Proceedings of the International Conference on Automation, Computing and Renewable Systems (ICACRS 2022)
IEEE Xplore Part Number: CFP22CB5-ART: ISBN: 978-1-6654-6084-2

2022 International Conference on Automation, Computing and Renewable Systems (ICACRS) | 978-1-6654-6084-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICACRS55517.2022.10029003

An Extensive Review of Machine Learning Techniques for EEG Signal Processing

Anita.M[1,]
Department of Computer Science and Engineering,
S.A. Engineering College,
Chennai, India.
anitam.engg@gmail.com

A.Meena Kowshalya[2,]
Department of Computer Science and Engineering,
Government College of Technology,
Coimbatore, India.
meenakowsalya@gct.ac.in

B.Maheswari[3,]
Department of Computer Science and Engineering,
Rajalakshmi Institute of Technology,
Chennai, India.
maheswari.b@ritchennai.edu.in

A.Muthuram[4],
Department of Aeronautical Engineering,
Hindustan Institute of Technology and Science,
Chennai, India.
masrimuthu@gmail.com

***Abstract* - Electrical brain activity is detected by signals in an Electroencephalogram (EEG). Based on their frequencies, EEG signals are usually put into one of five groups: delta, theta, alpha, beta, and gamma. These signals help find a pattern that can be used to predict when a person will have a seizure. Classifying a seizure is a very important job for a doctor, as it helps them figure out what kind of seizure it is and if there will be any other problems. The goal of seizure classification is to learn as much as possible about the EEG signals. Literature shows that there are a lot of EEG signal pre-processing techniques, selection methods, feature extraction from EEG signals, and classification algorithms that can be used to find out if someone is having a seizure. The most important thing about pre-processing EEG signals is that it improves the quality of raw experimental data, which leads to better datasets, better classifications, and better accuracy. This study gives an overview of recent EEG pre-processing methods, datasets that can be used for experiments, and EEG classification techniques that will help a beginner researcher build on and use the right techniques.**

Keywords - Electroencephalogram (EEG); Classification; Decomposition; Preprocessing; Seizure

I. INTRODUCTION

The primary application of electroencephalogram (EEG) analysis is in the evaluation and diagnosis of brain and nervous system disorders. An EEG can detect any shifts or irregularities in brain activity. It is more time-consuming and challenging to make an EEG diagnosis visually. Furthermore, the number of neurologists is small. In order to automatically and rapidly diagnose the brain-neuro disease, a tool can be used to detect epilepsy, seizure, and other neurological disorder. This facilitates diagnosis and lessens time spent waiting for results. The signal-to-noise ratio is the main obstacle when analysing EEG data (SNR). Section II examines preprocessing strategies that have been developed to address these difficulties. In Section III, discuss how to use algorithms and machine learning to extract features from EEG signals for further analysis. Figure 1 provides a high-level overview of how EEG signals can be categorised.

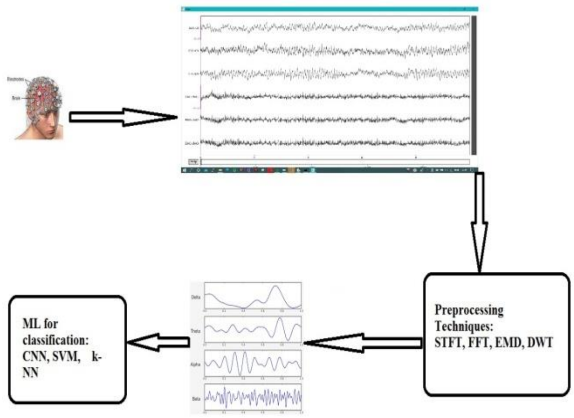

Fig.1. A schematic outline of EEG Signal Classification

II. LITERATURE SURVEY

The Neuronal oscillations in the nervous system aid the brain in its communication with its vast network of neurons. Neurological disorders are caused by abnormal electrical discharge in the brain. Together, millions of neurons produce signals and send them to the body's various systems. Non-linear data is abundant and adds complexity to these signals. Emotions can be detected by monitoring the continuous brain waves that humans constantly emit.

Electrical potential fluctuations brought on by neural networks in communication can be detected by means of EEG signals.

