Predicting the Health Impacts of Commuting Using EEG Signal Based on Intelligent Approach

Mhd Saeed Sharif
Computer Science and DT, ACE, UEL,
University Way, E16 2RD, London,
UK.
S.sharif@uel.ac.uk

Madhav Raj Theeng Tamang
Computer Science and DT, ACE, UEL,
University Way, E16 2RD, London,
UK.
u1430774@uel.ac.uk

Cynthia Fu
Department of Psychological Sciences,
UEL, Water Ln, E15 4LZ, London,
UK.
C.Fu@uel.ac.uk

Abstract— **Commuting to work is an everyday activity for many which can have a significant effect on our health. Commuting on regular basis can be a cause of chronic stress which is linked to poor mental health, high blood pressure, heart rate, and exhaustion. This research investigates the neurophysiological and psychological impact of commuting in real-time, by analyzing brain waves and applying machine learning. The participants were healthy volunteers with mean age of 30 years. Portable electroencephalogram (EEG) data were acquired as a measure of stress level. EEG data were acquired from each participant using non-invasive NeuroSky MindWave headset for 5 continuous activities during their commute to work. This approach allowed effects to be measured during and following the period of commuting. The results indicate that whether the duration of commute was low or large, when participants were in a calm or relaxed state the bio-signal alpha band exceeded beta band whereas beta band was higher than alpha band when participants were stressed due to their commute. Very promising results have been achieved with an accuracy of 97.5% using Feed-forward neural network. This work focuses on the development of an intelligent model that helps to predict the impact of commuting on participants. In addition, the result obtained from the Positive and Negative Affect Schedule also suggests that participants experience a considerable rise in stress after their commute. For modelling of cognitive and semantic processes underlying social behavior, the most of the recent research projects are still based on individuals, while our research focuses on approaches addressing groups as a complete cohort. This study recorded the experience of commuters with a special focus on the use and limitation of emerging computing technologies in telehealth sensors.**

Keywords—machine learning, wearable sensor, commuting, stress recognition, electroencephalogram.

I. INTRODUCTION

The commute to work involves a significant amount of time and effort in our daily work activities, impacting us both physically and mentally; it can reduce workplace performance and even increase absenteeism. Vitality Health, the University of Cambridge, RAND Europe, and Mercer produced a survey of more than 34,000 staff across all UK industries that explored the effect of commuting, along with versatility and research on employee health and productivity [1]. Longer commutes are associated with a negative effect on mental well-being, such as depression, financial worries, work-related stress, and sleep disorders. Similarly, commuters develop higher psychosomatic conditions, such as headache, back pain, digestive issues, high blood pressure and psychological disorders, when compared to short-distance commuters. It is particularly difficult for the commuters driving in adverse weather, traffic congestion and accidents as these all induce stress, which can lead to burnouts, depression, and illness [2]. There are various methods for recognizing stress levels that have been formulated; some are focused on physiological signals: variability of heart rate, blood pressure, cortisol, whilst most are based on encephalography (EEG) technology to identify stress levels. There are many reasons why EEG is an outstanding method for the study of human behavioural neurocognitive processes. EEG is used to rule out various health-related conditions like seizure, epilepsy, encephalitis, brain tumor, memory problem, stress, sleep disorders, stroke, and dementia [3]. In addition, EEG signals recorded from the prefrontal cortex are used in asymmetric analysis of the frequency-band powers in the various stress related research [4]. Therefore, electroencephalogram, that indicates the electrical activity in the brain plays vital role to detect and assess the mental stress.

EEG is useful to quantify electrical activity from the brain, as it shows how electric signals communicate with each other across various neurons in human brain network and is the neurological method of choice for recording brain-generated electrical activity through electrodes mounted on the surface of the scalp [5]. The EEG signal is of a highly dynamic and extremely random type and non-linear in nature.

Furthermore, EEG signals helps to identify human emotions, mental workload, vigilance, etc. This technology is broadly applied in the diagnosis of different types of mental diseases. Also, it's application has been done in various field such as bioengineering, psychophysiology, psychology, video games [6].

The signal, however, is often a combination of many underlying frequencies which are intended to represent those cognitive, affective, or attentive phases. EEG signals are categorized in five frequency bands and its associated to various activities as shown in Table I [7].

