International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 11 | November - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Using Machine Learning Techniques to Analyze Stress Levels in Electroencephalogram Data

Rodrigo Alexandre dos Santos[1,2]

[1]*Pontifical Catholic University of Minas Gerais, Belo Horizonte, MG, Brazil*
[2]*Department of Software Development, CPQD Foundation, Campinas, SP, Brazil*

---***---

Abstract - Research in mental health has proposed the application of technologies that allow the analysis of substantial amounts of data and the provisioning of mechanisms to monitor and diagnose the health status of individuals with the objective of improving their quality of life. Mental stress has been widely addressed in this type of research as it is widely acknowledged as a condition that leads to negative impacts on modern life. Several studies have been directed to the generation of datasets with data from physiological signals obtained through sensors and the public availability of such datasets for further research. One of the main approaches for the processing and analysis of these datasets has been the use of Machine Learning techniques. This study explores the use of Machine Learning techniques to analyze a dataset consisting of mental stress level classification data from electroencephalograms. Different Machine Learning models were compared. The Multilayer Perceptron model presented the best performance with an accuracy rate of 98.99% in the predictions. The results demonstrate the potential of Machine Learning techniques as aiding tools to health monitoring and diagnosis.

Key Words: machine learning, deep learning, mental stress, electroencephalogram

1. INTRODUCTION

Several pieces of research have been developed in recent years addressing the use of technologies for the monitoring and diagnosis of mental health in individuals. According to Kang and Chai [1], the main reasons for the increased interest in these studies stem from the fact that billions of people around the world currently suffer from mental health problems and the fact that a considerable increase in the costs of mental health management is expected in the forthcoming years.

The need to improve health care as to promote the mental well-being of individuals has led to the development and improvement of sensors capable of monitoring biological signals and recording data about the mental state of individuals. According to Sheik et al. [2], sensor-based mental health monitoring and diagnosis methods have advantages over traditional methods because they are more objective and accurate due to relying on the direct measurement of physiological data, whereas traditional methods are more subjective and prone to distortion because they usually rely on self-report interviews and questionnaires of individuals. Another advantage of sensor-based methods, as quoted by Can et al. [3], is that sensors allow data to be recorded at the exact moment when events occur, hence increasing the possibility of obtaining accurate, relevant information, whereas traditional methods may fail to record useful information due to individuals' forgetfulness.

As a result of the complexity and the large amount of data generated by the use of sensors, the development of software solutions for the processing and analysis of such data has also increased, aiming to extract insights that allow a better understanding of the physiological changes occurring in the body of individuals and support decision-making in the medical field. According to Can et al. [3], complex software solutions have been developed to explore the capabilities of the devices more efficiently and deal with the challenge of processing large volumes of data within a short period of time.

According to Hickey et al. [4], the mental disorders that have been the focus of most interest in this research are stress, anxiety, depression, schizophrenia, and sleep disorders. Stress has received great attention and has been widely approached because, according to Hickey et al. [4] and Long et al. [5], chronic stress can increase the risk of cardiovascular diseases, even resulting in death, which justifies the interest in developing mechanisms to monitor and prevent this disorder.

This study proposes the use of Machine Learning techniques to analyze stress levels in electroencephalogram data. The study aims to explore different Machine Learning models and verify which one achieves better performance in dealing with the problem under study.

2. BACKGROUND

2.1. Types of sensors

Wearable sensors are devices attached to the skin of individuals and designed to capture physical, chemical as well as biological signals. The use of such devices provides information about various vital signs such as body temperature, blood pressure, breathing rate, heart rate, and brain electrical activity. Some sensors can obtain vital signs through electrodes attached to certain parts of the body, such as the forehead, scalp, chest, wrist, and fingers. Typically, these are the larger pieces of equipment, restricted to medical clinics, laboratories, and hospitals. There are also smaller sensors such as smart watches, smart bracelets, and smart shirts; these sensors are versatile enough to be used anywhere and allow the monitoring of individuals as they perform common everyday activities. According to Luo and Gao [6], wearable sensors have been improved in recent years to increase the accuracy of their measurements through the use of materials with enhanced sensitivity and performance as well as provide individuals with a less invasive and more comfortable experience that allows the device to be used in everyday situations over longer periods of time.