Electroencephalograms (EEGs) are often used to find out what's going on with the brain's electrical activity. Electrodes are put on the head of the patient to record the electrical activity of the brain [15]. This is what scalp signals from an electroencephalogram (EEG) look like. One way to describe neurological disorders is to look at the EEG signals that have been recorded [2]. Visually analysing the EEG data takes a lot of time and can take a few hours.

The preprocessing technique [3] transforms the raw signals into time frequency pictures. Smaller and medium-sized hospitals often lack the trained and experienced neurologists necessary for EEG analysis [4]. Multiple spectral thresholds are used to create permutations of frequency sub-bands. From the EEG signal's characteristics, we can determine the synchronisation pattern that correlates with seizure activity. The signals in the time domain are used to create the 2-dimensional images. The electroencephalogram (EEG) is the best way to diagnose epileptic syndromes and a cheap way to keep track of people over time.

Seizure classification is accomplished using the extracted ranges as inputs to various ML algorithms like SVM, CNN, and other data mining techniques. While each technique has its place, hybrid methods have been shown to improve accuracy . This proposed review and analysis lays the groundwork for furthering the study of EEG signal classification.

II. APPROACHES OF EEG SIGNAL PREPROCESSING METHODOLOGIES

This section goes into detail about the different ways to analyse EEG data before they are processed.

A. SHORT TIME FOURIER TRANSFORM:

To transform original signal into time-frequency images, Gaowei Xu et al. [3] proposed an STFT method. Time-frequency signal analysis employing the Fourier transform. Here, we figure out the amplitude of a signal that doesn't stay the same over time and frequency. By processing the raw EEG data, we were able to get features in both the time domain and the frequency domain. The mu band consists of frequencies from about 4 to 14 hertz, while the beta band features those from about 16 to 32 hertz.

STFT gives better results because it can find transient changes in EEG signals, which have a short length of time. STFT uses a 30 second window that doesn't overlap to find small changes. By mapping the EEG signal's amplitude and phase onto a two-dimensional matrix, Mingyang Li et al. [5] made an image of the brain's electrical activity using STFT. The most common brain state frequencies are Delta (0-4Hz), Theta (4-8Hz), Alpha (8-15Hz), Beta (15-30Hz), and Gamma (30-60Hz).

STFT was used by S. Raghu et al. [6] to turn a time series of EEG signals into an image. They did this by using a band pass filter with an intensity range of 0.1 to 44 Hz. The final single image features of a spectrogram are the spectra of all the individual channels stacked on top of each other.

B. DISCRETE WAVELET TRANSFORM:

Using the DWT method, Hafeez Ullah Amin et al. [2] were able to separate the signal into its component wavelet coefficient rhythms. This paper approximates the original signal using 4 level wavelet decomposition techniques. The final result is a reorganised signal free of quality loss and with the insignificant coefficients removed.

For their time-frequency series, Mahendra Yadava et al. [7] employed discrete wavelet transform. In doing so, it separated the original signals into narrower bands whose coefficients varied. In this case, high & low filters were used. The signal was cleaned high and low frequency noise by two separate filters. By iteratively applying the process across four distinct levels, the signal was decomposed into approximation and wavelet coefficients. Alpha (8–13 Hz), Theta (4–8 Hz), Delta (1–4 Hz),and Gamma (13–30 Hz) represent the subsets of frequency spectrum at each of the four echelons (8-100 Hz).

C. EMPIRICAL MODE DECOMPOSITION (EMD):

Wonsik The EMD procedure was used by Wonsik Yang et al. [4] to generate the IMFs. Raw EEG signals are dynamic and non-linear. For starters, we identified the signal's local maximum and minimum at each time segment. The second step was to find the midpoint. This model met the following two requirements for determining the IMF: I one maxima between two zero crossings; and (ii) mean value become 0. Following this procedure, the range of audible frequencies was narrowed down to the following five categories: Beta, alpha, delta, theta and gamma.

An ensemble EMD decomposition strategy was proposed by Ahnaf Rashik Hassan et al. [8]. It generated a signal in the form of time frequencies. To do this, it used a technique called intrinsic mode functions (IMF) to split the input signal up into discrete frequency bands. As an extension of Hilbert transform, Hesam Shokouh Alaei et al. [9] proposed the Hilbert Huamg Transform (HHT), an approach that combines the two methods. Beta, alpha, delta, theta and gamma were the designated frequency bands after EMD decomposition of the IMF

III. APPROACHES OF MACHINE LEARNING METHODOLOGIES USED IN EEG SIGNAL DATASET

This section describes machine learning methods such as CNN, SVM, and the K-NN Algorithms, as well as an in-depth assessment of the EEG data.

