

Automated classification of eight different Electroencephalogram (EEG) bands using hybrid of Fast Fourier Transform (FFT) with machine learning methods

Nur Shahirah Md Nor ¹, Nurul Hashimah Ahamed Hassain Malim ^{1,*}, Nur Aqilah Paskhal Rostam ¹, J. Joshua Thomas ², Mohamad Azmeer Effendy ³ and Zurina Hassan ³

¹ School of Computer Sciences, Universiti Sains Malaysia, 11800 Pulau Pinang, Malaysia.

² UOW Malaysia, KDU Penang University College, 10400 Pulau Pinang, Malaysia.

³ Centre for Drug Research, Universiti Sains Malaysia, 11800 Pulau Pinang, Malaysia.

* Correspondence: nurulhashimah@usm.my; Tel.: +604 653 4645

Received: 23 August 2021; **Accepted:** 30 September 2021; **Published:** 2 March 2022 (Version 2: 5 March 2022)

Edited by: Indranath Chatterjee (Tongmyong University, South Korea)

Reviewed by: Heru Suhartanto (Universitas Indonesia, Indonesia);

Muhamad Kamal Mohammed Amin (Universiti Teknologi Malaysia, Malaysia)

<https://doi.org/10.31117/neuroscirn.v5i1.116>

Abstract: Analysing and processing the EEG dataset is crucial. Countless actions have been taken to ensure that the researcher in brain studies always achieves informative data and produces notable findings. There are several standard procedures to produce an informative result in analysing the EEG data. However, the techniques used in each standard procedure might be different for the researcher or data analyst because they have their preferences to suit the purpose of their experiments to adapt with the dataset collected. Not only the current manual method is time-consuming, but the main challenges are that researchers need to analyse only a small portion of the brain signals that are the most relevant to be observed through the analysis of several bands such as Very low, Delta, Theta, Alpha-1, Alpha-2, Beta-1, Beta-2, and Gamma. Therefore, one of the best alternatives is to automate the process of classifying the eight bands and extract the most relevant features. Hence, this paper proposed an automated classification method and feature extraction method through hybridising Fast Fourier Transform (FFT) with three different machine learning methods (KNN, SVM, and ANN) that can improve the efficiency of EEG analysis. Based on the result, the FFT + SVM method gives a 100% accuracy and successfully classified the bands into different of eight EEG bands accurately.

Keywords: Electroencephalogram analysis; Time Series Classification; Fast Fourier Transform; K-Nearest Neighbor; Artificial Neural Network;

©2022 by Md Nor *et al.* for use and distribution according to the Creative Commons Attribution (CC BY-NC 4.0) license (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

1.0 INTRODUCTION

Brain research has been conducted for decades and is still ongoing for a long time until today. Each year, different brain diseases and disorders affect millions of people, such as infections, seizures, trauma, tumours,

masses, increased pressure, autoimmune conditions, and neurodegenerative conditions (Siuly *et al.*, 2016). Even though brain studies are complicated, it gives people many advantages as the brain is an essential part of a human. Researcher in this field potentially

finds a cure, finding treatments to improve the quality of life or increase the revenue of the businesses. Hence, billions of dollars are granted in this field toward understanding the brain functions and their effects but the outcomes are scattered and inconclusive ([Chu, 2017](#)).

Towards the years of technology expanding era, brain research has integrated physiological and engineering innovations in experiments. In 1929, a German clinical neuropsychiatrist at the University of Jena, Hans Berger ([Duffy, 1981](#)), was the earliest person who recorded the electrical activity of the human brain and introduced the term electroencephalogram or EEG. The recorded EEG is from an electrical connection between the patient's brain attached to an electrode, and the other end of the electrode is plugged into the Electrocardiogram (ECG) machine. In Berger's findings, brain electrical activity consists of a mixture of rhythmic like sinusoidal fluctuations in voltage having a frequency of about 1 to 60 oscillations per second. He has also introduced two types of rhythms called bands: Alpha and Beta. Alpha is examined when the frequency is about ten oscillations per second, while if the frequency of waves is more than 15 oscillations per second, it is called Beta. Also, the mixture of rhythmic in EEG consists of different features in neurological disorders, and Berger was the first recorded epileptic seizure ([Duffy, 1981](#)).