Certain studies have also analyzed the use of sensors built into smartphones, such as the accelerometer, gyroscope, and global positioning system (GPS). Although this type of sensor is not considered wearable as it is not physically connected to the human body, it also allows the collection of data about the behavior and actions of individuals. According to Kulkarni et al. [7], some motivators for the use of smartphones in mental health monitoring and diagnosis are the widespread adoption of the devices in recent years, in addition to the feasibility of easily carrying the devices to different locations and allowing the monitoring of individuals while performing a variety of activities. While the use of these devices opens up a range of new possibilities for research, the authors also agree upon the existence of advantages and disadvantages in comparison to wearable sensors; and the choice for a particular type of sensor may depend on several factors such as data collection environments, the degree of freedom of the individuals, in addition to the type of signals to be monitored.

2.2. Mental Health and Stress

The term “stress” stands for “tension” or “pressure”. It was first used by the Hungarian physician Hans Selye in a 1936 publication. Stress can be defined as a response of the body to external stimuli (also called stressors), which are perceived as demands or threats from the external environment. The action of such stressors generates an emotional arousal that leads the brain to trigger a process of adaptation to the environment. This process generates the secretion of certain hormones that can cause emotional, behavioral, and physical changes in the individual.

According to Shahsavarani et al. [8], stress generated at low levels is considered useful and desirable since it is a mechanism of adaptation and reaction to the various situations to which the individual is exposed in everyday life. This mechanism enables the individual to overcome challenges and thrive in dynamic environments. However, stress generated at high levels is harmful to the individual because it can trigger psychological as well as physiological problems that compromise the quality of life.

Stress is commonly classified into acute stress (short-term) or chronic stress (long-term). This classification considers the amount of time the body is exposed to a given stressor. According to McEwen and Akil [9], during acute stress, a rapid activation of the mechanism that triggers the secretion of associated hormones is observed, and the process ceases within a short time in order to avoid an overload in the organism as a result from the strong action of hormones. During chronic stress, the organism is exposed to stressors for a long period of time and, although it is not as intense as acute stress, the long exposure to stressors can cause the mechanism to become deregulated and allow the generation of imbalances in the organism that can lead to hypertension, heart attack and stroke.

The brain communicates with other organs in the body through the Autonomic Nervous System (ANS). The ANS enables the involuntary control of vital functions such as heart rate, blood pressure, respiratory rate, bowel function, and skin temperature. The ANS is divided into two parts: the Sympathetic Nervous System (SNS) and the Parasympathetic Nervous System (PNS). When the organism is exposed to stressors, the SNS goes into action through the release of hormones that cause the physiological changes required for the organism to adapt to the external environment. At the end of the stressful event, the PNS goes into action, inhibiting the effects of the SNS, with the objective of returning the organism to a state of homeostasis (internal stability). The two systems operate together and a balance between them is necessary for a healthy functioning of the organism. According to Shahsavarani et al. [8], chronic stress can affect the balance between the SNS and the PNS, and an irregular functioning of the system may lead to organism exhaustion.

Stress detection has been based on the analysis of psychological signals, physiological signals, and behavioral symptoms. An example of a psychological signal is an individual’s own perception of his or her mental state by completing a questionnaire. An example of a physiological sign is the set of hormonal changes that take place in an individual’s body in response to a stressful event. An example of a behavioral symptom is a change in the speed at which an individual moves. While psychological signs and behavioral symptoms are more subjective, physiological signs are considered more concrete and precise.

