Pejanović, V., & Radaković, M. (2024). Classification of cognitive stress as a psychological indicator through machine learning, *SCIENCE International journal, 3*(1), 139-143.
doi: 10.35120/sciencej0301139p UDK: 004.3:004.85:[159.923.2:159.944.4
159.95:[004.3:004.85

CLASSIFICATION OF COGNITIVE STRESS AS A PSYCHOLOGICAL INDICATOR THROUGH MACHINE LEARNING

Vladimir Pejanović[1], Milan Radaković[2]*

[1]Faculty of technical sciences, University of Novi Sad, Trg Dositeja Obradovića 6, Novi Sad, Serbia,
e-mail: vladimirpejanovic@uns.ac.rs
[2]Faculty of Sport, University "Union – Nikola Tesla", Narodnih heroja 30, Belgrade, Serbia,
e-mail: radakovic.milan@fzs.edu.rs

Abstract: In this study, we explored the potential of Support Vector Machine (SVM) method for classifying levels of cognitive stress using EEG (Electroencephalogram) signals. The goal is to develop accurate models that would enable the prediction and understanding of not only the current mental state of the subjects, but also potential real-time interventions. In medical fields, the application can be seen in the treatment of attention, focus, hyperactivity, autism, and depression disorders. Additionally, there is an extremely high potential for application in areas such as psychology, sociology, education, economics, neuromarketing, security, and in enhancing workplace stress management, anxiety treatment, digital marketing, economic-financial forensics, as well as improving user experience in virtual environments and video games. The results have shown that it is possible to differentiate high and low levels of cognitive stress with satisfactory accuracy, opening the way for the application of these findings in various fields. Cognitive stress represents one of the fundamental cognitive processes that causes individuals to behave and think differently in certain situations than in their usual state of consciousness. Predicting, analyzing, and understanding the level of cognitive stress from EEG signals is of great importance in various fields, including neuroscience, psychology, education, professional sports, human-computer interaction, and many other areas. Machine learning represents a subgroup of artificial intelligence that uses statistical models, and functions to 'learn' and 'train' data resulting in corresponding output values. The brain-computer interface, through which data on cognitive stress, among other parameters and psychological categories, is collected, is based on the functioning of EEG devices. The prediction of cognitive stress represents the application of machine learning, recording and using brain EEG signals or extracted characteristics from EEG signals as input values, in order to predict the level of output values of cognitive stress, of high or low degree, reflecting the mental state of the subjects in real time

Keywords: cognitive stress, machine learning, interdisciplinary application, brain-computer interface, prediction
Field: Social Sciences, and Humanities

1. INTRODUCTION

Machine learning represents a subgroup of artificial intelligence that uses statistical models, and functions to 'learn' and 'train' data, resulting in corresponding output values.

EEG has its use, both in medical and non-medical purposes, representing, with a certain degree of credibility, different mental states, inner experiences, and user experiences.

The prediction of cognitive stress represents the application of machine learning, noting and using, as input values, EEG brain signals or extracted features from the EEG signal, with the aim of predicting the output values of cognitive stress, i.e., neurological and physiological reactions, of a higher or lower degree, which reflect the mental state of the subjects in real time.

2. THE SUBJECT AND THE GOAL OF THE RESEARCH

Predicting cognitive stress or output prediction refers to the process of using machine learning algorithms and models to predict an individual's level of cognitive stress.

The primary goal of the research is to enhance the accuracy of developed prediction methods that can provide valuable insights or enable real-time responses based on the obtained levels of cognitive stress (Guger et al., 2021).

There are a large number of machine and deep learning algorithms that can be used for EEG data classification. Advances in technology that enable such analyses can have a strong impact on social sciences through application in various social contexts.