The recorded EEG waveform is raw data that consists of unnecessary waves that the researcher does not need. For example, the initial EEG recorded of the rat brain will consist of the abnormal waveform. First, the rat might feel uncomfortable when the electrodes are placed in its brain, causing the rat to try to take off the electrodes from its brain. This abnormal waveform is called artefacts similar to noise data. After removing the artefacts by getting the correct time of EEG recording, analysing the data usually takes up hours or even days to complete. The difficulties of this experiment are; it is in a manual way where the standard EEG waveform is applied with the calculation of mean square values in excel and manually categories the bands into eight types of frequencies. The situation explained is a real-life situation and experiments taken by one of the students and researchers in the Center for Drug Research (CDR) of Universiti Sains Malaysia (USM). The group is currently experimenting with the drug habituation inside a body. The drugs used, such as morphine and methamphetamine, are used on the rats. The inconvenient problems faced by the group have

motivated the researchers to categorise the eight types of frequency bands collected from real lab experiment data using machine learning classification techniques.

The current situation in CDR USM has led to an impractical method in analysing EEG data for drug habituation experiments due to time-consuming processes and takes approximately up to 2 months to complete the process. For the researcher to analyse the data, they will use the Fourier series analysis method. Mean square values are often used to express the component amplitude of Fourier Series analysis. The resulting plot of the data is called a power spectrum. It can then analyse the waveform of each frequency series into five types of bands: Delta, theta, Beta, alpha, and gamma band. These five bands will give a different type of analysis in brain studies by depending on the investigation or experiments that have been done to achieve the objectives.

Brain electrical activity consists of a mixture of rhythmic, sinusoidal-like fluctuations in voltage, having a frequency of about 1 to 60 oscillations per second. The frequency of brainwaves is below 1 Hz up to 100 Hz. Band categories may vary in the experiments conducted. Such as, a researcher in CDR USM are using eight bands in their research for drug habituation which is very low (0 – 1.24 Hz), Delta (1.25 – 4.5 Hz), theta (4.75 – 6.75 Hz), alpha-1 (7 – 9.5 Hz), alpha-2 (9.75 – 12.5), beta-1 (12.75 – 18.5 Hz), beta-2 (18.75 – 35 Hz) and Gamma (35.25 – 45 Hz). In analysing the EEG data, there are several standard procedures to produce an informative result (Al-Fahoum & Al-Fraihat, 2014). However, the techniques used in each standard procedure might be different for the researcher or data analyst because usually depending on the preferences to adapt with the dataset collected.

The procedure will usually start with data collection, where different tools or machines are used to record the EEG data. For example, if the experiments require data of the human brain, they must use the electrodes scalp to get the data, and if for the animal, a different tool will be used where the electrodes must be inside the animal brain. The typical tool used by researchers to record human brain EEG is an Emotiv EPOC headset ([Balaji et al., 2017](#); [Bastos et al., 2020](#); [Chan et al., 2015](#); [Murugappan et al., 2014](#); [Murugappan et al., 2013](#)). Meanwhile, to record an animal brain, such as rats or rabbits, the researcher usually uses similar tools for all animals: stainless steel electrodes placed inside the animal brain ([Kortelainen et al., 2012](#)). Minor

surgery is needed to open the outer layer of the head to the skull so that a plastic plug is placed in the skull and connected to the electrodes. The plastic plug and electrodes are fixed using dental cement. The animal then will be given time to rest for recovery before starting the EEG recording. Feature extraction is a process to reduce the loss of important information embedded in the signal ([Al-Fahoum & Al-Fraihat, 2014](#)).

The primary objective of feature extraction is to achieve pertinent information from the original data and represent it in a lower dimensionality space ([Kumar & Bhatia, 2014](#)). A recent study has shown many methods for extracting the features from EEG signals such as Fast Fourier Transform (FFT), Time-Frequency distribution (TFD), Wavelet Transform (WT), Auto-Regressive Method (ARM), and Eigenvector Methods (EM) ([Al-Fahoum & Al-Fraihat, 2014](#); [Asadur Rahman et al., 2020](#); [Chao et al., 2020](#); [Petrovska et al., 2020](#)). These methods have their advantages and disadvantages, which give different purposes for different signal datasets. It depends on what feature needs to extract in a specific dataset.