Using an image-based classification, S. Raghu et al. [6] proposed using CNN for seizure prediction. An initial step involved transforming the EEG signal over time into a spectrogram image stack. The SVM classifier was determined with the aid of an image feature extractor. However, the overlap method was able to solve the issue of uneven datasets. To better predict seizures, Wenbin Hu et al. [10] implemented a CNN-based method. Downsampling was used to eliminate extraneous data and background noise. The classifier for epilepsy is constructed by SVM based on the extracted features of CNN. Dropout was used to discard some of the randomly chosen feature to stop the overfitting issue.

Automatically identifying useful features of an EEG signal was achieved by Zuyi Yu et al., [11] using the CNN method. Extracted features provided crucial data for seizure event prediction. However, (BLDA) Bayesian linear discriminant analysis by regularisation fixed the overfitting issue. Successful feature extraction from invariant video patterns in VGGNet was achieved by Hengjin Ke et al. [12] using CNN (Visual Geometry Group). Images with dissimilar features were chosen in an adaptive fashion. Early stopping, in which units are dropped from training at random, however, eliminates the overfitting issue.

The CNN algorithm with optimization technique was proposed by Mohammad-Parsa Hossein et al. [14] and used for the extraction of unsupervised feature patterns. Overfitting, which causes deep learning methods to stall out at unhelpful "local maxima," was eliminated via optimization using PCA, ICA, and a DSA to locate optimal solutions. To extract and classify features from EEG signal time series data and images.

A CNN model for seizure prediction was proposed by Xiaobin Tian et al. [1], which leveraged multi-view characteristics to extract deep view features. The dimensionality of this feature is reduced, making it more discriminatory. The different perspectives combine elements from the TD, the FD, and the TFD. Overfitting is a problem with the highly unbalanced data, but this is mitigated by the removal of non-seizure data that was previously used.

The CNN model was proposed by Hengjin Ke et al. [12] due to its ability to automatically extract features and share their weights, allowing for the capture of correlated EEG signals within an image. The feature patterns were processed using this technique after each iteration. However, a 10-fold CV approach was applied during training to eliminate the overfitting issue.

In classification of seizure analysis, Wonsik Yang et al. [4] used an SVM model for the feature selection process. The pattern recognition problem was solved effectively by SVM, which located the boundary that correctly categorised the most closely related training samples. SVM's recursive feature elimination (RFE) process mitigated the overfitting issue.

SVM and k-NN algorithms were used by MarziehSavadkoohi et al. [13] to categorise the preprocessed signal. In this case, the SVM algorithms used the resulting features. The programme used the Euclidean distance to determine how far apart each sample was. Using a nearest-to-greatest distance measure between each data point, an SVM classifier determined the optimal hyper plane line. Cross validation helped get rid of the issue of overfitting. For the purpose of seizure classification, Mingyang Li et al. [5] proposed a SVM with a RBF algorithm. SVM classification was used to determine to which state the EEG Signal belonged by incorporating the features in a sequential order. When it comes to EEG recognition, the SVM algorithm excels.

IV. COMPARISON OF EEG SIGNAL CLASSIIFICATION METHODOLOGIES

In this section, the examined methods are summarised, and their advantages and disadvantages are contrasted. Table 1 shows that various preprocessing techniques for EEG signals, including the Fourier Transform (which is further subdivided into the FFT, STFT, DWT, and Empirical Mode Decomposition, etc., offer improved accuracy, sensitivity, and specificity while requiring less computational effort. In Table 2 we see a comparison of EEG-based machine learning techniques. The reader can use this table to better comprehend the various machine learning methods and the purpose of the study.

