Analysis of EEG Signals Using Machine Learning Algorithms

2024 International Conference on Advances in Modern Age Technologies for Health and Engineering Science (AMATHE) | 979-8-3503-7156-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/AMATHE61652.2024.10582072

Garima Chandel
Department of Electronics and Communication Engineering
Chandigarh University
Mohali, India
chandelgarima5@gmail.com

Tarun Aggarwal
Department of Mechatronics Engineering
Chandigarh University
Punjab, India
tarunaggarwal0211@gmail.com

Tejinderpal Singh
Department of Mechatronics Engineering
Chandigarh University
Punjab, India
tejinderpal0520@gmail.com

Samarpreet Singh
Department of Mechatronics Engineering
Chandigarh University
Punjab, India
samar.cu23@gmail.com

Karanvir Singh
Department of Mechatronics Engineering
Chandigarh University
Punjab, India
abhisandhu1308@gmail.com

Harpreet Singh
Department of Mechatronics Engineering
Chandigarh University
Punjab, India
harpreetsingh29790@gmail.com

***Abstract*— Epilepsy is a threatening neurological disease which causes sudden bursts in the electrical activities of the brain known as epileptic seizures. Due to its less availability in terms of treatment, and the data is also very complex to be dealt with, it has become a matter of immense concern. Also, the rate of patients for a single doctor is very high in low- and middle-income countries. Epilepsy can be overcome if it is diagnosed at early stages. So, to achieve the same, an automated machine learning based model is proposed in this study which uses the EEG patterns received from the patient. The model is capable of telling whether the patient is epileptic or not by using the data from the electrical activities of the patient's brain. The model uses Supervised Machine learning and the random forest (RF) classifier to do the analysis of EEG signals for seizure detection. Python library pandas is used to deal with the data in the spreadsheet form and matplotlib is used for visualization purposes. The model has predicted epilepsy with a good accuracy of 99.96%, specificity of 99.2% and the sensitivity came out to be 100%.**

Keywords*— *Epilepsy, Epileptic Seizure, EEG Signals, Supervised Machine learning, Random Forest Classifier.

I. Introduction

Now-a-days, the study of neurotechnology has grown very much. It is due to the advancements in brain screening techniques which help in diagnosing neurological diseases. It encompasses the techniques which visualize the brain activities and its functions and helps in controlling and improving the brain's conditions. Therefore EEG (Electroencephalogram) has become extremely popular. EEG is a non-invasive process which extracts the brain information and hence helps in detecting epilepsy [1].

Epilepsy is a lethal neurological disease in which a patient encounters severe uncontrolled bursts in the brain activities, known as seizures. It happens due to changes in the electrical activities of the brain. Epileptic seizures occur among people of any age and background. However, the information like frequencies of these epileptic seizures differ among people [2]. Loss of consciousness, confused speech, weakness, anxiety, muscle rigidity, twitching, jerking movements of body parts etc. are the symptoms of epilepsy. Focal seizures affect only a specific part of the brain hence, they are also known as partial focal seizures. Focal Seizures are also of two types: simple partial and complex partial. In simple partial seizures, a person is unable to communicate but still is conscious. But in complex partial seizures, a person becomes unconscious and hence gets confused and starts behaving differently. Generalized seizures are more lethal than focal seizures as they affect the whole brain [3].

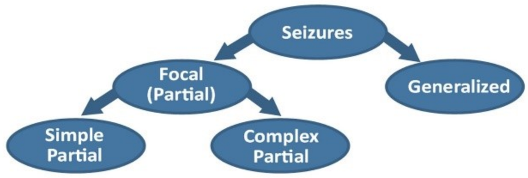

Fig. 1. Types of Epileptic Seizures and then further classification of Focal Seizures

Genetic factors, brain injuries, strokes, brain tumours, infections, down syndrome are the causes of epileptic seizures. Sometimes the cause of seizure is unknown, and in such cases, it is called idiopathic. The seizures can occur at any time and last from seconds to a few minutes [4].