TABLE I. EEG FREQUENCY BAND AND ACTIVITY

Band	Frequency	Activity
Delta	0.5 – 4 Hz	Deep Sleep
Theta	4 – 8 Hz	Drowsiness, Light Sleep
Alpha	8 -13 Hz	Relaxed
Beta	13 – 30 Hz	Active Thinking, Alert
Gamma	More than 30 Hz	Hyperactivity

II. RELATED WORK

Individuals drive long distances to work show less frequent involvement in strong to intense physical activity and decreased cardiorespiratory fitness (CRF), increased body mass index (BMI) and blood pressure [8]. The increased strain that commuting imposes on individuals and their families may outweigh the job and the benefits they might achieve. Assessing and managing stress is difficult because each person experiences stress in different ways in their everyday life. Measures of stress have often been questionnaire-based, such as Cohen's Perceived Stress Scale, and Stress Response Inventory [9].

Over the past three decades, machine learning has made enormous progress due to expeditious advancement in computer technology, efficient machine learning algorithms and most notably, big data. Although contemporary approaches to automated EEG interpretation have employed ultra-modern machine learning algorithms using high-dimensional models, these approaches had not previously been used in EEG analysis due to a lack of large databases to train these systems.

The extraction of features remained the most significant step in EEG data classification. The most common methods for features extraction used are wavelet analysis, discrete wavelet transform, fourier analysis and statistical analysis. These features can be classified using well-known classifiers, such as SVM, KNN, random forest, ANN, Bayesian modelling, k-means clustering [10].

Such research focuses on a methodology that is simple, effective, and easy to execute in hardware to evaluate commute effect on work productivity using EEG signal analysis. A considerable amount of research had focused on developing processing pipelines for cleaning, extracting relevant features, and classifying EEG data. SVM is widely used and has been proven to be one of the most accurate classification techniques using EEG signals.

The development of emotion recognition systems discusses the use of features derived from EEG signals to create an emotion recognition system, which has been focused on a model of Arousal or Valence. EEG signals are decomposed into various frequency bands using Discrete Wavelet Transform (DWT). Energy and entropy are calculated and used for further classification methods. Emotional states were identified with high accuracy, k-nearest neighbour with 78% accuracy, support vector machine with 91.3%, and artificial neural network with 82% accuracy, and energy and entropy were extracted and used for further classification methods [11].

In a study, an EEG signal were classified applying the wavelet transform and SVM to detect epileptic seizure activity from background electroencephalographs [12]. Various features like energy, entropy and standard deviation were calculated to classify the EEG. In that work, accuracy of nearly 91.2% was acquired in classification between abnormal and normal signal.

Many studies have exhibited higher accuracy in analysing EEG analysis using a Decision Tree classifier. Hybrid methodology of fast Fourier transform and decision tree classifier have been used to detect epileptic seizure [13]. It was found out that the classifier successfully obtained an accuracy of 98.7% [14]. Using an automated Brain Computer Interface feature, a machine learning approach based on Decision Tree (DT) to automate the feature selection showed almost 90% of accuracy [15].

Artificial neural network (ANN) is another powerful algorithm in the analysis and data mining process. In data classification, an artificial neural network showed a greater positive predictive value to identify epilepsy than alternative methods [16]. Guo et al. [17], used various automation and computer techniques that were employed to develop a neural network classifier that accurately identifies whether the subject being associated is normal or epileptic with overall classification accuracies from 97.8 to 99.6%.

Emotions can be classified using relative wavelet energy for pre-processing and normal and depressed patients are classified using Artificial feedforward neural network. Using ANN with 20 features, overall 98.11% accuracy was achieved [18].

Similarly, K-nearest neighbours (KNN) algorithm is one of the popular supervised machine learning algorithms which can be used for classification and regression problems. It is easy to implement and understand and can work with good speed on a smaller size of data. Rahman et al. [19] provides an overview of the impact of mental workload on mental stress in humans through the EEG signal selection function and classification to identify mental stress. Several EEG signals were obtained from people with specific mental workload and analysed using spectrum analysis and extracted characteristics using the classification KNN. Using KNN classifier the mean power of the beta band was selected as a function for classifying mental stress. With a maximum accuracy of 91.26%, the assimilation of mean power of beta band and KNN classifier significantly outperforms the conventional Fuzzy K-Nearest Neighbour classifier (FKNN) with maximum accuracy of 81.98% and the more methods previously proposed for the mental stress recognition.