According to Baran [10], Long et al. [3], and Can et al. [3], solutions towards the detection of physiological stress signals have been based mostly on the analysis of data from the brain, heart, skin or blood, and the most common tests are: electroencephalogram (EEG), electrocardiogram (ECG), electromyogram (EMG), electrodermal activity (EDA), skin temperature (ST), galvanic skin response (GSR), blood pressure (BP), blood volume pulse (BVP), thermal imaging (TI) and salivary cortisol.

Several studies have been conducted with the participation of volunteers monitored in controlled environments while performing a stress-inducing task and having their physiological signals recorded. Natasha et al. [11] cite as examples of stress-inducing tasks: solving complex mathematical tests, vehicle driving simulators, and the color and word test (widely known as the Stroop test). Dham et al. [12] cite as examples of stress-inducing tasks: viewing videos on a given topic, public presentations, and playing numerical calculation games.

2.3. Electroencephalogram and Brainwaves

The Electroencephalogram (EEG) is a device that makes it possible to record the electrical activity of the brain through the use of electrodes connected to the scalp. As iMotions [13] puts it, when compared to other brain imaging analysis techniques, the EEG is considered as a low cost, lightweight, portable, non-invasive, and passive recording system. Also according to iMotions [13], an important advantage of the EEG is its high time resolution, which allows hundreds of data instances to be recorded in a one-second time interval.

The EEG measures voltage fluctuations resulting from ionic current inside neurons. The signal obtained represents the frequency of neural oscillation and is measured by the Hertz unit (Hz). According to Kumar and Bhuvaneswari [14], researchers have defined frequency bands for signal classification. The so-called bands or brainwaves are divided into five categories: delta waves (1-4 Hz), theta waves (4-8 Hz), alpha waves (8-12 Hz), beta waves (13-25 Hz) and gamma waves (25-100 Hz).

Research has shown that each type of brainwave is associated with certain mental states and is produced mostly in certain areas of the brain. According to Kumar and Bhuvaneswari [14], the generation of each type of wave causes the production of certain hormones, which allows inferring, for instance, the association of the types of waves with certain disorders or diseases.

The EEG is basically composed of electrodes, conductive gel, signal amplifier and signal converter (analog/digital). The device is considered flexible because it allows different configurations of the number of electrodes and their positioning on the scalp. According to iMotions [13], the most used electrode configuration standard is the so-called 10-20 System, which was proposed in 1957 by Heber Jasper and was recognized by the American Encephalographic Society in 1994 as the gold standard. As shown in Fig 1, this system defines 21 positions on the scalp and each position is identified by an acronym that represents the region or brain lobe where the electrode is positioned.

2.4. Datasets and Machine Learning

Data captured by wearable sensors have been made available through public datasets. According to Garg et al. [15] and Zainudin et al. [16], the initiative of making such datasets available has favored the work of researchers by allowing them to conduct their studies even without having participated in the data collection experiments.

Machine Learning techniques have been used to process datasets and provide learning through data analysis. The main goal of using Machine Learning in stress analysis has been the production of models that are able to learn patterns from the data and allow making predictions about the level of stress in individuals. According to Mozgovoy [17], Kyamakya et al. [18] and Garg et al. [15], several studies have applied the main Machine Learning algorithms in dataset analysis and evaluated their performance through metrics established in the literature in order to indicate which algorithm produces better accuracy in its predictions. According to Ahuja and Banga [19], and Kene and Thakare [20], several studies have reported results with accuracy rates over 80% in the prediction of stress levels, which was considered a satisfactory result in this context. According to Priya et al. [21], several researchers have applied different Machine Learning algorithms and obtained different accuracy rates depending on the scenario, which shows that a single algorithm is not the best to every situation, and that such studies have been conducted following an empirical approach.

