EEG-Based Driver Drowsiness Detection: A Data-Driven Approach for Enhancing Road Safety

1st Preeti Paschapuri
Department of Electronics and Communication Engineering
KLE Technological University, Dr. M. S. Sheshgiri Campus
Belagavi,India
02fe22bec053@kletech.ac.in

2nd Diksha Nalawade
Department of Electronics and Communication
KLE Technological University, Dr. M. S. Sheshgiri Campus
Belagavi, India

3rd Rahul Kundargi
Department of Electronics and Communication
KLE Technological University, Dr. M. S. Sheshgiri Campus
Belagavi, India

4th Basavaraj Mujagoni
Department of Electronics and Communication
KLE Technological University, Dr. M. S. Sheshgiri Campus
Belagavi, India

5th Dattaprasad A. Torse
Department of Electronics and Communication
KLE Technological University, Dr. M. S. Sheshgiri Campus
Belagavi, India

***Abstract*—Early detection of drowsiness condition is critical in alerting the driver before mishap. In this paper, our aim is to develop a drowsiness detection model using analysis of EEG. The EEG signals from the F8 channel was selected for 60 sec duration as the data from this channel can effectively detect drowsiness in the driving scenarios. A dataset that was sampled at 128 Hz composed of 8,064 samples was preprocessed to eliminate muscular and ocular artifacts with root mean square error. The Power feature extraction of independent components of the signal was carried out using db4 Discrete Wavelet Transform. The trained SVM on power features was tested on the unclassified and clustered EEG data using Kmeans clustering, an efficiency of 97% was achieved resulting in reliable drowsiness detection.**

***Index Terms*—EEG, Drowsiness, Outlier Detection, Segmentation, Butterworth Filter, Independent Component Analysis, Discrete Wavelet Transform, Support Vector Machine, radial basis function kernel.**

I. Introduction

Road accidents are a worse case of human safety, where it directly harms humans in any of the scenarios considered. If we closely look at accidents, they are caused by some factors such as drunk driving, drowsiness, etc., among these factors, drowsiness is found to be one of the leading causes in serious road accidents [1, 2, 3]. Police reported that accidents in cars caused by drowsiness have contributed 3% in the total of road accidents[4].

Drowsiness is an intermediate state of sleepy and awake. It may be caused by many factors; for example, long driving, lack of sleep, consuming alcohol, taking more drugs,driving in the early morning and monotonous environments. According to 'National Road Travel Protection Management' (NHTPM) analysis around 1,00,000 crashes are direct [5] outcome of driver drowsiness each year, it has become a major concern to overcome drowsiness.

Many methods for detecting drowsiness in drivers are given in the Academic Literature, recorded by Stancin et al. [6] & Mohammedi et al. [7], these techniques are classified into three main categories, explained by Rayan [8]. First technique uses driver's physiological signals that are Electroencephalogram (EEG), Electrocardiogram (ECG) and Electrooculogram (EOG) activities, [9]. Second one is based on measurements of vehicle parameters, where drowsiness detection is made by measuring the movements of the steering wheel, pressure on acceleration paddle and others [10]. Third and the last category is based on the facial movements of the driver [11]. According to analysis in [12] [13] Physiological method best fits for detecting Drowsiness effectively. The method used in this study is using EEG signals, A study [14] showed an innovative of C3-O1 channel to analyze alpha and beta waves. The model was evaluated using different performance metrics. The model proposed in this paper contains Support Vector Machine (SVM), which was proved to be efficient [15].