For classification, Hu et al. [20] proposed a procedure that uses a combination of correlation-based features with the data mining algorithm k-nearest neighbour. It was found that for the valence aspect split up into 3 classes, 83% of correct classification rate (CCR) was achieved on attention recognition in EEG.

III. METHODOLOGY

Our research study focuses on the development of an intelligent model that can predict the impact of commuting on participants during their commute to work by applying the machine learning techniques; KNN, SVM, decision tree and Feed-forward neural network. These techniques were applied to the quantitative data from EEG signal and qualitative data collected from the PANAS form.

Support vector machine is a supervised algorithm which provides data analysis for classification and regression. The basic concept behind SVM is to find a hyperplane that divides the datasets into groups. Support vectors are the points which lie close to all groups. This technique is widely used as it gives good accuracy with less computation power. Similarly, KNN algorithm is one of the popular supervised classification techniques. In this technique, an object is categorized by its neighbour's majority vote, assigning the object to the most accepted class of its nearest k neighbour. This model uses an entire dataset as training set.

Decision tree is mainly used as non-linear classifier. It can be used in big data too due to its rapid and straightforward classification and training processes. This classifier is like tree chart in which every decision retrieves a particular class before achieving the confirmed class. Test outcome is represented by branch, a node represents a test on a feature utilizing the transformation rules and a class category is kept by the leaf.

Feed-forward neural network is a computational deep learning model which is massive and powerful in calculation. This network is formed of three interconnected layers made up of neurons; input layer, hidden layer, and output layer. This network uses data sample to acquire the relevant information.

The following steps were implemented to carry out this experiment. Fig. 1 represent the flow chart of this study.
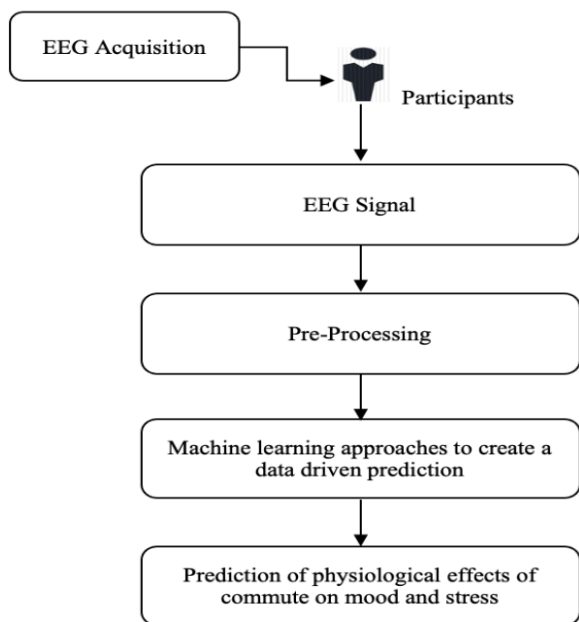

Fig. 1. Data collection and study design

A. Subject Selection

16 healthy participants were recruited. Each participant commutes to work for 5 working days. The contributors were healthy, and they provided an informed consent.

B. Measurement Protocol

Data were collected during the commute to work from participants using non-invasive wearable technologies. NeuroSky MindWave EEG headset was used to collect the EEG signal during the commute as illustrated in Fig. 2. The

eegID application was utilized to collect the obtained EEG values from the device.
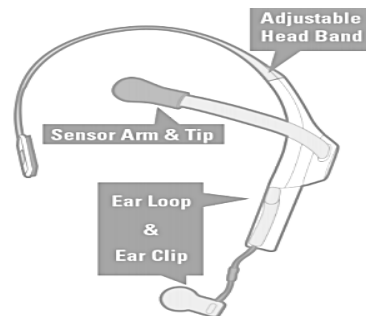

Fig. 2. EEG Device.

A questionnaire form PANAS has been employed as a subjective self-report quantitative measure. Each participant completed the required questions for 5 consecutive days before and after commute to work. This method is applied extensively to measure the positive and negative affect in community and clinical environment [21]. This form is comprised of various words representing feelings and emotions as show in the Table II.

TABLE II.   PANAS FORM.