There are four phases of seizures: preictal phase, ictal phase, interictal phase and postictal phase. Preictal phase is the phase of the sensations which a patient experiences before the seizure. It is also known as aura. It leads to various sensations and emotions which are a sign to show that the seizure is imminent. Ictal phase is the phase in which a patient encounters seizure and hence leads to alteration in the consciousness. It remains for seconds to a few minutes. Ictal phase is also known as Middle or seizure phase. Postictal phase is the recovery phase after the seizure. In this phase the patient may experience dizziness, fatigue, headache, and some other symptoms. Interictal phase is the phase in which a person having epilepsy does not experience any seizure. The patient seems to be normal in this phase. This phase lasts from days to a few months. Interictal phase is the most crucial phase because in this phase the patient's activities are monitored and examined by doctors and hence it is also known as subclinical phase [4].

According to the World Health Organization (WHO), there are 50 million people worldwide having epilepsy and hence again makes it a matter of concern and also making it one of the most common neurological diseases. About 70-80% people having epilepsy are living in low- and middle-income countries. WHO is also stating that about 5 million people are found with the symptoms of epileptic seizures. It is also found that the proportion of active epileptic patients with the normal people is between 4 and 10 per 1000 people. WHO also estimates that about 70% of the people who are suffering from epilepsy can live seizure free if they are properly examined and treated [3].

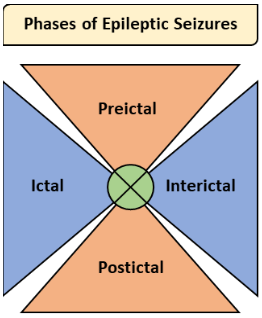

Fig. 2. Different types of Phases of Epileptic Seizures.

EEG is the readily available method used for detecting such seizures. In the EEG process, some electrodes are attached to the scalp area of the patient. The patient is asked to do some activities like deep breathing, walking, and some other functions to collect data from the various activities. The electrodes attached to the scalp collect the data and show the output patterns on the display screen which is then studied by the expert neurologists to conclude whether the patient is suffering from epilepsy or not. An automated system is made based on supervised machine learning which is based on a classifier to do the classification. This system extracts and selects the features upon which it is supposed to work and then applies the classifier and the algorithm and hence classification is done between the epileptic patients and the normal ones.

Many researches are made previously on this topic to classify EEG signals received from epileptic and healthy patients using the same data i.e., Bonn University dataset [2]. Below a quick review is given of some of the existing solutions for better understanding and visualization.

Many of the authors have used different-different classifiers- Logistic Model Tree (LMT), CNN, Linear Classifier Wavelet Transform, Deep Neural Networks (DNN), KNN and DWT, Time and Frequency Domain Method, Random Forest etc. Some authors have used Machine learning while some have used deep learning algorithms. But Supervised machine learning's analysis is easy and is more flexible whereas deep learning is much more complex and difficult to operate [5].

Tzimourta et. al. used the classifier random forest based on the discrete wavelet transform (DWT). In this technique the author has decomposed the 5 levels in EEG segments and hence has extracted 5 features in the form of wavelet coefficients and then the extracted features are used to train the random forest classifier [4].

Liu et. al. has presented an algorithm for seizure detection based on Deep Forest (DF) model and variational modal decomposition (VMD) [6]. This model is decomposed over the EEG readings and constructed over the time-frequency distributions of the EEG signals.

Kantipudi et al., has used Finite Linear Haar wavelet-based Filtering (FLHF) to filter the input signals and used Fractional Dimension (FD) Analysis to extract the relevant features. Then they have used Grasshopper Bio-Inspired Swarm Optimization (GBSO) and Temporal Activation Expansive Neural Network (TAENN) technique to do the classification between the normal and seizure affected [7].

Kunekar et al., has used both machine learning and deep learning algorithms as they help in fully automated feature extraction and classification. Their model have used Long Short Term Memory (LSTM) approach to find the best results for epileptic seizure detection [8].

Another technique used by Gini et al., has used an improved optimization algorithm based on random forest classification [9]. The author has extracted the spectral features from EEG samples from intrinsic mode functions. He then has done the feature selection and then application of the algorithm is done. He has implemented his work on the MATLAB platform. Kabir et al., has used Logistic Model Trees (LMT) to do the classification between healthy and epileptic patients [10]. Zhou et al., has used deep learning classifier i.e., Convolutional Neural Networks (CNN) [11], Ahammad et al., has used a Linear classifier based on Wavelet transform [12]. Akyol et al., has used Deep neural networks (DNN) [13]. Other Techniques and classifiers used are: Ibrahim et al., discrete wavelet transform (DWT) and K. neural networks (KNN) [14], Fasil et al., time domain method [15], Acharya et al., gaussian mixture model (GMM) [16], Duan et al., deep metric learning [17], Xiang et al., Fuzzy Entropy based on classification algorithm [18], Rao et al., random forest (RF) classifier [19], Sood et al., random forest (RF) classifier [20].