II. Literature Survey

Arif et al.,[16] proposed a study that dealt with structured methods to identify drowsiness from a passive brain-computer interface system (pBCI). EEG signals of 12 participants recorded using an 16-channel EEG headset. Data was collected of the subjects drowsy under controlled conditions and that was sampled at a rate of 125 Hz. The raw EEG data were prepared for analysis after preprocessing to remove noise and artifacts. The signals were smoothed using Gaussian filters, notch filters used to eliminate the electrical interference at 50 Hz as well as 60 Hz. Further, band-reject filters were used to eliminate low-frequency artifacts $(< 0.4Hz)$ and high frequency artifacts $(> 40Hz)$. Finally, EEG signals were band-limited from 0.5 to 40 Hz to cover all frequencies, which is necessary for identifying the drowsy states: delta, theta, alpha, beta, and gamma. Finally, required spectral 3 features are extracted in the form of power spectral density (PSD) and band power ratios (BPRs). Feature selection techniques were applied to reduce the dimensionality using minimum redundancy maximum relevance (MRMR), Chi-square tests, and it dealt with structured methods to identify drowsiness from a pBCI system. These methods led to the identification of prominent features for drowsiness detection. For classification purposes, several machine learning algorithms such as decision trees, discriminant analysis, logistic regression, naïve Bayes, support vector machines (SVM), k-nearest neighbors (kNN), and ensemble models were tested. Finally, the ensemble classifier achieved maximum accuracy of 85.6%, where $F8$ was determined to be the best single-channel for drowsiness detection. S. Yu et al.,[17] studied using EEG recorded from the CAP Sleep Database, a publicly available signal. This database contained 108 EEG recordings from healthy subjects and those with sleep disorders also 16 recordings of healthy individuals were used. For the raw EEG data for analysis, FFT is used to calculate the Power Spectral Density of each epoch for feature extraction. Normalization techniques are used to address variability across subjects, ensuring consistency in data from all the recordings. SVM was used to make classifications. To reduce computational cost and enhance accuracy, feature extraction focused on a combination of traditional and sub-band EEG features. The two primary feature sets analyzed were: Traditional Broadband Features (standard EEG bands delta, theta, alpha, beta) and Sub-Band Features (Features split into 1 Hz sub-bands based on neuro-scientific findings). SVM classifiers were trained and tested on these features, using the RBF kernel for handling non-linear separations in data. Through feature selection experiments optimal classification accuracy was achieved using reduced subsets of only 9 sub-bands. The proposed SVM model showed high precision and recall. Accuracies of over 97% with the sub-band feature set. Li et al.,[18] built a system that used a wearable dry-electrode EEG headband integrated with a smartwatch with Bluetooth to collect live data. The headband made use of two signal electrodes positioned at the occipital regions (O1 and O2) and a ground. The acquired EEG signals, capturing relevant frequency bands $: 4\text{–}7Hz, : 8\text{–}12Hz, and : 13\text{–}30Hz)$, were amplified, filtered, and wireless transmitted to the smartwatch using Bluetooth Low Energy (BLE). The collected EEG was preprocessed on the smartwatch through a digital band-pass filter $(4\text{–}30Hz)$ to reject noise in the power line interference. Feature extraction utilized Fast Fourier Transform (FFT) analysis on 2-second Hamming windows for computing relative power ratios in the , , and bands. The classification task was handled by a Support Vector Machine-based Posterior Probabilistic Model that gave a continuous drowsiness probability between 0 and 1. Thus fine detection resolution can be gained. The model was trained by leave-one-subject-out cross-validation wherein the best achieved accuracy was using the radial basis function (RBF) kernel. Threshold probabilities were established for early and full drowsiness warnings. In the model a high accuracy was achieved: alert states with 91.25% and full drowsy warnings with 91.92%. Islam A. Fouad et. Al.,[19] used a dataset provided by Jianliang Min et al. was processed in MATLAB, the data was obtained from 12 men aging between 19-24 driving in a simulated environment, a cap with 32 electrodes placed according to the 10-20 system was sampled at 1000 Hz after driving for some time the signals captured at the last 5 min were considered under drowsy state, finally the data was divided into training and testing in 40% to 60% ratio. The data was filtered by a band pass filter between 0.15 and 45 Hz sampled at 1000 Hz, the EEG data of each subject had 1000 points X 30 channels, and was again filtered using a band pass filter with 0.15-45 Hz of range, it consisted of 600 segments. The data was further Normalized using "Z-Score Normalization". Classification was done using different algorithms like Diagonal discriminant analysis (Naive Bayes), DiagLDA, Support vector machine (SVM), K-Nearest Neighbors (KNN), Random Forest classifier with 10 trees was implemented. When compared, the accuracies of different classifiers SVM with linear kernel function performed better than SVM with RBF kernel function. Additionally, the most effective electrode positioning for each subject was located. KNN and SVM resulted in a maximum accuracy of 97.11% and 47.63% on a particular electrode respectively. It was inspected that the electrodes F4 (84.53%), CPZ (83.18%) and O2 (81.32%) were most performing and lead to highest average accuracies among all the classifiers. These electrodes were considered for further studies, to compute accuracy, sensitivity, specificity, and precision. SVM (RBF), KNN and Random Forest resulted in high rates in the confusion matrix and the AUC was more than 0.99. Mohammedi et al.,[14] proposed a system that uses Relative Power Ratio and the analyzes the alpha and beta activities to detect the state of the driver. The raw EEG of 11 adult males was recorded during driving using an EEG headset; the collected data was then sampled at 128 Hz and was filtered using Butterworth filter and band pass filter with cutoff frequency of 0.5 Hz and 30 Hz. They made use of ICA to cancel any Artifacts and Discrete wavelet transformation (DWT) from MATLAB a powerful way to analyze non-stationary signals used to separate each epoch band and the power of each band was determined for calculating the relative power of each frequency band for every single epoch, Daubechies4 (db4) mother wavelet was used. This method provided both time and frequency representation of the signal. Furthermore, binary SVM from the LibSvm toolkit is used for classification.The previously used data for classification was again used for training the model. The system can even adapt to new users if a user has previously used the headset since it can use the general drowsiness detection algorithm apart from the adapted ones for reliability. SVM Classifiers with -1 (power of alpha is dominant over the power of beta) and 1 (power of beta is dominant over the power of alpha) classes were classified. If the classification resulted in value more than or equal to 0 then the sample was classified to be drowsy and the onboard alarm was used to warn the driver, also the other drivers using V2R communication using VANETs, so that other vehicles could communicate between different RSU to share drowsiness data and act accordingly. This system's performance measure was F-score 99.88%, Accuracy 99.87%, Recall 99.77%, Precision 100%, False Negative Rate 0.22%, False Positive Rate 0.00%, Specificity 100%. Pasaribu et al,.[19,20] after Data collection, Preprocessing and feature extraction again using Wavelet transform, SVM for classification which used 5-fold cross validation and Quadratic kernel with a highest accuracy of 84.5%. Analyzation of different kernel functions like Linear 71.8%, Quadratic 84.5%, Cubic 52.3%, Fine Gaussian 72.2%, Medium Gaussian 70.2%, Coarse Gaussian 59.6% was also carried out. Thilagaraj et al., [5] whose system filtered the collected raw EEG using Butterworth Filter, The Statistical Feature Parameters extracted are Mean, Standard Deviation, Entropy, Skewness, Kurtosis using gaussian radial foundation characteristic and the polynomial feature.