1 Very Slightly or Not at All	2 A Little	3 Moderately	4 Quite a Bit	5 Extremely
_____ 1. Interested			_____ 11. Irritable	
_____ 2. Distressed			_____ 12. Alert	
_____ 3. Excited			_____ 13. Ashamed	
_____ 4. Upset			_____ 14. Inspired	
_____ 5. Strong			_____ 15. Nervous	
_____ 6. Guilty			_____ 16. Determined	
_____ 7. Scared			_____ 17. Attentive	
_____ 8. Hostile			_____ 18. Jittery	
_____ 9. Enthusiastic			_____ 19. Active	
_____ 10. Proud			_____ 20. Afraid	

C. Data Processing

In this research, four machine learning techniques has been applied to objectively verify our proposed hypothesis as "While the average of Alpha band of EEG Power Spectrum exceeds that of the Beta band while in a relaxed state of mind whereas the Beta band exceeds Alpha when in a stressful state of mind".

The classifiers selected for the proposed approach are SVM, KNN, Decision Tree and Artificial Neural Network. For each of the machine learning algorithm, the data was initially processed before using it as training and testing data to validate the proposed research hypothesis. To fulfil the hypothesis two datasets have been prepared by choosing related parameters that will yield a concise effect of stress. The first dataset was created from the main objective parameters such as alpha and beta frequency bands from EEG signals whereas, the other dataset was comprised of key objective values and other subjective data acquired from the subjects such as height, weight, age, alcohol intake per week, cigarettes

per day, weather (temperature). The target data points were represented by binary numbers. In the target data values, 0 indicates participant in relax state after commute, and 1 means stressed after commute. If the average of Beta band exceeds average of Alpha band, the participant is considered as stressed after commute. Meanwhile, when the average of Beta is lower than the average of Alpha band, it indicates participant is in relax state.

IV. IMPLEMENTATION

Four different techniques are used to design an intelligent approach which predicts the impact of commute. KNN, SVM, decision tree and feed-forward neural network are four different techniques applied in the data to acquire the result that can be used to test our research hypothesis. A combination of four classifiers were used in this research to enhance the classification accuracy of stress analysis. Accuracy is the primary operation of model performance, which is how accurately the predictions of the model match to the examined data. However, it is assumed that the more numbers of parameters in the model the more chances to overfit. Hence, the accuracy of each model must be measured properly. It has been verified that the accuracy of model evaluated on the training data which is an overall basic performance of the model. Therefore, a small part of dataset needs to be collected as a validation dataset for model accuracy evaluation. In addition, we have applied the fivefold cross validation to estimate how the model is expected to perform in general when used to make prediction on data not used during the training of the model. All four techniques were implemented using the software MATLAB. It is very popular computing software used by many researchers for data analysis, algorithms development, and for creating machine learning models. In addition, the classifier model was trained via a fivefold cross-validation method which helps to protect against the overfitting of data. There are many parameters in machine learning to check model performance which are discussed below.

A. Cross Validation

This technique can be employed in applied artificial intelligence specifically to evaluate the capability of a machine learning model to learn on unattended datasets. It uses a small proportion of data to approximate how the model is supposed to normally operate when used to make observations about data which is not being used during training phase. This procedure is to split the sample into k folds and then generalize the performance and enhance data records, thereby increasing the model's efficiency Therefore, it is referred as *k*-fold validation [22]. The dataset will be divided into k-number of subsets that is called folds. Then training will be performed on all dataset leaving only one that is (k-1) subset to evaluate performance of trained model. Due to this overfitting can be reduced during training phase. We can choose splits of 3, 5, 10 or any K number. There are several techniques with which we can build these folds. It is necessary to choose proper number of k for data sample otherwise it may result into misinterpretation of the given dataset. The value of k is usually 5 or 10. The value 10 is broadly recommended if we are confused to select a value for the dataset. When training the model Cross validation reduces

overfitting of data, while no validation method does not guarantee of overfitting.

B. ROC/AUC Curve

Measuring efficiency is an important activity at Machine Learning. So, we can count on an AUC-ROC Curve when it comes to binary classification task performance. ROC (Receiver Operating Characteristics) and AUC (Area Under the Curve) curve helps to evaluate how well a model can perform probability between positive and negative classes at different threshold setting. ROC is a prediction curve, and AUC is a level of separability assessment. This is one of the most significant performance metrics for verifying the efficiency of any classification model [23, 24]. ROC has two parameters True Positive Rate and False Positive Rate. The graph of ROC is plotted against the FPR with TPR, where FPR and TPR are plotted on the x-axis and y-axis respectively. TPR indicates ratio observation correctly predicted to be positive. The larger AUC values the better model performance at prediction right values. AUC values range between 0 to 1. If AUC value is 1 this shows 100% correct prediction, 0 means 100% wrong prediction of the model and AUC with 0.5 value means it doesn't have distinguish capacity of prediction [25].