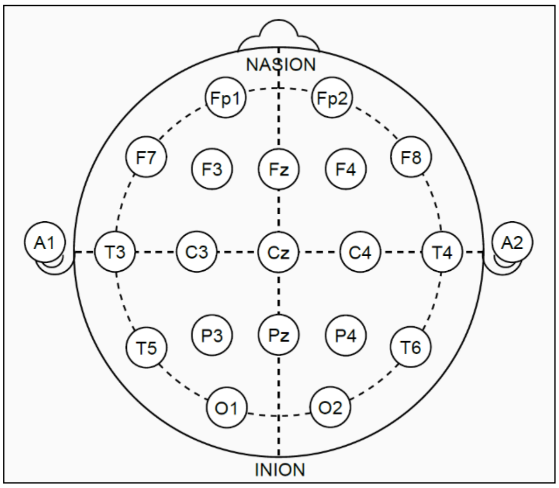

Figure 1 - Positioning System 10-20.

2.5. SAM 40 Dataset

SAM 40 is a dataset derived from research developed by Ghosh et al. [22] at the BCI Lab GU laboratory (Guwahati, India). The dataset presents an EEG data collection from 40 individuals monitored throughout 2019. The audience participating in the experiment consisted of students in the lab aged between 18 and 25 years. Data were recorded while subjects performed three stress-inducing tasks: color and word testing (Stroop), arithmetic question-solving, and symmetrical mirror image identification. According to Ghosh et al. [22], the SAM 40 was made publicly available in order to aid research in the field of Brain-Computer Interfaces (BCI) and contribute to pattern discovery and insights in stress identification.

The device used in the experiment was the Emotiv EPOC Flex Gel Kit (Fig 2). The EEG was set up according to the 10-20 electrode placement standard with a 32-channel recording pattern. The sampling frequency employed was 128 Hz. As shown in Fig 3, the electrodes were connected at the following positions: CZ, FZ, Fp1, F7, F3, FC1, C3, FC5, FT9, T7, CP5, CP1, P3, P7, PO9, O1, PZ, OZ, O2, PO10, P8, P4, CP2, CP6, T8, FT10, FC6, C4, FC2, F4, F8, and Fp2.

Figure 2 - Emotiv EPOC Flex Gel Kit.

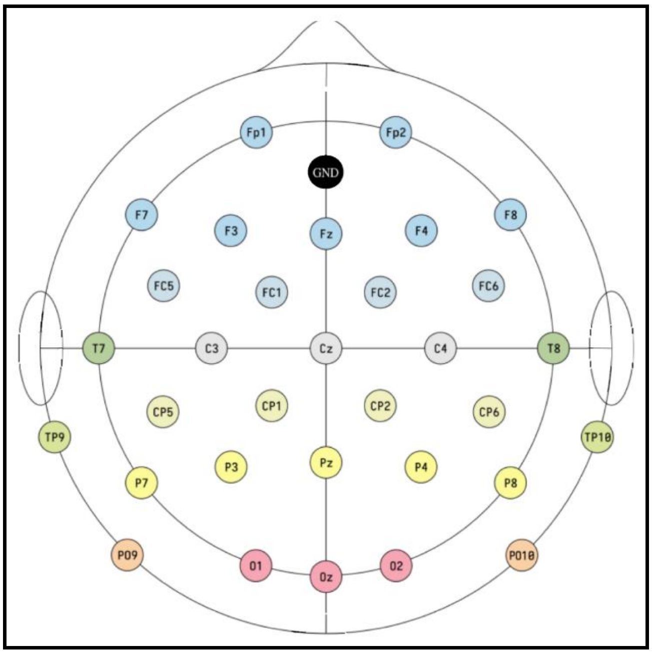

Figure 3 - EEG experiment set-up.

According to Ghosh et al. [22], each individual underwent a sequence of activities for three times (three trials), the steps of which can be described as follows:

1. The individual remained in a state of relaxation listening to relaxing music for 25 seconds;
2. Instructions on performing the first task were presented to the individual for 10 seconds;
3. The individual performed the Stroop test for 25 seconds;
4. The individual relaxed for 5 seconds;
5. Instructions on performing the second task were presented to the individual for 10 seconds;
6. The individual performed the symmetrical mirror image identification test for 25 seconds;
7. The individual relaxed for 5 seconds;
8. Instructions on performing the third task were presented to the individual for 10 seconds;
9. The individual performed arithmetic tests for 25 seconds;
10. The individual rated each of the three tasks on a scale from 1 to 10, where 1 represents the lowest stress level and 10 represents the highest stress level.