A combination of 2 such algorithms is proposed in this study.

III. Methodology

This project focuses mainly on detecting drowsiness using recorded EEG data. The following block diagram shows the methodology that is employed in the project.

The Block diagram of the proposed system is shown in Fig. 1.

image.png

Fig. 1. Block Diagram of the Proposed System

A. Data Collection

The EEG data used in this study is publically available on Kaggle [https://www.kaggle.com/datasets/samnikolas/eeg-dataset]. From the previous studies F8 channel was best known to capture drowsy data, so only the F8 column in the dataset was extracted for this study.

B. Preprocessing and Segmentation

Preprocessing was done to minimize the noise and to reduce misleading data. Removal of outliers is important because they can affect the model's performance and lead to misclassifications. The outliers can also be the result of external noise or Muscular, Ocular artifacts. To carry out this Interquartile Range (IQR) method is used because the extreme values contributed by the artifacts lie outside the calculated range of IQR, which are later removed and the middle range lying data is preserved for further analysis. The Single channel data were divided into smaller fixed-size segments or epochs, this resulted in 29 such epochs.

C. Butterworth Filtering and Independent Component Analysis (ICA) Application

Along with the Butterworth filtering, band pass filtering was also done to get rid of frequency components that lie outside the 0.5Hz to 50 Hz range when sampled at a rate of 256Hz, since they do not have any significance in the detection of drowsiness. Additionally Butterworth Filtering to smoothen and preserve information in the epochs was performed. FastICA is used to clean all the filtered epochs by decomposing them into independent components by doing which noise can be identified and separated after which the signal is reconstructed using inverse transform.

D. Obtaining Wavelet features and power

The study uses Discrete Wavelet Transform with Daubechies-4 (db4) and 4 levels for power and feature extraction of epochs. It computes wavelet coefficients by analyzing multiple frequencies.