The organisation of paper is as follows: Section II is describing about dataset used in present study, followed by Methodology in Section III. Result and discussion was carrying out in Section IV. Finally work is concluded in Section V.

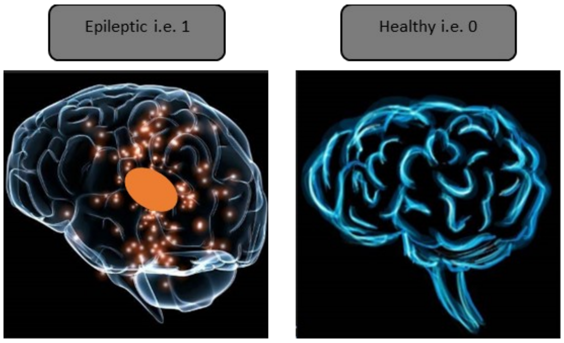

Fig. 3. Brain Image of Epileptic patient and healthy person.

II. DATASET DESCRIPTION

Bonn University has provided the dataset for classification of EEG signals received from patient's activities [2]. It has 178 attributes where X1, X2, X3, -----, X178 are referred to as variables which record the data received from patient's activities at 11500 instances. 'y' is the target variable which gives the output as either 1 or 0 and according to which the patients are classified into epileptic or healthy (non-epileptic).

The patients corresponding to output 1 are considered epileptic and the patients categorized under 0 are healthy.

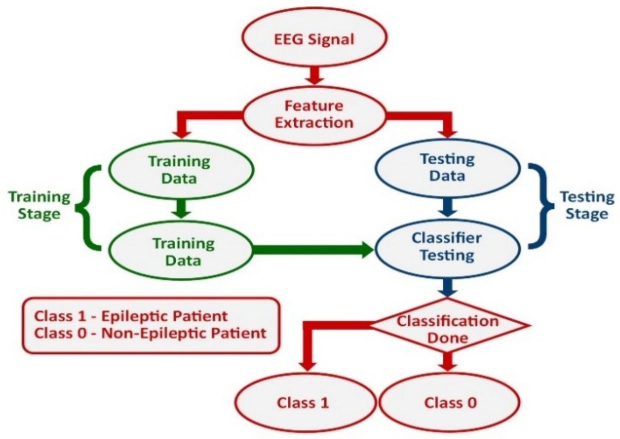

Fig. 4. Flowchart depicting the methodology process

III. Methodology

This work have used machine learning algorithms to do the classification between the healthy and epileptic patients by the help of Bonn University dataset. The dataset provided by Bonn University is raw, so firstly converted it to an understandable form and then applied the classifier to do the classification. It is important to convert the data because Machine Learning algorithms can be easily applied and the outcomes will be accurate and precise also. Also, the algorithms can be applied easily. The prediction regarding the epilepsy among the patients is made by using supervised machine learning algorithms. In Supervised machine learning algorithms, labelled data for training of the machine has been used. As here predefined inputs have been used, there are high chances of getting accurate predictions. Hence the efficiency in the real time scenarios is also very high.

As shown in the Fig. 4, firstly the extraction of the data has been done from the dataset. Data extracted from the different patients according to the instances and the attributes. Then the pre-processing of data has been done. Pre-processing involves filtration of received data and standardizing it. Then the data has been divided into two sections: training data and testing data. 70-80% of the extracted data is used for the training purpose while the remaining 20-30% of the extracted data is used for the testing. The testing set depicts the performance of the model while the training set is used for the training purpose.

Then to the extracted features, the classifier is applied. These features are very crucial for the application of classifiers. Random Forest classifier has been used. Random Forest classifier consists of decision trees. These trees are fetched with the several data and then mean is calculated to find the prediction. The result is then classified into two categories: 1 and 0. The category '1' depicts the patients who are suffering from epilepsy or having the symptoms of epileptic seizures while category '0' represents healthy or non-epileptic persons.