The tasks were performed in the same order during the second and third attempts. However, they were presented with slightly different content from the one presented on the first attempt in order to avoid any distortion derived from repetition and prior knowledge of the content.

The dataset went through a pre-processing stage before being made publicly available. According to Ghosh et al. [22], filtering techniques were employed to remove noise from the data that might have been caused by the influence exerted on the EEG signals by muscle and eye movements.

3. METHODS

3.1. Dataset Description

The SAM 40 dataset was obtained from Ghosh et al. [22]. The original set had 12 files for each of the 40 experiment participants, totaling 480 files. Each file stored the record from the execution of one of the three tasks performed by each individual or the record of the moment when the individual remained in a state of relaxation. The 480 files were unified, generating a single file to be used in this research.

The dataset generated gathered 1,536,000 samples with 32 features, where each feature represents the record of brain activity obtained by each of the electrodes used in the experiment. The value stored in each feature is the measurement in microvolts of the electrical activity recorded by the electrode. In addition, the dataset has the label that stores the classification of each sample. The rating is the individual's perception of the level of stress experienced while performing a particular task. The range of values for the classes varies from 0 to 10, where zero (0) indicates no stress and the range from 1 to 10 indicates the perception of the level of stress from the least intense (level 1) to the most intense (level 10).

3.2. Preprocessing

Data underwent an exploratory analysis, which pointed to lack of data inconsistency problems in the dataset, such as missing data, zeroed values, duplicate samples, or noisy data. It also pointed to the fact that the features presented previous values in similar scales, so the need to perform special treatments for scale normalization was not required.

Certain features presented high correction among them. The features with multicollinearity were removed from the dataset in order to avoid instability problems during model training. Twelve features were removed from the dataset. The resulting dataset preserved the 20 features obtained by electrodes CZ, FZ, Fp1, F3, FC1, FC5, FT9, T7, CP5, P3, P7, PO9, PZ, O2, P4, CP6, FT10, FC6, F8, and Fp2.

Table 1 - Dataset unbalance analysis.

Class	Quantity of Samples	%
0	384,000	25.00
1	60,800	3.96
2	80,000	5.21
3	185,600	12.08
4	172,800	11.25
5	246,400	16.04
6	182,400	11.88
7	102,400	6.67
8	76,800	5.00
9	32,000	2.08
10	12,800	0.83
Total	1,536,000	100.0

Visual analysis showed that all the features presented a standardized normal distribution, and that the data were symmetrically distributed, indicating that the dataset was not affected by distortions in the data or by outliers. Upon these analyses, the dataset was considered ready for training the Machine Learning models.

3.3. Model Training

Eight Machine Learning models were selected to be trained with the SAM 40 dataset, aiming to compare the performance of different models and verify which of the models best adapts to the dataset. The selected models were: Naive Bayes, K-Nearest Neighbors, Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, eXtreme Gradient Boosting and Multilayer Perceptron. The scikit-learn framework and XGBoost library implementations were used. The models were trained with cross-validation (3-Fold) and tuning strategies. Additionally, a data augmentation strategy (increase in the number of samples) was used to train the Multilayer Perceptron model. Table 2 presents the accuracy rates obtained from model training.

Table 2 - Accuracy obtained from model training.

Model	Training Accuracy (%)	Test Accuracy (%)
Naive Bayes	11.2	11.2
K-Nearest Neighbors	36.7	11.1
Logistic Regression	9.9	9.3
Decision Tree	100.0	10.3
Random Forest	100.0	13.3
Gradient Boosting	19.4	13.8
eXtreme G. Boosting	48.2	14.5
Multilayer Perceptron	98.87	98.99

4. RESULTS AND DISCUSSION

The “traditional” Machine Learning models showed poor performance. An underfitting problem was observed when training Naive Bayes, K-Nearest Neighbors, Logistic Regression, Gradient Boosting and eXtreme Gradient Boosting models. An overfitting problem was observed when training the Decision Tree and Random Forest models. A plausible explanation for the poor performance of these models is that the SAM 40 dataset presents complex relations among its features and its classes are not linearly separable, which made it difficult to learn and generalize the data for such models.