The two most common algorithms used for the extraction wave dataset are Fast Fourier Transform (FFT) and Wavelet-based algorithm. The FFT algorithm is sophisticated and has become well-known due to its efficiency in calculating the Discrete Fourier transform (DFT) ([Heideman et al., 1984](#)), the formula for evaluating the N Fourier coefficients from a sequence of N numbers. While Wavelets are mathematical functions that represent data or other functions regarding the averages and differences of a prototype function ([Murugappan et al., 2013](#)) and is both a band-pass filter and a denoiser for decomposing and isolating EEG signals to obtain desired subbands, for example, extracting only alpha, beta, and gamma frequencies as they are more related to emotion elicitation.

Comparing FFT and Wavelet algorithm, even though Wavelet can consistently analyse irregular data patterns, Wavelet is difficult to find and select an actual mother wavelet. Wavelet is not suitable for the experiment related to the changes that affect drug habituation because its ability in extracting the small changes of the noise-corrupted signal is considered too detailed for the experiment. FFT, however, offer many benefits if compared to wavelet such as it can serve as a good tool for stationary signal processing and suitable for frequency domain analysis method, which

is more appropriate for a narrowband signal such as a sine wave and has an enhanced speed over virtually all other available methods in real-time applications. Moreover, CDR USM has been used the FFT method to conduct the analysis. According to their standard guidelines, FFT is suitable for EEG in drug habituation. Its feature in the EEG signal extracts the raw time-domain signal into the frequency domain and finds the Power Spectral Density (PSD) with units of V^2/Hz . This characteristic suits the primary purpose of our research that aimed to automate a laborious and time-consuming process.

Consequently, there are two detailed objectives of this research to be achieved, which are (i) To propose a hybrid of FFT and classification method and (ii) To investigate by experiment the effect of different classification using machine learning methods to improve the overall automated classification process. Many brain studies have combined machine learning techniques to achieve the objectives proposed as accurately as possible. Machine learning is an algorithm that allows the system to learn to predict the outcomes accurately ([Mohammed et al., 2016](#)). Moreover, most researchers have focused on the supervised classification method to classify the bands. There are different methods of classification such as k-Nearest Neighbor (KNN), Support Vector Machine (SVM), Decision trees, Naïve Bayes, and Neural Network (NN). Various past works related to classification and wave dataset was to categories epilepsy and non-epilepsy of brainwaves ([Anwar et al., 2020](#); [Sarmast et al., 2020](#); [Si, 2020](#)), categories in emotional states ([Masrurroh et al., 2019](#); [Selvaraj et al., 2013](#); [Shu et al., 2020](#)), classification in heart disease ([Almustafa, 2020](#); [Gárate-Escamila et al., 2020](#); [Javeed et al., 2020](#)), wave height prediction ([Berbić et al., 2017](#); [Li & Liu, 2020](#)) and automatic sleep stage classification ([Chriskos et al., 2017](#); [Giannakaki et al., 2017](#)).

For the past years, machine learning algorithms have helped in brain imaging and the computational neurosciences developed to signalise task-relevant brain states and distinguish them from non-informative brain signals ([Barranco-Gutiérrez, 2020](#); [Lemm et al., 2011](#); [Michel et al., 2019](#)). Hence, similar to the feature extraction explained above, different classification methods, as discussed previously, such as KNN, SVM, and ANN, will be experimented with and compared, and the best method will be the final method. As for EEG analysis, Fast Fourier will be the feature extraction method.

2.0 MATERIALS AND METHODS

This section discusses the implementation of the methodology used to achieve our objectives. The illustration of the methodology is shown in **Figure 1**.

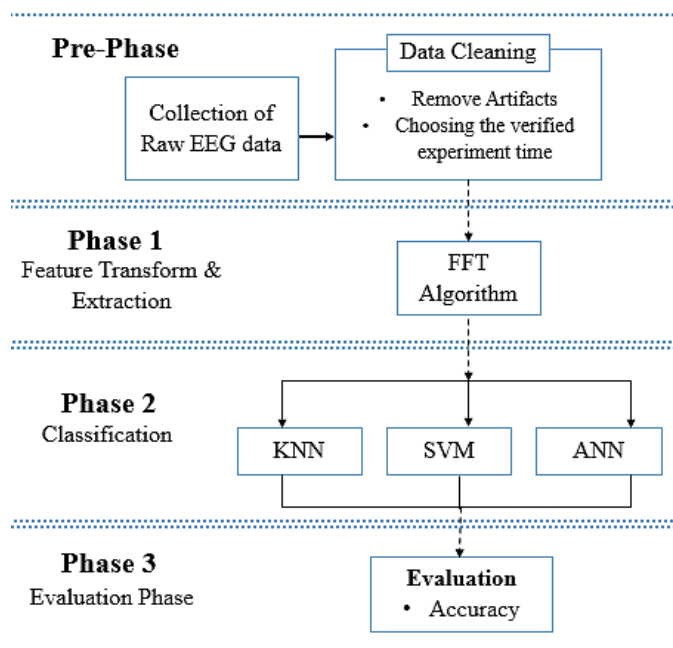


Figure 1: The overall framework of the processes involved in all the experiments conducted in this study.