TPR = True positive values / True Positive +False Negative
And False Positive Rate which is incorrectly predicted observation to be positive.
FPR =   False Positive / True Negative + False Positive

C. Confusion Matrix

This is a performance metric for classification of machine learning tasks where results can be two or multiple classes. It is a table of four main expected and real combinations of values. The number of observations that are true and false is summarized with count values and broken down by class. It gives us a detail of the model classification not only into the mistakes made by a classifier but also the types of error made.

Fig. 3.   Confusion matrix

This Fig. 3 is of confusion matrix contains two classes actual values against predicted values where observations classes are P (Positive values) and N (Negative values).

TP = Real and estimated positive values are similar.
TN = Real and estimated negative values.
FP = Real observation stands negative but predicted value is positive.
FN = Actual observation is positive, but prediction value holds negative.

This technique will further help to analyse more powerful classification statistics, such as classifier's precision, specificity, sensitivity, and recall [26].

V. RESULT AND ANALYSIS

We applied the most effective machine learning techniques to optimize our hypothesis, "While the average of Alpha band of EEG Power Spectrum exceeds that of the Beta band while in relaxed state of mind whereas the Beta band exceeds Alpha when in stressful state of mind". The performance of our stress analysis method was evaluated by creating two datasets and numerous approaches were tested to educate and assess the datasets using five-fold cross validation and no validation techniques. The confusion matrix and ROC curves were examined to check efficiency of the model. The outcome of each algorithm is compared using visualization graph, confusion matrix and ROC curves to measure efficiency of the mode.

We achieved the accuracy of 97.5% using the feed-forward neural network as presented in Fig. 4 confusion matrix. Out of 80 predicted values, two values were misclassified.

Fig. 4. Evaluation of model using the Confusion matrix

Similarly, ROC curve was plotted for Feed forward neural network as shown in Fig. 5 to test the true positive rate against the false positive rate.
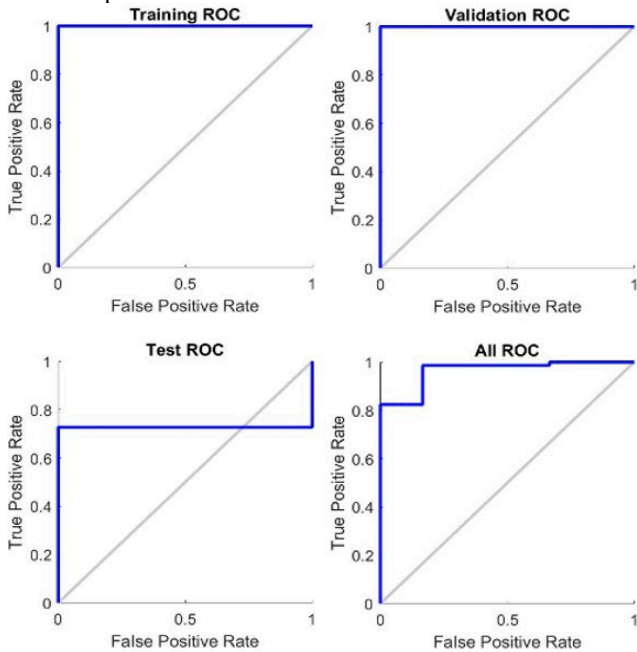

Fig. 5. Plotting the TPR against FPR using ROC curve

a) Performance Comparision Based on Confusion Matrics

The deployed matrix for each of the machine learning techniques is plotted against the true positive ratio positive estimated values, false negative ratio, false discovery rates, and the overall observations against the real class and the estimated values for the class. The results in Table III show the accuracy levels achieved from all the techniques applied on both datasets.

TABLE III. CLASSIFICATION ACCURACY USING CONFUSION MATRIX WITH AND WITHOUT VALIDATION.

Algorithm	Dataset	Validation	Accuracy
KNN	Dataset I	5-fold cross-validation	97.5%
		No validation	96.0%
	Dataset II	5-fold cross-validation	92.5%
		No validation	95.0%
SVM	Dataset I	5-fold cross-validation	96.3%
		No validation	97.0%
	Dataset II	5-fold cross-validation	93.8%
		No validation	98.8%
Decision Tree	Dataset I	5-fold cross-validation	91.3%
		No validation	97.5%
	Dataset II	5-fold cross-validation	97.5%
		No validation	98.8%
Feed-forward neural network	Dataset I	No validation	97.5%
	Dataset II	No validation	98%

b) PANAS Result

Each participant was asked to complete the form before and after the commute to work. PANAS form comprises 20 items which demonstrate how the participant feels. A 5-point Likert scale is used for scoring.