The Multilayer Perceptron model obtained the best performance among all the models evaluated with an accuracy rate of 98.99% on the test set. A plausible explanation for the good performance of this model is the ability of neural networks to approximate complex functions and learn complex patterns from the data. An accuracy of 98.99% can be considered satisfactory in this scenario due to the complexity of the problem under study.

Table 3 presents the detailed metrics (precision, recall, and F1-Score) of the Multilayer Perceptron training. High rates were obtained for these metrics, which points to a good management of the model over the dataset class predictions, and its performance was not affected by false positives or false negatives.

Table 3 - Multilayer Perceptron training metrics.

Class	Precision	Recall	F1-Score
0	0.99	0.99	0.99
1	0.99	0.99	0.99
2	0.99	0.99	0.99
3	0.99	0.99	0.99
4	0.99	0.99	0.99
5	0.99	0.99	0.99
6	0.99	0.99	0.99
7	0.99	0.99	0.99
8	0.99	0.99	0.99
9	0.99	0.99	0.99
10	0.99	0.99	0.99

Fig 4 shows the Multilayer Perceptron architecture with which the best accuracy was reached. A 5-hidden layer architecture (with 100, 300, 500, 300 and 100 neurons) was used. Table 4 shows the hyperparameters setup used for training.

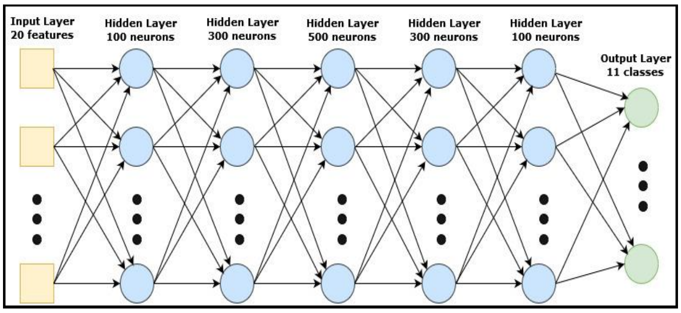

Figure 4 - Multilayer Perceptron architecture.

Table 4 - Hyperparameters setup for training.

Parameter	Value
Optimizer	Adam
Activation function	reLU
Learning rate	0.001
Epochs	300

5. CONCLUSIONS

Machine Learning techniques are an excellent tool for analyzing and learning data from sensors. Its use in problem-solving for the health area, such as stress level analysis, for example, expands the possibilities of providing technology whose application allows the discovery of insights about the problem at issue, and promotes the improvement of mental well-being as well as quality of life for individuals.

The Multilayer Perceptron proved to be a suitable Machine Learning model to deal with the problem under study by reaching a satisfactory performance in learning and predicting the SAM 40 dataset information. As future research, it is proposed to explore the training of other neural network architectures with the SAM 40 dataset in order to try to improve the accuracy rate on test set or to achieve good results with the use of less complex architectures as well as lower computational costs.