2.1 Pre-Phase

Data Collection

The data collection recorded from the rat's brain is provided by CDR USM. The wave pattern of the data is based on time-domain representation. The recording length is 90 minutes to testify before and after the rat consumes drugs, and only 10 seconds will be taken into the calculation for analysis. In this experimental study, data collected for one rat reading consists of 4 channel regions. The experiment used one rat to test the habituation effect such as addiction. Hence, the experiment will only need one rat. However, if more than one rat is involved, all rats from the same channel regions will be average to get the final reading. LabChart software is used to record the wave data from the four electrodes that were attached from the rats' brains. Major regions of the rat's brain are the Frontal Cortex, Parietal Cortex, Right Hippocampus, and Left Hippocampus, also called sensory ([Kondo et al., 2001](#)). The EEG data consists of time and amplitude, which made up the wave.

The data structure in LabChart software into the data input for classifier algorithm is mainly in waveform.

The recorded waves can be analysed based on their amplitude (microvolt), frequency (Hz/sec), duration and contour (spike, sharp, slow, arch, spike-wave, poly spike) and their relation to different states of the wake-sleep cycle and age of the subject (rat).

Data Cleaning

In cleaning the data, the first strategy is eliminating noise or artefacts. Usually, this strategy avoided environmental noises such as AC power lines, lighting, and other sources from electronic equipment which produces electromagnetic. This equipment will be able to track the EEG recording. The rat will be placed in a black box without light for the experiment to imitate a rat habitat that always lives in a darker place. To avoid any distraction, the area of the experimental study prohibit any movement from human or anything hence no one is allowed to go to the experimental area as it will distract the nature of the rats.

However, since the rats are in pain due to the insertion of the electrode in their brain, the uncomfortable feeling, artefacts or noise are still tough to avoid. Usually, CDR USM will use visual inspection to reject the artefact. This method is one of the most useful methods and is preferred the most for the researchers in signal data. Meanwhile, an experiment was done by Lan et al. ([2017](#)) for emotion recognition using EEG, and they removed twenty-one EEG channels using this method due to the signal quality issues such as loose electrode contacts to the subject. Anusha et al. ([2012](#)) and Tzallas et al. ([2017](#)) use a similar method in the pre-processing stage for experiments to classify normal and epileptic EEG signals using Artificial Neural Network (ANN) and Brain-Computer Interface, respectively.

The way CDR USM identifies the noise data is based on the large and dense amplitude signal and the large spike shown in the EEG signal. Due to the expertise in identifying the noisy data, the CDR USM group usually manually monitors and studies the rats' movement while the EEG signal is recorded. Thus, after several experiments, the group successfully learned to remove the noise by visual inspection. As shown in **Figure 2**, waves that the group will use to analyse the data for their study, which is the verified time for the drug to habituate, affect the brain based on the group related work studies.

After choosing the specific time and the regular waves, the data type will be changed to a suitable format in numerical form as shown in **Figure 3** and **Figure 4**, such

as the *.txt file that can be loaded into the RStudio. The loaded EEG dataset does not define the column correctly. Therefore, V1 until V5 column in Figure 21 shows the raw EEG dataset and the standard variables used in RStudio. The loaded raw EEG data consist of 794,555 entries which approximately 90 minutes of the experiment takes place, as shown in **Figure 3**.