E. K-Means Clustering for Labeling

The dataset used in this study lacks labels, which could be manually added after examining by a medical professional, but the system does this using K-means clustering, the model is fed with the wavelet powers, then it groups them into two i.e. drowsy or alert clusters. These newly labeled epochs were then used to train the SVM model. The training set consists of 60% of the labeled epochs and the remaining 40% is used for testing purposes.

F. Support Vector Machine

The SVM model with RBF kernel was trained. Additionally to address the class imbalance, SMOTE (Synthetic Minority Oversampling Technique) was used to generate synthetic samples for minority class.

IV. Results and Discussion

RawEEGSignal.png

Fig. 2. Raw EEG Data

In Fig. 2, the raw EEG data of the F8 region used for the study which is shown in the time domain.

The narrow box in Fig. 3. indicates a very small IQR (a measurement of the spread of data in a box plot) and varies less in the central portion of the data. There are numerous outliers above and below the whiskers (the two lines that extend from the ends of a box), suggesting that the data contains extreme values away from the center. The median is centered in the box, showing symmetry in the central part of the data.

BoxplotforOutliers.png

Fig. 3. Boxplot for Outliers

AfterRemovingOutliers.png

Fig. 4. Signal after Outliers Removal

The Fig. 4. shows the EEG signal after the removal of outliers in it. It can be seen that before the removal of outliers the maximum amplitude was somewhere around 60uV but after the removal of artifacts the range of the amplitude is observed to be around 20uV, which indicates the removal of unwanted and misleading data points.

In Fig. 5., the signal is segmented for clear analysis. Segmentation is done considering a window size and step size of 1. This results in 29 different segments which are used to train the SVM model and these capture local data patterns that have meaningful information about the dependencies.

EEGSignalSegmentation.png

Fig. 5. EEG Signal Segmentation

After Independent Component Analysis on the number of epochs obtained after segmentation, Band pass Butterworth Filtering was performed on them. The epochs then were free of unwanted frequency ranges and noise. The first 5 of those epochs are shown in Fig. 6.