TABLE 1
COMPARISON OF PREPROCESSING METHOD AND THE FREQUENCY RANGES

Authors	Preprocessing Methodology	Dataset	Frequency bands
Wonsik Yang et al. 2020	EMD	Severance Children's Hospital (Yonsei University, Seoul, South Korea)	delta(0.5−3.5Hz) theta (3.5−8 Hz) alpha (8−13 Hz) beta (13−30 Hz) gamma (>30 Hz)
Ahnaf Rashik Hassan et al. 2019	EMD	University of Bonn's	--
Hesam Shokouh Alaei et al. 2019	HHT combination of Empirical Mode Decomposition (EMD) and Hilbert transform	Center of University Hospital of Freiburg, Germany	gamma(30-70Hz) beta (13–30Hz) alpha (8–13Hz) theta (4–8 Hz) delta (0.5–4 Hz)
Gaowei Xu, et al. 2019	STFT	University of Bonn & University Hospital of Freiburg	Mu (4-14Hz) beta (16-32 Hz)
Mingyang Li et al. 2019	STFT	Department of Epileptology, Bonn University	gamma(30-60Hz) beta(15-30Hz) alpha(8-15Hz) theta(4-8Hz) delta(0-4Hz)
Hafeez Ullah Amin et al. 2020	DWT	Bonn University	---
Mahendra Yadava et al. 2018	DWT	University of Bonn and University Hospital of Freiburg	gamma (8-100 Hz) alpha (8-13 Hz) theta (4-8 Hz) delta (1-4 Hz)

Table 2
COMPARISON OF MACHINE LEARNING METHODS USING EEG

Authors	Approach	Dataset
S. Raghu et al. 2020	CNN, SVM	Publicly available epilepsy data-set
Wenbin Hu et al. 2019	CNN , SVM	CHB-MIT database
Zuyi Yu et al. 2018	CNN	Center for Epilepsy - Freiburg University Hospital, Germany
Wonsik Yang et al. 2020	SVM	Severance Children's Hospital (Yonsei University, Seoul, South Korea,
Hengjin Ke et al. 2017	CNN	The CHB-MIT scalp EEG dataset
Marzieh Savadkoohi et al. 2020	SVM, k-NN	University of Bonn,
Mingyang Li et al. 2019	SVM	Department of Epileptology, Bonn University 5 Dataset
Mohammad-Parsa Hossein et al. 2017	CNN	Mayo clinic, University of Pennsylvania and sponsored by the American Epilepsy Society

Xiaobin Tian et al. 2019	CNN	The CHB_MITs dataset Boston Children's Hospital

V. CONCLUSION

The objective of this survey is to point out the difficulty of pre-processing EEG signals and the advantages of extracting frequency bands for identifying and classifying neural abnormalities like epilepsy and seizures. The information from the dataset was used to analyse and summarise a number of the best existing approaches to classification. The analysis stands out from the literature in several key respects. First, we give special attention to the pre-processing techniques and the rationale behind the transformation of the frequency bands into ranges. Some of these approaches directly apply the EEG signals to machine learning algorithms, while others pre-process the EEG signals before feeding them into the machine learning algorithms for feature selection methods, all of which are summarised here. From these results, the proposed study can infer that the most precise prediction can be achieved by employing pre-processed methods rather than raw signals. Multiple pre-processing techniques for EEG signals (like STFT, FFT, and WT) can be used to create frequency bands. From the granular bands, the bare minimum of features is chosen to produce meaningful outcomes. The key is to avoid choosing superfluous details that will increase processing time and produce meaningless patterns. If fewer features are used than necessary, a low-dimensional dataset is produced, which hinders effective process discovery. The need to develop one's own hybrid model of epilepsy prediction grows as more is discovered about how to categorise and forecast seizures.

REFERENCES

[1] XiaobinTian, Zhaohong Deng, Wenhao Ying, Kup-Sze Choi, Dongrui Wu, Bin Qin, Jun Wan, Hongbin Shen and Shitong Wang, "Deep multi-view feature learning for EEG-based epileptic seizure detection," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, pp. 1-11, 2019. DOI: 10.1109/TNSRE.2019.2940485

[2] Hafeez Ullah Amin, Mohd Zuki Yusoff and Rana Fayyaz Ahmad, "A novel approach based on wavelet analysis and arithmetic coding for automated detection and diagnosis of epileptic seizure in EEG signals using machine learning techniques," *Biomedical signal processing and control, Elsevier*, Vol. 56, pp. 1-10, 2020. https://doi.org/10.1016/j.bspc.2019.101707

[3] Gaowei Xu, Xiaoang Shen, Sirui Chen, Yongshuo Zong, Canyang Zhang, Hongyang Yue, Min Liu, Fei Chen and Wenliang Che, "A deep transfer convolutional neural network framework for EEG signal classification," *Special Section on Data-Enabled Intelligence for Digital Health, IEEE Access*, Vol. 4, pp. 1-10,2016.DOI:10.1109/ACCESS.2019.2930958