Then the accuracy, specificity and sensitivity of the trained model is calculated [21]. The accuracy of the model depicts how accurately or precisely the trained model predicts or classifies between epileptic and non-epileptic (healthy) patients. Hence accuracy is a very important aspect.

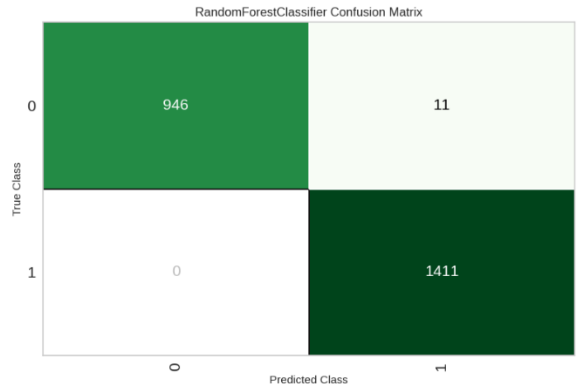

Fig. 5 Results for Healthy and epileptic EEG classes

Sensitivity is the term in which the model correctly predicts the true positive (TP) instances. True Positive (TP) cases are those in which the model correctly predicts the seizures. Specificity refers to the term in which the model correctly predicts the True Negative (TN) instances. True Negative (TN) cases are those in which the system correctly determines the non-seizures. A confusion matrix is plotted between the true and the detected labels. Confusion matrix shows performance of proposed algorithm.

TABLE I. Comparison Table Showing Results Comparison With State of The Art Methods.

Author	**Year**	**Methodology Used**	**Acc. (%)**	**Spec. (%)**	**Sen. (%)**
Acharya et al., [16]	2012	Gaussian Mixture Model	99%	—	—
Ahammad et al., [12]	2014	Linear Classifier & Wavelet Transform	84.2%	—	98.5%
Xiang et al., [18]	2015	Fuzzy Entropy & rf model	100%	100%	100%
Kabir et al., [10]	2016	Logistic Model Trees	95.33%	92%	99%
Zhou et al., [11]	2018	CNN	96.7%	96.8%	96.7%
Ibrahim et al., [14]	2018	Discrete Wavelet Transform and KNN	94.6%	—	—
Akyol et al., [13]	2020	Deep Neural Networks	97.17%	—	93.11%
Duan et al., [17]	2021	Deep Metric Learning	98.6%	100%	—
Fasil et al., [15]	2023	Time Domain Method	99%	—	—
This Work	—	**Random Forest (rf)**	**99.96%**	**99.2%**	**100%**

IV. RESULT AND DISCUSSION

With the use of the above methodology, three parameters were calculated based on presented classifier i.e. RF classifier. The three parameters are accuracy, sensitivity and specificity. The accuracy, sensitivity and specificity is calculated with the help of confusion matrix. There are two classes in the confusion matrix which comprises further total 4 subsets which are True Negative, True Positive, False Negative and False Positive. The confusion matrix of proposed model is shown in Fig. 5..

a) Accuracy: Accuracy describes whether the true or actual values match the predicted class or not. It is calculated by taking the difference between the true (actual) class and the predicted class. Accuracy is the ratio of truly predicted class to the total number of instances as described in (1).

$$\text{i.e., Accuracy}= \frac{(T.P.+ T.N.)}{(T.P.+ T.N.+ F.P.+ F.N.)} \tag{1}$$

The accuracy of trained model using supervised machine learning and random forest classifier is 99.96%.

b) Specificity: Specificity is calculated by taking the ratio of true negative instances to the sum of true negative and false positive instances as given in (2).

$$\text{i.e., Specificity} = \frac{(T.N.)}{T.N.+ F.P.} \tag{2}$$

Supervised machine learning model has achieved a specificity of 99.2%.

c) Sensitivity: Sensitivity is calculated by taking the ratio of true positive instances to the sum of true positive and false negative instances as shown in (3).

$$\text{i.e., Sensitivity} = \frac{(T.P.)}{(T.P.+ F.N.)} \tag{3}$$

The model has achieved a sensitivity of 100% using RF classifier. Hence, The detection algorithm used random forest classifier, the model got an accuracy, specificity and sensitivity of 99.96 %, 99.2 % and 100 % respectively.