Fig. 7. Shows the first epoch after cleaning with Independent

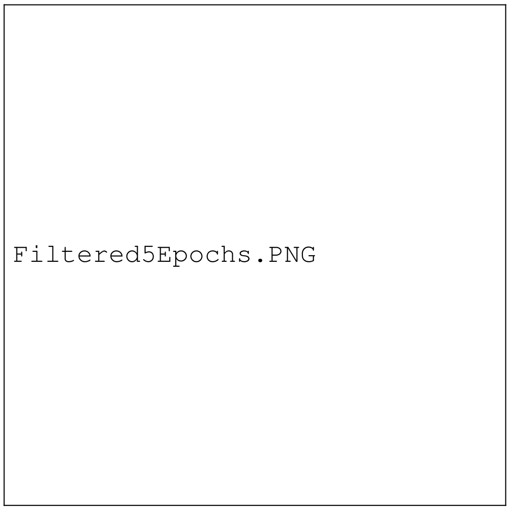

Fig. 6. Filtered Signal Epochs

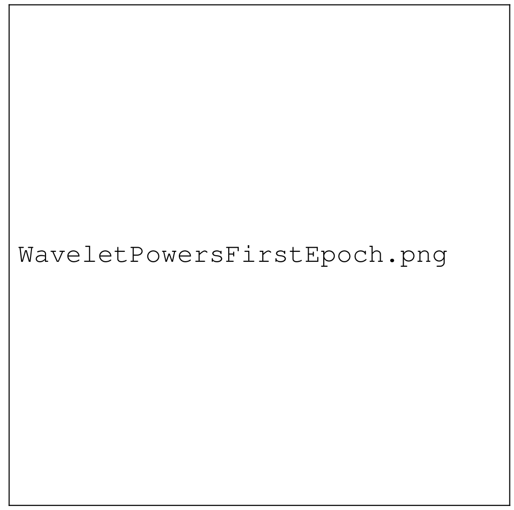

Fig. 8. Wavelet Powers First Epoch

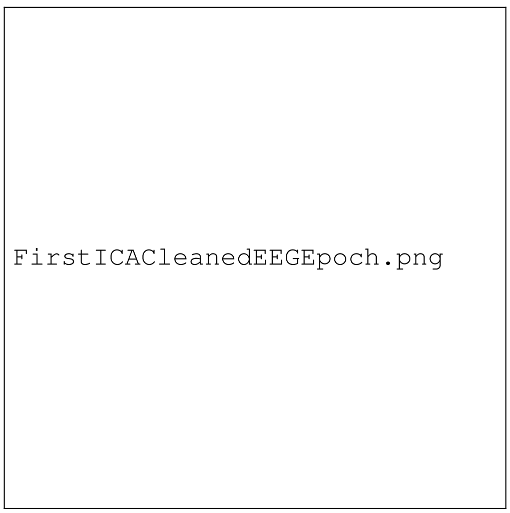

Fig. 7. First ICA Filtered Epoch

Fig. 9. K-Means Cluster Centers

Component Analysis which was done for preserving underlying brain signals. After cleaning, the signal appeared smooth without any abrupt spikes that indicated the absence of noise and artifacts. No data were lost during this process.

Fig. 8. represents the wavelet powers for the first epoch, on the X-axis Wavelet Decomposition Levels whereas on the Y-axis the power levels are shown. Level 2 exhibits the maximum power which indicates the signal is significant in that frequency band. After Level 3 the power drops indicating lower frequency components.

Fig. 9. represents K-means clustering centroids performed on the Wavelet Power features. The algorithm identified 2 clusters, cluster 1 (blue) and cluster 2 (orange). X-axis corresponds to Wavelet Decomposition Levels whereas, Y-axis corresponds to average power. Cluster 1 has lower power as compared to Cluster 2. This may possibly mean different states of the brain. The SVM model was able to achieve an accuracy of 97%. Precision for the alert and drowsy classes was 0.50 and 1.00 respectively. Recall for the alert and drowsy classes was 1.00 and 0.00 respectively. F1-score for the alert and drowsy classes was 0.67 and 0.00 respectively.

V. Conclusion

In conclusion, this study effectively utilized discrete wavelet decomposition and clustering techniques to analyze EEG signals, demonstrating a reliable method for differentiating between drowsy and alert states. Although an impressive accuracy 97% was achieved with the SVM model, there remains scope for improvement to address the needs of real-world applications. In scenarios such as driver monitoring, occupational safety, and healhealth care, ahigher precision of drowsiness detection systems is essential. Future research work is proposed on the integration of advanced machine learning algorithms, hybrid models, and larger, more diverse datasets to overcome current limitations. Furthermore, using

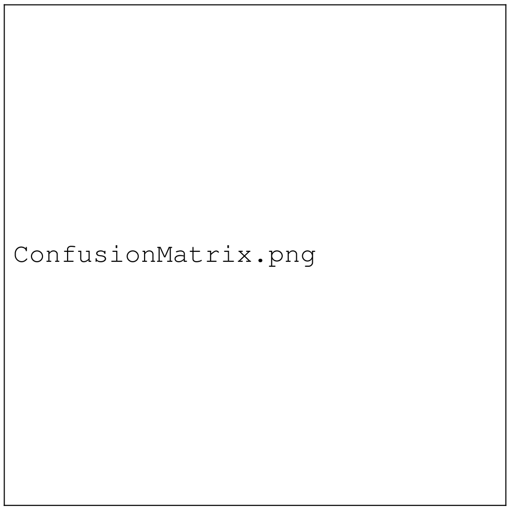

Fig. 10. Confusion Matrix

real-time signal processing and multimodal approaches, such as combining EEG with other physiological signals, could further improve detection accuracy.

REFERENCES

[1] Phan, Anh-Cang & Trieu, Thanh-Ngoan & Phan, Thuong-Cang. (2023). Driver drowsiness detection and smart alerting using deep learning and IoT. Internet of Things. 22. 100705. 10.1016/j.iot.2023.100705

[2] S. Kaplan, M. A. Guvensan, A. G. Yavuz and Y. Karalurt, "Driver Behavior Analysis for Safe Driving: A Survey," in IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 6, pp. 3017-3032, Dec. 2015, doi: 10.1109/TITS.2015.2462084

[3] Forsman PM, Vila BJ, Short RA, Mott CG, Van Dongen HP. Efficient driver drowsiness detection at moderate levels of drowsiness. Accid Anal Prev. 2013 Jan;50:341-50. doi: 10.1016/j.aap.2012.05.005. Epub 2012 May 28. PMID: 22647383.

[4] Lyznicki JM, Doege TC, Davis RM, Williams MA. Sleepiness, driving, and motor vehicle crashes. Council on Scientific Affairs, American Medical Association. JAMA. 1998 Jun 17;279(23):1908-13. doi: 10.1001/jama.279.23.1908. PMID: 9634264.