Processing EEG signals for the purpose of predicting both cognitive stress and other mental states and brain activities is becoming increasingly interesting in both scientific and commercial applications,

*Corresponding author: vladimirpejanovic@uns.ac.rs

and it refers to an interdisciplinary and multidisciplinary approach that includes social sciences such as psychology and sociology. Moreover, in non-medical contexts it has a wider use, in terms of multidisciplinarity, and helps in neuromarketing, brain-computer video games, monitoring mental state and human-computer interaction. The integration of artificial intelligence and machine learning with EEG technology opens a world of new possibilities, promising to reshape the way we understand the human brain and interact with the world around us. As research in this field continues to advance, society will benefit greatly from the numerous applications and progress offered by predictions using EEG signals as input values.

Brain-computer interface (BCI), the use of EEG signals and the application of machine learning techniques represent one of the not so common areas of science with the epithet 'multidisciplinary'. The importance of this topic can be viewed from the aspect of understanding how our most important organ functions. The human brain still represents a research challenge for scientists. Research of this, so to speak, microcosm, the most complex and still largely unexplored organ, can be achieved through various tools and methods that are available today and that relate to an interdisciplinary approach that includes social sciences.

Several metrics are used to assess the performance of prediction models, including accuracy, sensitivity, specificity, precision, Matthew's correlation coefficient, which are usually used to evaluate the results of predicting levels of cognitive stress for each class of output EEG signals.

3. RESEARCH METHODOLOGY

The subject of this research is the development of a machine learning based model for predicting the level of cognitive stress based on EEG signals.

The level of cognitive stress is divided into two classes, a class of lower or moderate stress (below 0.9) and a class of higher stress (above 0.9). The research aims to assess the precision and efficiency of the SVM (Support Vector Machine) model in predicting the level of cognitive stress based on EEG signals.

The Laboratory for Biomedical Engineering and Instrumentation at the Faculty of Technical Sciences of the University of Novi Sad has provided artificially generated EEG data with labeled levels of cognitive stress. In this way, interdisciplinarity and cooperation between technical and social sciences in the research of cognitive stress are encouraged.

Model training - The SVM model, a powerful supervised learning algorithm, was used for training the predictive model. This particularly gives a new aspect to the application of techniques from the field of artificial intelligence in social sciences, especially in psychology and sociology.

Preprocessing - Before entering the EEG data into the SVM model, several preprocessing steps were applied to improve the data quality and reduce noise.

4. COGNITIVE STRESS

Cognitive stress represents one of the fundamental cognitive processes that cause individuals to behave and think differently in certain situations than in the usual state of consciousness. Predicting, analyzing, and understanding the levels of cognitive stress obtained from EEG signals is of great importance in various fields, including neuroscience, psychology, education, professional sports, and human-computer interaction, as well as many other areas. An excessive amount of information represented in an extremely dynamic and fast environment is present in our everyday life, and the identification of neural processes in EEG signals associated with measuring cognitive stress can be useful for predicting, treating, and eliminating certain disorders and diseases related to the given cognitive process. Also, the identification and prediction of these processes and cognitive states, like cognitive stress, can be useful for an individual's interaction in a virtual environment.

Cognitive stress is one of the concepts in neuroscience, psychology, biology which is among the most interesting for researchers. The mentioned sciences strive to objectively and rationally detect and classify the processes of cognitive stress and its adaptation to internal and external stimuli, whether somatic or psychological (Blanco et al., 2019; GAILLARD, 1993; Staal, 2004). Staal described two traditional models for stress: The first, caused by stimulation and the other, caused by response (Staal, 2004). The stimuli cause model defines the concept of stress as the impact od different influences ('stressors') that destabilize the 'normal' functioning of an individual (Staal, 2004). Unlike the stimulus-based model, the other model defines the concept of stress by distinguishing types of actions, activities (behavioral, cognitive, and affective) that are a consequence of the impact of stress (Staal, 2004). Current research has not established comprehensiveness for either of these two models, so a third model has been developed. That model defines the concept of stress as a result of mismatch between an individual's awareness of task demands and their awareness of the resource capacities needed to cope with the given tasks (Blanco et al., 2019; GAILLARD, 1993; Staal, 2004).