Scoring Instructions:
For positive affect, we added score values at one, three, five, nine, ten, twelve, fourteen, sixteen, seventeen, and nineteen at Table II. Values range is between 10 to 50. Similarly, negative affect is calculated by adding up the scores on lines 2, 4, 6, 7, 8, 11, 13, 15, 18, and 20 from Table II. Negative affect scores range between 10 to 50. The average of the positive value of affect and the negative ones for both cases of commuting are illustrated in the Table IV.

TABLE IV. POSITIVE AND NEGATIVE AFFECT AFTER COMMUTE

Avg Pre-Positive Affect	Avg Post-Positive Affect	Avg Pre-Negative Affect	Avg Post-Negative Affect
33.75	30.88	12.9	14.92

The result at the table above shows that participant's positive affects were higher before commute which indicates they were stressed after commute. Also, the post-commute negative affect was higher than the baseline negative affect, which indicates the participants showed increase stress following the commute.

VI. Conclusions

We investigated the impact of commute on human bio signal through the development of various intelligent techniques. We designed a model that could predict the commute impact on our bio signals. Multiple algorithms analysing EEG signals of participants revealed that Beta band of EEG spectrum exceeded that of Alpha band during the commute. Also, we used the PANAS form to validate the effects of commute (as this varies subjectively from one participant to another). This questionnaire helped to acquire the needed subjective data from each participant.

We researched and selected the best and most suitable algorithm for both datasets. Using the Feedforward network, an accuracy level of 97.5 % in the objective data, that comprised of key parameters. The other data set has an accuracy level of 98%. This data comprised of the key values and other data acquired from the subjects using PANAS form. The developed neural network provided a high level of accuracy in comparison to the other three techniques used throughout the research. The quantitative result achieved from the machine learning approaches confirmed the proposed hypothesis "While the average of Alpha band of EEG Power Spectrum exceeds that of the Beta band while in relaxed state of mind whereas the Beta band exceeds Alpha when in stressful state of mind". Similarly, the result obtained from the PANAS form also suggests that participants experience a considerable rise in stress after their commute.

The findings in this study helped to create an intelligent model which can help us to predict the health impacts of regular commuting on people using EEG signal. Our future work would be focusing on the generalisation of the developed model across all London and other cities.

References

[1] Ma, L. and Ye, R., 2019. Does daily commuting behavior matter to employee productivity?. *Journal of Transport Geography*, 76, pp.130-141.

[2] A. Sano and R. W. Picard, "Stress Recognition Using Wearable Sensors and Mobile Phones," 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, 2013, pp. 671-676, doi: 10.1109/ACII.2013.117.

[3] Blocka, K., 2018. *EEG (Electroencephalogram): Purpose, Procedure, and Risks*. [online] Healthline. Available at: <https://www.healthline.com/health/eeg>.

[4] Ahn, J., Ku, Y. and Kim, H., 2019. A Novel Wearable EEG and ECG Recording System for Stress Assessment. *Sensors*, 19(9), p.1991.

[5] S. A. Awang, M. P. Paulraj and S. Yaacob, "Implementing eigen features methods/neural network for EEG signal analysis," 2013 7th International Conference on Intelligent Systems and Control (ISCO), 2013, pp. 201-204, doi: 10.1109/ISCO.2013.6481149.

[6] X. Hou, Y. Liu, O. Sourina, Y. R. E. Tan, L. Wang and W. Mueller-Wittig, "EEG Based Stress Monitoring," 2015 IEEE International Conference on Systems, Man, and Cybernetics, 2015, pp. 3110-3115, doi: 10.1109/SMC.2015.540.

[7] N. Sulaiman, B. S. Ying, M. Mustafa and M. S. Jadin, "Offline LabView-based EEG Signals Analysis for Human Stress Monitoring," 2018 9th IEEE Control and System Graduate Research Colloquium (ICSGRC), 2018, pp. 126-131, doi: 10.1109/ICSGRC.2018.8657606.