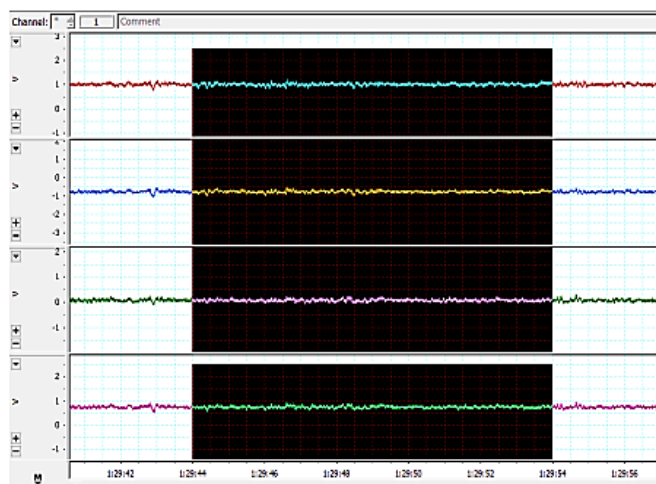


Figure 2: Example of the waves used for analysis.

	V1	V2	V3	V4	V5
1	Range=	10.000 V	10.000 V	10.000 V	10.000 V
2	0.0	-0.597	0.078	0.504	0.953
3	0.001	-0.592	0.077	0.507	0.945
4	0.002	-0.601	0.057	0.490	0.927
5	0.003	-0.576	0.071	0.496	0.926
6	0.004	-0.568	0.071	0.483	0.913
7	0.005	-0.548	0.090	0.491	0.914
8	0.006	-0.560	0.070	0.456	0.889
9	0.007	-0.568	0.061	0.439	0.879
10	0.008	-0.568	0.058	0.425	0.869
11	0.009	-0.590	0.039	0.392	0.846
12	0.01	-0.612	0.026	0.369	0.828
13	0.011	-0.607	0.030	0.362	0.831
14	0.012	-0.630	0.013	0.334	0.803
15	0.013	-0.581	0.072	0.393	0.853
16	0.014	-0.693	-0.035	0.285	0.755
17	0.015	-0.656	-0.001	0.321	0.781
18	0.016	-0.469	0.189	0.511	0.941
19	0.017	-0.492	0.170	0.502	0.913
20	0.018	-0.518	0.142	0.460	0.889
21	0.019	-0.521	0.138	0.446	0.902
22	0.02	-0.545	0.118	0.420	0.883
23	0.021	-0.543	0.126	0.429	0.897
24	0.022	-0.563	0.116	0.434	0.899
25	0.023	-0.570	0.111	0.445	0.898

Figure 3: An example of raw EEG dataset

Then, the columns were renamed based on the brain region as "Time", "FrontalCortex", "ParietalCortex", "RightHippocampus", "LeftHippocampus". Hence the outcome of the EEG dataset after the Pre-Phase is shown in **Figure 4**.

2.2 Phase 1: Preprocessing and Feature Selection

After the cleansing part, the FFT algorithm extracts the number of points in the raw EEG signal and transforms it from time-domain representation into frequency domain representation. The time-domain representation is shown in **Figure 5**.

Figure 6 is the frequency domain representation after we implemented FFT into the dataset in R. According to CDR USM, windowing functions reduce the importance of data at the edges of the window in waveform and prevent spurious peaks arising from edge effects. The outcome has changed to frequency-domain representation with a maximum frequency value of 500 Hz. The figure has four different colours, which shows four different brain regions. The green colour represents Frontal Cortex, the red colour represents Parietal Cortex, the blue colour represents Right Hippocampus, and the orange colour represents Left Hippocampus.

	Time	FrontalCortex	ParietalCortex	RightHippocampus	LeftHippocampus
2	0.0	-0.597	0.078	0.504	0.953
3	0.001	-0.592	0.077	0.507	0.945
4	0.002	-0.601	0.057	0.490	0.927
5	0.003	-0.576	0.071	0.496	0.926
6	0.004	-0.568	0.071	0.483	0.913
7	0.005	-0.548	0.090	0.491	0.914
8	0.006	-0.560	0.070	0.456	0.889
9	0.007	-0.568	0.061	0.439	0.879
10	0.008	-0.568	0.058	0.425	0.869
11	0.009	-0.590	0.039	0.392	0.846
12	0.01	-0.612	0.026	0.369	0.828
13	0.011	-0.607	0.030	0.362	0.831
14	0.012	-0.630	0.013	0.334	0.803
15	0.013	-0.581	0.072	0.393	0.853
16	0.014	-0.693	-0.035	0.285	0.755
17	0.015	-0.656	-0.001	0.321	0.781
18	0.016	-0.469	0.189	0.511	0.941
19	0.017	-0.492	0.170	0.502	0.913
20	0.018	-0.518	0.142	0.460	0.889
21	0.019	-0.521	0.138	0.446	0.902
22	0.02	-0.545	0.118	0.420	0.883
23	0.021	-0.543	0.126	0.429	0.897
24	0.022	-0.563	0.116	0.434	0.899
25	0.023	-0.570	0.111	0.445	0.898
26	0.024	-0.570	0.117	0.459	0.912

Figure 4: An example of the outcome of EEG dataset after pre-phase.

