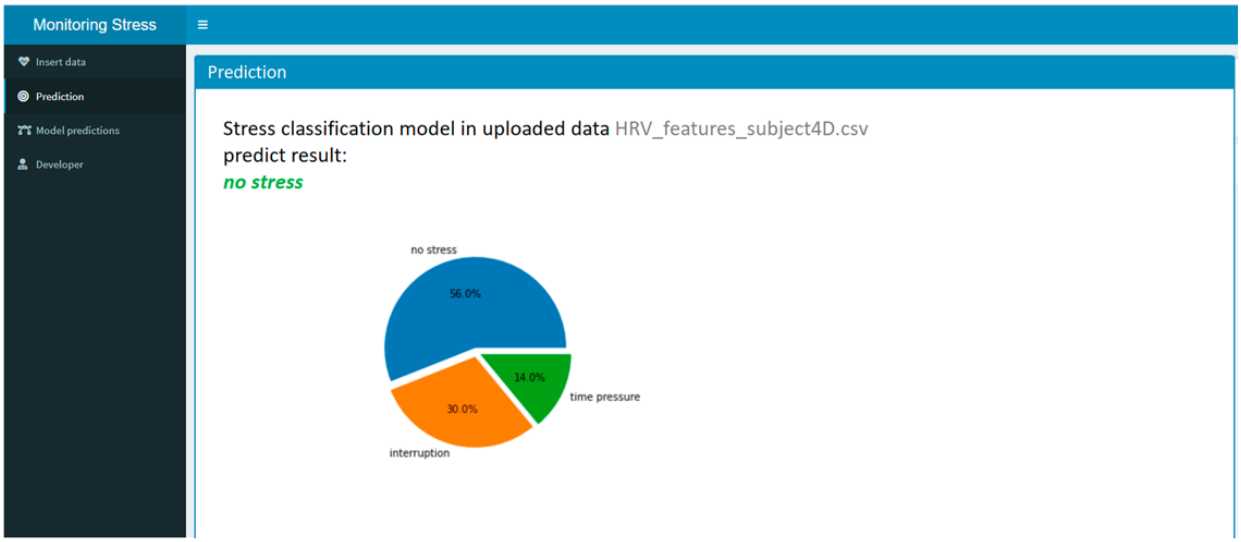

Fig. 7. Result of program application operation

Conclusions

On the base of SWELL-KW data base 369288 cases were analyzed, they were studied on the subject of the stress state, four classification models, enabling to predict the stress state with high

accuracy were constructed. The classifiers for the prediction of the stress presence were constructed: artificial neural network, method of k-nearest neighbors, random forest and decision tree. As a result the random forest model with the accuracy 98 % of the prediction of the presence or absence of the stress was obtained, this is the most accurate result as compared with the operation of Sriramprakash. S, where the accuracy of the classification model was 92.75 %. On the base of this model software with user interface was developed, it enables to download the data of electrocardiogram and obtain the conclusion, regarding the stress level.

REFERENCES

1. Medically-oriented design for explainable AI for stress prediction from physiological measurements [Electronic resource] / Dalia Jaber, Hazem Hajj, Fadi Maalouf, Wassim El-Hajj // BMC Med Inform Decis Mak. – 2022. – Access mode: https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-022-01772-2.

2. Choi M. Wearable device-based system to monitor a driver's stress, fatigue, and drowsiness / M. Choi, G. Koo, M. Seo, SW. Kim // IEEE Trans Instrum Meas. – 2017. – № 67. – P. 34 – 45.

3. Thermal Comfort and Stress Recognition in Office Environment [Electronic resource] / K. Nkurikiyeyezu, K. Shoji, A. Yokokubo, L. Guillaume // International Conference on Health Informatics. – 2019. – Access mode to the resource: https://www.researchgate.net/publication/330754493_Thermal_Comfort_and_Stress_Recognition_in_Office_Environment.

4. The SWELL Knowledge Work Dataset for Stress and User Modeling Research / Saskia Koldijk, Maya Sappelli, Suzan Verberne et al.]. // ICMI '14: Proceedings of the 16th International Conference on Multimodal Interaction. – 2014. – P. 291 – 298.

5. Koldijk S. Detecting Work Stress in Offices by Combining Unobtrusive Sensor / S. Koldijk, M. Neerincx, W. Kraaij // IEEE Transactions on Affective Computing. – 2018. – № 2. – P. 227 – 239.

6. Driver stress level detection using HRV analysis [Electronic resource] / N. Munla, M. Khalil, A. Shahin, A. Mourad // International Conference on Advances in Biomedical Engineering (ICABME). – 2015. – Access mode to the resource: https://ieeexplore.ieee.org/document/7323251.

7. Machine learning-based signal processing using physiological signals for stress detection [Electronic resource] / A. Ghaderi, J. Frounchi, A. Farnam // Iranian Conference of Biomedical Engineering (ICBME). – 2015. – Access mode to the resource: https://ieeexplore.ieee.org/abstract/document/7404123.

8. EEG signals classification using a new radial basis function neural network and jellyfish meta-heuristic algorithm [Electronic resource] / H. Rastegar, D. Giveki, M. Choubin // Evolutionary Intelligence. – 2022. – Access mode to the resource: https://link.springer.com/article/10.1007/s12065-022-00802-2.

9. Naser A. Classification of pleasant and unpleasant odor imagery EEG signals [Electronic resource] / A. Naser, O. Aydemir // Neural Computing and Applications. – 2022. – Access mode to the resource: https://link.springer.com/article/10.1007/s00521-022-08171-8.

10. Ekubo E. Using machine learning to predict low academic performance at a Nigerian university / E. Ekubo, B. Esiefarienrhe. // The African Journal of Information and Communication (AJIC). – 2022. – № 30. – P. 1 – 33.

11. What UI/UX design is – where to start learning from [Electronic resource] // IT STEP ACADEMY. – 2022. – Access mode to the resource : https://kiev.itstep.org/blog/what-is-ui-ux-design-where-to-start-learning. (Ukr).

Editorial office received the paper 14.12.2022.
The paper was reviewed 24.12.2022.

Shevaga Diana – Student group БС-11мп, e-mail: dinashevaha@gmail.com.

Gorodetska Olena – Cand. Sc. (Eng.), Associate Professor with the Department of Biomedical Cybernetics.

Dobrovska Ludmyla – Cand. Sc. (Pedag.), Associate Professor with the Department of Biomedical Cybernetics.

National Technical University of Ukraine "Igor Sikorskyi Kyiv Polytechnic Institute".