REFERENCES

1. Kang, M., Chai, K. (2022). Wearable sensing systems for monitoring mental health. Sensors, 22(3), pp. 994. doi: 10.3390/s22030994
2. Sheikh, M. et al. (2021). Wearable, environmental, and smartphone-based passive sensing for mental health monitoring. Frontiers in Digital Health, 3(1), pp. 662811. doi: 10.3389/fdgth.2021.662811
3. Can, Y. et al. (2019). Stress detection in daily life scenarios using smart phones and wearable sensors: a survey. Journal of Biomedical Informatics, 92(1), pp. 103139. doi: 10.1016/j.jbi.2019.103139
4. Hickey, B. A. et al. (2021). Smart devices and wearable technologies to detect and monitor mental health conditions and stress: a systematic review. Sensors, 21(10), pp. 3461. doi: 10.3390/s21103461
5. Long, N. et al. (2022). A scoping review on monitoring mental health using smart wearable devices. Mathematical Biosciences and Engineering, 19(8), pp. 7899-7919. doi: 10.3934/mbe.2022369
6. Luo, H., Gao, B. (2021). Development of smart wearable sensors for life healthcare. Engineered Regeneration, 2(1), pp. 163-170. doi: 10.1016/j.engreg.2021.10.001
7. Kulkarni, P. et al. (2022). Opportunities for smartphone sensing in e-health research: a narrative review. Sensors, 22(10), pp. 3893. doi: 10.3390/s22103893
8. Shahsavarani, A. M. et al. (2015). Stress: facts and theories through literature review. International Journal of Medical Reviews, 2(2), pp. 230-241.
9. Mcewen, B. S, Akil, H. (2020). Revisiting the stress concept: implications for affective disorders. Journal of Neuroscience, 40(1), pp. 12-21. doi: 10.1523/JNEUROSCI.0733-19.2019
10. Baran, K. (2021). Stress detection and monitoring based on low-cost mobile thermography. Procedia Computer Science, 192(1), pp. 1102-1110. doi: 10.1016/j.procs.2021.08.113
11 Natasha, P. et al. (2018). Detection of mental stress using EEG signals. International Journal of Engineering and Techniques, 4(1), pp. 323-331. doi: 10.29126/23951303/IJET-V4I1P46
12. Dham, V. et al. (2021). Mental stress detection using artificial intelligence models. Journal of Physics, 1950(1), pp. 012047. doi: 10.1088/1742-6596/1950/1/012047
13. iMotions. (2022). Electroencephalography: the complete pocket guide. Available: https://imotions.com/guides/ [Accessed: Jul. 30, 2022]
14. Kumar, J. S., Bhuvaneswari, P. (2012). Analysis of electroencephalography (EEG) signals and its categorization: a study. Procedia Engineering, 38(1), pp. 2525-2536. doi: 10.1016/j.proeng.2012.06.298
15. Garg, P. et al. (2021). Stress detection by machine learning and wearable sensors. ACM, pp. 43–45. doi: 10.1145/3397482.3450732
16. Zainudin, Z. et al. (2021). Stress detection using machine learning and deep learning. Journal of Physics, 1997(1), pp. 012019. doi: 10.1088/1742-6596/1997/1/012019
17. Mozgovoy, V. (2021). Longitudinal estimation of stress-related states through bio-sensor data. Applied Computing and Informatics, 17(1). doi: 10.1108/ACI-03-2021-0070
18. Kyamakya, K. et al. (2021). Emotion and stress recognition related sensors and machine learning technologies. Sensors, 21(7), pp. 2273. doi: 10.3390/s21072273
19. Ahuja, R., Banga, A. (2019). Mental stress detection in university students using machine learning algorithms. Procedia Computer Science, 152(1), pp. 349-353; doi: 10.1016/j.procs.2019.05.007
20. Kene, A., Thakare, S. (2022). Prediction of mental stress level based on machine learning. Machine Intelligence and Smart Systems, pp. 525-536. doi: 10.1007/978-981-16-9650-3_41
21. Priya, A. et al. (2020). Predicting anxiety, depression and stress in modern life using machine learning algorithms. Procedia Computer Science, 167(1), pp. 1258-1267. doi: 10.1016/j.procs.2020.03.442
22. Ghosh, R. et al. (2022). SAM 40: dataset of 40 subject EEG recordings to monitor the induced-stress while performing Stroop color-word test, arithmetic task, and mirror image recognition task. Data in Brief, 40(1), pp. 107772. doi: 10.1016/j.dib.2021.107772