[4] Wonsik Yang, Minsoo Joo, Yujaung Kim, Se Hee Kim and Jong-Moon Chung, "Hybrid machine learning scheme for classification of BECTS and TLE patients using EEG brain signals," *IEEE Access*, Vol.8, pp. 218924218325,2020.DOI:10.1109/ACCESS.2020.3038948

[5] Mingyang Li, Xiaoying Sun, Wenzhen Chen, Yun Jiang and Tao Zhang, "Classification epileptic seizures in EEG using time-frequency image and block texture features," IEEE Access, Vol. 8, pp. 9770-9781, 2019.DOI:10.1109/ACCESS.2019.2960848

[6] S. Raghu, Natarajan Sriraam, YasinTemel, ShyamVasudeva Rao and Pieter L. Kubben, "EEG based multi-class seizure type classification using convolutional neural network and transfer learning," *Neural Networks, Elsevier,* Vol. 124, pp.202-212,2020.https://doi.org/10.1016/j.neunet.2020.01.017

[7] Mahendra Yadava, Pradeep Kumar, Rajkumar Saini, Partha Pratim Roy and Debi Prosad Dogra, "Analysis of EEG signals and its application to neuromarketing," *Multimed Tools Applications, Springer*, Vol. 76, pp. 19087–19111, 2018. DOI 10.1007/s11042-017-4580-6

[8] Ahnaf Rashik Hassan and Abdulhamit Subasi, Yanchun Zhang, "Epilepsy seizure detection using complete ensemble empirical mode decomposition with adaptive noise," *Knowledge-Based Systems, Springer,* Vol. 191, pp. 1-12, 2019. https://doi.org/10.1016/j.knosys.2019.105333

[9] Hesam Shokouh Alaei, Mohammad Ali Khalilzadeh and Ali Gorji, "Optimal selection of SOP and SPH using fuzzy inference system for on-line epileptic seizure prediction based on EEG phase synchronization," *Australasian Physical & Engineering Sciences in Medicine, Springer,* Vol.42, no. 4, pp. 1049-1068, 2019. https://doi.org/10.1007/s13246-019-00806-w

[10] Wenbin Hu,Jiuwen Cao, ·Xiaoping Lai and Junbiao Liu, "Mean amplitude spectrum based epileptic state classification for seizure prediction using convolutional neural networks," *Journal of Ambient Intelligence and Humanized Computing, Springer,* 2019. https://doi.org/10.1007/s12652-019-01220-6

[11] Zuyi Yu, Weiwei Nie, · Weidong Zhou, Fangzhou Xu, Shasha Yuan,Yan Leng and· Qi Yuan, "Epileptic seizure prediction based on local mean decomposition and deep convolutional neural network," *The Journal of Supercomputing, Springer,* Vol. 76, no. 5, pp. 3462-3476,2018.https://doi.org/10.1007/s11227-018-2600-6

[12] Hengjin Ke, Dan Chen, Xiaoli Li, Yunbo Tang, Tejal Shah and Rajiv Ranjan, "Towards brain big data classification: Epileptic EEG identification with a lightweight VGGNet on global MIC," *IEEE ACCESS, Special section on cyber-physical-social computing and networking,* Vol. 6, pp. 14723-14733,2017.DOI:10.1109/ACCESS.2018.2810882

[13] Marzieh Savadkoohi, Timothy Oladunni and LaraThompson, "A machine learning approach to epileptic seizure prediction using Electroencephalogram (EEG) signal," *Biocybernetics and Biomedical Engineering, Elsevier*, Vol. 40, no. 3, pp. 1328-1341, 2020. https://doi.org/10.1016/j.bbe.2020.07.004

[14] Mohammad-Parsa Hosseini, Dario Pompili, Kost Elisevich, and Hamid Soltanian-Zadeh, "Optimized deep learning for EEG big data and seizure prediction BCI via internet of things," *IEEE Transactions on Big Data*, Vol. 3, no. 4, pp. 392-404, 2017. DOI: 10.1109/TBDATA.2017.2769670

[15] Simon, Judy, and A. Naveen Kumar. "A Novel Signal Processing Based Driver Drowsiness Detection System." Journal of IoT in Social, Mobile, Analytics, and Cloud 3, no. 3 (2021): 176-190. DOI: 10.1109/TBDATA.2017.2769670