[5] Thilagaraj, M., Pallikonda Rajasekaran, M., & Ramani, U. (2021). Identification of drivers' drowsiness based on features extracted from EEG signal using SVM classifier. *3C Tecnología. Glosas de Innovación Aplicadas a la Pyme*, 579–595. https://doi.org/10.17993/3ctecno.2021.specialissue8.579-595

[6] Stancin I, Cifrek M, Jovic A. A Review of EEG Signal Features and their Application in Driver Drowsiness Detection Systems. Sensors (Basel). 2021 May 30;21(11):3786. doi: 10.3390/s21113786. PMID: 34070732; PMCID: PMC8198610.

[7] Mohammedi, Mohamed & Omar, Mawloud & Bouabdallah, Abdelmadjid. (2018). Automatic removal of ocular artifacts in EEG signals for driver's drowsiness detection: A survey. 10.1109/SaCoNeT.2018.8585680.

[8] Rajan SP (2015) A survey on detecting methods of driver drowsiness. Int J Innov Eng Manag 4(1):2319–3344

[9] Khushaba RN, Kodagoda S, Lal S, Dissanayake G. Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm. IEEE Trans Biomed Eng. 2011 Jan;58(1):121-31. doi: 10.1109/TBME.2010.2077291. Epub 2010 Sep 20. PMID: 20858575.

[10] Awais, M., Badruddin, N., & Drieberg, M. (2014). A simulator based study to evaluate driver drowsiness using electroencephalogram. In 2014 5th International Conference on Intelligent and Advanced Systems (ICIAS) The Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/icias.2014.6869485

[11] Warwick, Brandy & Symons, Nicholas & Chen, Xiao & Xiong, Kaiqi. (2015). Detecting Driver Drowsiness Using Wireless Wearables. 585-588. 10.1109/MASS.2015.22.

[12] De Gennaro L, Ferrara M, Bertini M. The boundary between wakefulness and sleep: quantitative electroencephalographic changes during the sleep onset period. Neuroscience. 2001;107(1):1-11. doi: 10.1016/s0306-4522(01)00309-8. PMID: 11744241.

[13] A. Čolić, O. Marques, and B. Furht, "Driver Drowsiness Detection," Springer International Publishing, 2014, doi: 10.1007/978-3-319-11535-1.

[14] Mohammedi, Mohamed & Mokrani, Juba & Mouhoubi, Abdenour. (2024). An automated and highly efficient driver drowsiness detection and alert system using electroencephalography signals for safe driving. Multimedia Tools and Applications. 83. 87299-87322. 10.1007/s11042-024-19797-2.

[15] Islam A. Fouad(2023)A robust and efficient EEG-based drowsiness detection system using different machine learning algorithms;https://doi.org/10.1016/j.asej.2022.101895

[16] Arif S, Munawar S and Ali H (2023), Driving drowsiness detection using spectral signatures of EEGbased neurophysiology. Front. Physiol. 14:1153268. doi: 10.3389/fphys.2023.1153268

[17] S. Yu et al., "Support Vector Stancin I, Cifrek M, Jovic A. A Review of EEG Signal Features and their Application in Driver Drowsiness Detection Systems. Sensors (Basel). 2021 May 30;21(11):3786. doi: 10.3390/s21113786. PMID: 34070732; PMCID: PMC8198610.

[18] G. Li, B. -L. Lee and W. -Y. Chung, "Smartwatch-Based Wearable EEG System for Driver Drowsiness Detection," in IEEE Sensors Journal, vol. 15, no. 12, pp. 7169-7180, Dec. 2015, doi: 10.1109/JSEN.2015.2473679.

[19] Islam A. Fouad, A robust and efficient EEG-based drowsiness detection system using different machine learning algorithms, Ain Shams Engineering Journal, Volume 14, Issue 3, 2023, 101895, ISSN 2090-4479, https://doi.org/10.1016/j.asej.2022.101895.

[20] Pasaribu, Novie & Halim, Timotius & Ratnadewi, Ratnadewi & Prijono, Agus. (2021). EEG signal classification for drowsiness detection using wavelet transform and support vector machine. IAES International Journal of Artificial Intelligence (IJ-AI). 10. 501. 10.11591/ijai.v10.i2.pp501-509.