Proposed Model's calculated accuracy is better than many previously researches based on random forest classification as shown in Table I. Sood et al., proposed the random forest based model on the same Bonn University dataset. That roposed model has achieved an accuracy of 91.4% and a sensitivity of 90.25% [20]. Rao et al., also proposed his model and achieved an accuracy of 94.1% [19]. Tzimourta et al., also proposed the model using classifier random forest based on Discrete wavelet transform and achieved an accuracy of 95% [4]. Gini et al., used an improved optimization algorithm based on RF classification and got an accuracy of 97.76% [9].

Random Forest is an emerging algorithm based on machine learning. Despite the fact that it is slow, still it has a number of wide applications. It can handle large datasets. Means that It can work when there are thousands of input variables and even when some of the data is missing. Despite having all these things, it used to maintain the accuracy [22].

V. CONCLUSION

Epilepsy is a very dangerous neurological disorder and around 50 million people are suffering from this lethal disease. Also, around 70-80% of the patients suffering with epilepsy live in low- and middle-income countries, and also the doctor availability for the patients is less in those countries, so it becomes difficult for their diagnosis and treatment. WHO also states that 70% of the patients will become normal if they get treatment at early stages of this disorder. As the availability of doctors is less, so it becomes necessary to have an automated disease detection model which will detect the disease at early stages by providing the dataset to the model and the model will tell whether the person is suffering from epilepsy or not by using supervised machine learning algorithm and random forest classifier. The limitation of this model is that it is capable of doing classification of short duration EEG signal-based data and also the artifacts like muscle movement, environmental conditions and other factors have bad impact on the results. But in future it will be doing classification on both short and long duration data and will be more flexible and hence can-do real-time EEG signal analysis. The model will be dealing with more complex data. This proposed model will lead to advancements in EEG technology.

REFERENCES

[1] H Anwar, QU Khan, N Nadeem, I Pervaiz, M Ali, FF CheemaDiscoveries, "Epileptic seizures," *2020•ncbi.nlm.nih.gov*, Accessed: Apr. 20, 2024. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7305811/

[2] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Physical Review E, 2001•APS*, vol. 64, no. 6, p. 8, 2001, doi: 10.1103/PhysRevE.64.061907.

[3] G. Chandel, A. Kaur, S. Grover, S. Garg, and G. Singh, "Machine Learning Based Model For Seizure Detection Using Eeg Signals," *sifisheriessciences. Journal of Survey in Fisheries Sciences, 2023•sifisheriessciences.com*, vol. 10, no. 1, pp. 2279–2289, 2023, Accessed: Apr. 20, 2024. [Online]. Available: http://www.sifisheriessciences.com/index.php/journal/article/download/1171/563

[4] K. D. Tzimourta *et al.*, "A robust methodology for classification of epileptic seizures in EEG signals," *Health Technol (Berl)*, vol. 9, no. 2, pp. 135–142, Mar. 2019, doi: 10.1007/S12553-018-0265-Z.

[5] X. Liu *et al.*, "Epileptic seizure detection based on variational mode decomposition and deep forest using EEG signals," *mdpi.com Brain Sciences, 2022•mdpi.com*, Accessed: Apr. 20, 2024. [Online]. Available: https://www.mdpi.com/2076-3425/12/10/1275

[6] X. Liu, J. Wang, J. Shang, J. Liu, L. Dai, and S. Yuan, "Epileptic Seizure Detection Based on Variational Mode Decomposition and Deep Forest Using EEG Signals," *Brain Sciences 2022, Vol. 12, Page 1275*, vol. 12, no. 10, p. 1275, Sep. 2022, doi: 10.3390/BRAINSCI12101275.

[7] M. Kantipudi, N. Kumar, R., "An improved GBSO-TAENN-based EEG signal classification model for epileptic seizure detection," *nature.com Scientific Reports, 2024•nature.com*, Accessed: Apr. 20, 2024. [Online]. Available: https://www.nature.com/articles/s41598-024-51337-8

[8] P. Kunekar, M. K. Gupta, and P. Gaur, "Detection of epileptic seizure in EEG signals using machine learning and deep learning techniques," *Journal of Engineering and Applied Science*, vol. 71, no. 1, Dec. 2024, doi: 10.1186/S44147-023-00353-Y.