Furthermore, the same task that causes high cognitive stress in one individual may be routine and stress-free for another. Multiple aspects of the problem have been considered to determine the optimal threshold. The process of reaching the most optimal threshold value using experiences from other fields has been analyzed. It has been determined that many approaches to determining the optimal threshold value that have been applied in other fields can be used directly within the boundaries of the obtained result values (Vizer et al., 2009; Shanahan & Roma, 2003). Optimizing threshold values is a problem of upmost importance (Vizer et al., 2009; Shanahan & Roma, 2003).

A stress level threshold of 0.9 was taken, which is an extremely low level on most stress scales, assuming that higher boundary values of the results indicate a higher level of stress (Crosswell & Lockwood, 2020; Toolbox | UCSF SMN, n.d.). Specifically, a score of 0.9 is considered low on most stress scales due to its scaling. In another context, for example, in some clinical or research contexts, a score of 0.9 may indicate a significant level of stress that requires attention or intervention. Some scales have a narrow range (e.g., from 0 to 10), while others have a wider range (e.g., from 0 to 100). On a scale with a narrow range, a score of 0.9 would be moderate to low stress. The device that made the assessment of cognitive stress based on the database we had available used a narrow scaling range. Taking all this into account, as well as the levels of cognitive stress recorded in the data files, we believe that the level, i.e., the boundary we have chosen to distinguish high from low levels of cognitive stress is optimal.

Simulators of brain activity, such as the SEREEGA package (Simulating Event Related EEG Activity; SEREEGA is an open-source tool package based on Matlab intended for generating simulated EEG data), or the European Union's Human Brain Project (HBP), and MindScope at the Allen Brain Institute, are becoming increasingly present and significant in scientific research (Krol et al., 2018; Einevoll et al., 2019). This further emphasizes the importance of developing tools and projects that enable a better understanding of brain activity, which, on the other hand, can have applications and significance in fields such as psychology and education.

5. DESCRIPTION OF THE MODEL BASED ON SUPPORT VECTOR MACHINE

In the field of biomedical research, Support Vector Machine (SVM) methods are used in the classification of microarray gene expression profiles (Noble, 2006). SVM is also used in other fields with exceptional success, for example, in the field of financial-economic forensics, in machine learning for detection and recognition of the occurrence of false credit card numbers by examining thousands of financial reports on the use of both false and real credit cards (Noble, 2006). Also, SVM can be trained to detect handwritten numbers, based on a database of a large number of scanned images with handwritten zeros and ones (Noble, 2006). The Support Vector Machine method is trained and learned to distinguish between two classes, in line with its epithet of binary classification, and taking into account the unlabeled parameter, i.e., the input value, it predicts to which of the two classes it will correspond (Noble, 2006).

The generally accepted term in literature for a plane that separates data sets located in a multi-dimensional space is called a hyperplane. The hyperplane, essentially, represents a line separating two classes, i.e., two groups of data (Noble, 2006). In the research on this topic, attempts are made to develop an SVM model that can handle errors in the data by allowing some of the expression profiles, as anomalies, to fall on the wrong side of the hyperplane (Noble, 2006). However, the soft margin in the SVM model allows some data points to pass through the margin or the boundary value of the hyperplane without the final result being compromised by errors (Noble, 2006).

Classification in SVM is an example of supervised learning. Known labels help to show whether the system is working correctly or not. This information indicates the desired response, confirming the accuracy of the system, or is used to help the system learn to act in the correct, desired way. The SVM classification process involves identifying those input values that are associated with the established two classes (Cristianini & Shawe-Taylor, 2000). Feature selection and SVM classification, as two significant processes in machine learning. They have application even when predicting unknown input values is not used (Cristianini & Shawe-Taylor, 2000; Guger et al., 2021). They can be used for detection and identification of data groups involved in any process that uses class differentiation (Cristianini & Shawe-Taylor, 2000; Guger et al., 2021).