[8] Hoehner, C., Barlow, C., Allen, P. and Schootman, M., 2012. Commuting Distance, Cardiorespiratory Fitness, and Metabolic Risk. *American Journal of Preventive Medicine*, 42(6), pp.571-578.

[9] G. Jun and K. G. Smitha, "EEG based stress level identification," 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2016, pp. 003270-003274, doi: 10.1109/SMC.2016.7844738.

[10] Chen, W., Wang, Y., Cao, G., Chen, G. and Gu, Q., 2014. A random forest model based classification scheme for neonatal amplitude-integrated EEG. *BioMedical Engineering OnLine*, 13(Suppl 2), p.S4.

[11] O. Bazgir, Z. Mohammadi and S. A. H. Habibi, "Emotion Recognition with Machine Learning Using EEG Signals," 2018 25th National and 3rd International Iranian Conference on Biomedical Engineering (ICBME), 2018, pp. 1-5, doi: 10.1109/ICBME.2018.8703559.

[12] R. Panda, P. S. Khobragade, P. D. Jambhule, S. N. Jengthe, P. R. Pal and T. K. Gandhi, "Classification of EEG signal using wavelet transform and support vector machine for epileptic seizure diction," 2010 International Conference on Systems in Medicine and Biology, 2010, pp. 405-408, doi: 10.1109/ICSMB.2010.5735413.

[13] Polat, K. and Güneş, S., 2007. Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform. *Applied Mathematics and Computation*, 187(2), pp.1017-1026.

[14] Duan, L., Ge, H., Ma, W. and Miao, J., 2015. EEG feature selection method based on decision tree. *Bio-Medical Materials and Engineering*, 26(s1), pp.S1019-S1025.

[15] Ganglberger, W., Gritsch, G., Hartmann, M., Fürbass, F., Perko, H., Skupch, A. and Kluge, T., 2017. A Comparison of Rule-Based and Machine Learning Methods for Classification of Spikes in EEG. *Journal of Communications*, pp.589-595.

[16] Kamruzzaman, S. and Sarkar, A., 2011. A New Data Mining Scheme Using Artificial Neural Networks. *Sensors*, 11(5), pp.4622-4647.

[17] Guo, L., Rivero, D., Dorado, J., Rabuñal, J. and Pazos, A., 2010. Automatic epileptic seizure detection in EEGs based on line length feature and artificial neural networks. *Journal of Neuroscience Methods*, 191(1), pp.101-109.

[18] PUTHANKATTIL, S. and JOSEPH, P., 2012. CLASSIFICATION OF EEG SIGNALS IN NORMAL AND DEPRESSION CONDITIONS BY ANN USING RWE AND SIGNAL ENTROPY. *Journal of Mechanics in Medicine and Biology*, 12(04), p.1240019.

[19] Rahman, Tatiur & Ghosh, Apu & Shuvo, Md. Maruf Hossain. 2015. Mental Stress Recognition using K-Nearest Neighbor (KNN) Classifier on EEG Signals.

[20] B. Hu, X. Li, S. Sun and M. Ratcliffe, 2018. "Attention Recognition in EEG-Based Affective Learning Research Using CFS+KNN Algorithm," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 15, no. 1, pp. 38-45, 1 Jan.-Feb. doi: 10.1109/TCBB.2016.2616395.

[21] Merz, E.; Malcarne, V.; Roesch, S.; Ko, C.; Emerson, M.; Roma, V.; Sadler, G. 2013. Psychometric properties of Positive and Negative Affect Schedule (PANAS) original and short forms in an African American community sample. J. Affect. Disord. 151, 942–949

[22] International Journal of Engineering and Advanced Technology, 2019. Research of Machine Learning algorithms using K-fold cross validation. 8(6S), pp.215-218.

[23] Sharif, M. S., Abbod, M., Al-Bayatti, A., Amira, A., Alfakeeh, A. S. and Sanghera B. 2020. "An Accurate Ensemble Classifier for Medical Volume Analysis: Phantom and Clinical PET Study," in IEEE Access, vol. 8, pp. 37482-37494, doi: 10.1109/ACCESS.2020.2975135.

[24] Hong, C. and Choi, S., 2020. ROC curve generalization and AUC. *Journal of the Korean Data And Information Science Society*, 31(4), pp.477-488.

[25] Bradley, A., 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7), pp.1145-1159.

[26] Hong, C., 2021. Confusion plot for the confusion matrix. *Journal of the Korean Data And Information Science Society*, 32(2), pp.427-437.