[9] A. Gini, M. Q.- Management, "An improved optimization algorithm for epileptic seizure detection in EEG signals using random forest classifier," *webology.org Management, 2021•webology.org*, vol. 18, 2021, doi: 10.14704/WEB/V18SI04/WEB18132.

[10] E. Kabir Siuly Yanchun Zhang, "Epileptic seizure detection from EEG signals using logistic model trees," *braininformatics.springeropen.com*, vol. 3, no. 2, pp. 93–100, Jun. 2016, doi: 10.1007/s40708-015-0030-2.

[11] S. Geng and W. Zhou, "Nonlinear feature comparision of EEG using correlation dimension and approximate entropy," *Proceedings - 2010 3rd International Conference on Biomedical Engineering and Informatics, BMEI 2010*, vol. 3, pp. 978–981, 2010, doi: 10.1109/BMEI.2010.5639306.

[12] N. Ahammad, T. Fathima, P. J.-B. research international, and undefined 2014, "Detection of epileptic seizure event and onset using EEG," *hindawi.com BioMed research international, 2014•hindawi.com*, Accessed: Apr. 20, 2024. [Online]. Available: https://www.hindawi.com/journals/bmri/2014/450573/abs/

[13] K Akyol. "Stacking ensemble based deep neural networks modeling for effective epileptic seizure detection," *Elsevier Expert Systems with Applications, 2020•Elsevier*, Accessed: Apr. 20, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417420300658

[14] S. Ibrahim, R. Djemal, "Electroencephalography (EEG) signal processing for epilepsy and autism spectrum disorder diagnosis," *Elsevier*, Accessed: Apr. 20, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0208521617301973

[15] O. K. Fasil and R. Rajesh, "Epileptic seizure classification using shifting sample difference of EEG signals," *J Ambient Intell Humaniz Comput*, vol. 14, no. 9, pp. 11809–11822, Sep. 2023, doi: 10.1007/S12652-022-03737-9.

[16] U. Acharya, S. Sree, A. Alvin, J. S.-E. S. with, and undefined 2012, "Use of principal component analysis for automatic classification of epileptic EEG activities in wavelet framework," *Expert Systems with Applications, 2012•Elsevier*, Accessed: Apr. 20, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417412002849

[17] L. Duan, Z. Wang, Y. Qiao, Y Wang, Z Huang, B Zhang, "An automatic method for epileptic seizure detection based on deep metric learning," *ieeexplore.ieee.orgL IEEE Journal of Biomedical and Health Informatics, 2021•ieeexplore.ieee.org*, Accessed: Apr. 20, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9664005/?casa_token=WwAFT51uPzcAAAAA:u-eL-XrIpsbUEV4ruHSQbK9kjLpwSsjRRmakt_fPGWlWcfjRz9kMQVPv7TMu4i48RlVCDT1JIeXVxLU

[18] J Xiang, C Li, H Li, R Cao, B Wang, X Han, J Chen, "The detection of epileptic seizure signals based on fuzzy entropy," *Elsevier Journal of neuroscience methods, 2015•Elsevier*, vol. 243, pp. 18–25, 2015, doi: 10.1016/j.jneumeth.2015.01.015.

[19] M Rao, "Decoding imagined speech using wearable EEG headset for a single subject," *ieeexplore.ieee.org2021 IEEE International Conference on Bioinformatics and, 2021•ieeexplore.ieee.org*, Accessed: Apr. 20, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9669361/

[20] V. Sood, S LalIlkogretim, "A REVIEW ON EPILEPSY SEIZURE DETECTION TECHNIQUES.," *researchgate.net Online, 2021•researchgate.net*, doi: 10.17051/ilkonline.2021.02.60.

[21] G. Chandel, P. P. M. Shanir, O. Farooq, and Y. U. Khan, "A simplified method for classification of epileptic EEG signals," *Int J Biomed Eng Technol*, vol. 25, no. 1, pp. 60–76, 2017, doi: 10.1504/IJBET.2017.086552.

[22] Y. Liu, H. Liu, B. Yang, "Automatic sleep arousals detection from polysomnography using multi-convolution neural network and random forest," *ieeexplore.ieee. IEEE Access, 2020•ieeexplore.ieee.org*, Accessed: Apr. 20, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9205893/