The disadvantages of SVMs are that they are not suitable for processing large amounts of data, and they do not perform best if there is more noise in the dataset (Hammad et al., 2020; Support Vector Machines | Dremio, 2023; Pan et al., 2008; Cooney et al., 2019; Guger et al., 2021). The biggest disadvantage of the

SVM model includes, as can be concluded from its binary data processing approach, the fact that SVM supports only binary classification problems. A certain modus operandi of multiclassification is, however, possible in the SVM model (Noble, 2006). The simplest approach to this problem is to train multiple one-against-all classifiers.

6. RESEARCH RESULTS

The investigation of the SVM classifier potential for classifying low and high levels of cognitive stress based on EEG signals was conducted using a Matlab software tool called Matlab Online. It is necessary at the beginning of the code to perform a data type conversion, i.e., reading from .xlsx to .mat extension, which involves converting string values to numerical values. This represents a significant precondition for efficient and functional code, as well as for valid and precise results.

The following features (feature extraction) of EEG data were observed in the development of the Support Vector Machine model: Mean value method (mean feature), Centroid method (centroid feature), Spectral feature extraction method, and Higuchi Fractal Dimension (HFD) method. The largest percentage increase in accuracy was achieved by introducing the mean value feature. The next largest percentage increase, which followed, was thanks to the spectral feature extraction method, as well as the centroid method. A significantly smaller percentage increase in precision was obtained by introducing the Higuchi Fractal Dimension (HFD) method.

7. CONCLUSIONS

The potential of the Support Vector Machine method in classifying cognitive stress levels based on artificially generated EEG signals was examined in the paper. The obtained results show that it is possible to make a classification between two levels of stress, low and high, with an average accuracy of 75.33% and specificity of 93.93%. A significant improvement was achieved through the selection of appropriate EEG data features compared to just using preprocessed EEG data as input signals in the developed model. Binary classification, on which the SVM model is based, shows that SVM is a good choice for the classification process and for presenting output prediction values for easier understanding of the results by the user. The research results can be applied in medical, but also non-medical fields. As for non-medical fields, it can be applied for various purposes such as: improving learning, stress management especially at the workplace, treatment of anxiety disorders, and in various situations and environments that cause the occurrence of cognitive stress in individuals. In the case of non-medical uses, the most interesting is the application in virtual environments and video games, where the system can read and/or predict the brain parameters of the user and by feedback system affect the user himself, contributing to a greater user experience and a more realistic environment in virtual reality or video game.

8. LITERATURE

Blanco, J., Vanleer, A., Calibo, T., & Firebaugh, S. (2019, January 25). Single-Trial Cognitive Stress Classification Using Portable Wireless Electroencephalography. Sensors, 19(3), 499. https://doi.org/10.3390/s19030499

Cinciripini, P. M. (1986, November). Cognitive stress and cardiovascular reactivity. II. Relationship to atherosclerosis, arrhythmias, and cognitive control. American Heart Journal, 112(5), 1051–1065. https://doi.org/10.1016/0002-8703(86)90320-0

Cooney, C., Folli, R., & Coyle, D. (2019, October). Optimizing Layers Improves CNN Generalization and Transfer Learning for Imagined Speech Decoding from EEG. 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). https://doi.org/10.1109/smc.2019.8914246.

Cristianini, N., & Shawe-Taylor, J. (2000, March 23). An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. https://doi.org/10.1017/cbo9780511801389

Crosswell, A. D., & Lockwood, K. G. (2020, July). Best practices for stress measurement: How to measure psychological stress in health research. Health Psychology Open, 7(2), 205510292093307. https://doi.org/10.1177/2055102920933072

Einevoll, G. T., Destexhe, A., Diesmann, M., Grün, S., Jirsa, V., de Kamps, M., Migliore, M., Ness, T. V., Plesser, H. E., & Schürmann, F. (2019, May). The Scientific Case for Brain Simulations. Neuron, 102(4), 735–744. https://doi.org/10.1016/j.neuron.2019.03.027

GAILLARD, A. W. K. (1993, September). Comparing the concepts of mental load and stress. Ergonomics, 36(9), 991–1005. https://doi.org/10.1080/00140139308967972

Guger, C., Allison, B. Z., & Gunduz, A. (Eds.). (2021). Brain-Computer Interface Research. SpringerBriefs in Electrical and Computer Engineering. https://doi.org/10.1007/978-3-030-79287-9

Hammad, M. A., Jereb, B., Rosi, B., & Dragan, D. (2020, February 1). Methods and Models for Electric Load Forecasting: A Comprehensive Review. Logistics & Sustainable Transport, 11(1), 51–76. https://doi.org/10.2478/jlst-2020-0004

Harding, T., & Zimmermann, E. (1989, July). Psychiatric Symptoms, Cognitive Stress and Vulnerability Factors. British Journal of Psychiatry, 155(1), 36–43. https://doi.org/10.1192/bjp.155.1.36

doi: 10.35120/sciencej0301139p UDK: 004.3:004.85:[159.923.2:159.944.4
159.95:[004.3:004.85

Krol, L. R., Pawlitzki, J., Lotte, F., Gramann, K., & Zander, T. O. (2018, May 18). SEREEGA: Simulating Event-Related EEG Activity. https://doi.org/10.1101/326066

Mandić, S., Pejanović, V., Đerić, J., Evetović, N., Miljuš, S., & Sovilj, P. (2023). Development and experimental platform in teaching activities in the field of brain computer interface. Vrnjačka Banja, Serbia: XXIX Conference Development Trends: University in front of new challenges.

Nixon, R. D., Nehmy, T., & Seymour, M. (2007, November). The effect of cognitive load and hyperarousal on negative intrusive memories. Behaviour Research and Therapy, 45(11), 2652–2663. https://doi.org/10.1016/j.brat.2007.06.010

Noble, W. S. (2006, December). What is a support vector machine? Nature Biotechnology, 24(12), 1565–1567. https://doi.org/10.1038/nbt1206-1565

Pan, Y., Jiang, J., Wang, R., & Cao, H. (2008, July). Advantages of support vector machine in QSPR studies for predicting auto-ignition temperatures of organic compounds. Chemometrics and Intelligent Laboratory Systems, 92(2), 169–178. https://doi.org/10.1016/j.chemolab.2008.03.002

Pejanović, R., & Vujić, V. (2016). The methodology of economic research and design work on academic studies. Novi Sad, Serbia: Akademska knjiga,.

Pejanović, V., Mandić, S., Đerić, J., Evetović, N., Miljuš, S., & Sovilj, P. (2023). Advanced tools in education research: the brain-computer interface and the stroop effect. Vrnjačka Banja, Serbia: XXIX Conference Development Trends: University in front of new challenges.

Rosa, H. (2013). Social Acceleration. Columbia University Press.

Schleifer, L. M., Ley, R., & Spalding, T. W. (2002, April 11). A hyperventilation theory of job stress and musculoskeletal disorders*. American Journal of Industrial Medicine, 41(5), 420–432. https://doi.org/10.1002/ajim.10061

Shanahan, J. G., & Roma, N. (2003). Improving SVM Text Classification Performance through Threshold Adjustment. Machine Learning: ECML 2003, 361–372. https://doi.org/10.1007/978-3-540-39857-8_33

Slijepčević, P. (2018). Svetac i grešnik, kako zloupotrebljavamo nauku i budućnost čovječanstva. Novi Sad, Serbia: Akademska knjiga

Staal, M. A. (2004). Stress, cognition, and human performance: A literature review and conceptual framework. Retrieved from https://human-factors.arc.nasa.gov/flightcognition/Publications/IH_054_Staal.pdf

Support Vector Machines | Dremio. (2023, December 29). Dremio. Retrieved from https://www.dremio.com/wiki/support-vector-machines/

Toolbox | UCSF SMN. (n.d.). UCSF SMN. Retrieved from https://www.stressmeasurement.org/measurement-toolbox

Vizer, L. M., Zhou, L., & Sears, A. (2009, October). Automated stress detection using keystroke and linguistic features: An exploratory study. International Journal of Human-Computer Studies, 67(10), 870–886. https://doi.org/10.1016/j.ijhcs.2009.07.005