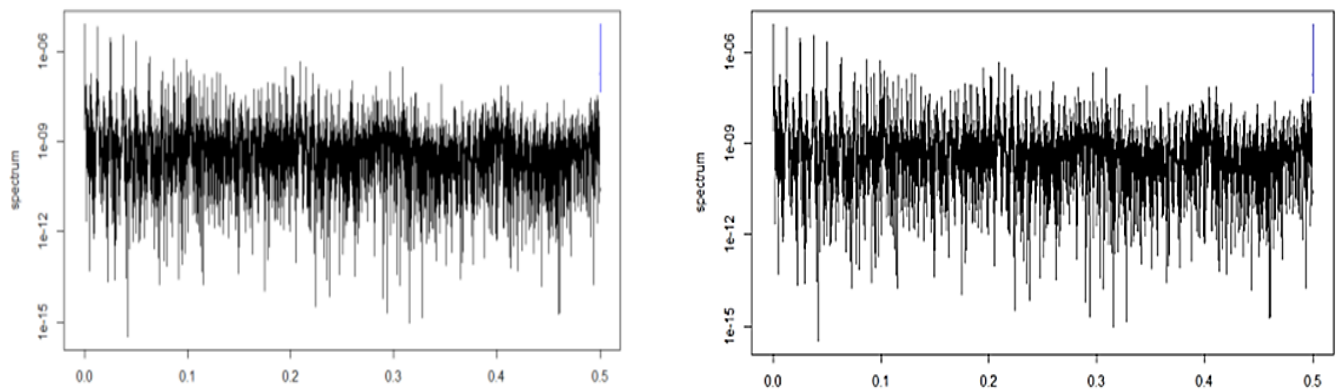


Figure 5: Raw periodogram time-series representation.

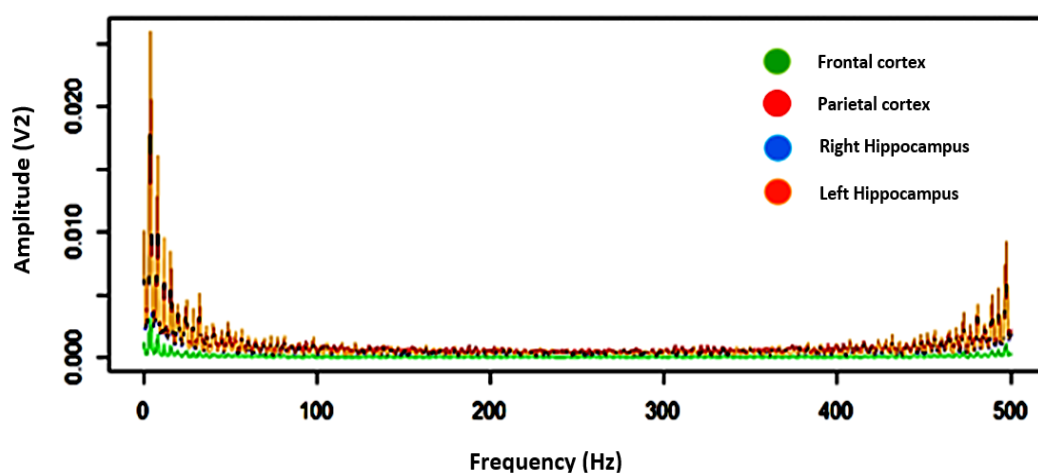


Figure 6: Frequency-domain representation.

2.3 Phase 2: Classification Phase

Support Vector Machine (SVM) is a supervised machine learning algorithm used for classification or regression problems. The SVM method came from the idea of splitting points of different classes in clouds of points with a line as the optimally distanced from the classes (Berbić et al., 2017). K-Nearest Neighbor (KNN) is another approach in wave dataset classification. KNN classifies objects based on the closest training examples in the feature space, also called a type of passive learning (Imandoust & Bolandraftar, 2013). This classification method has been used in many applications such as problem-solving, classification, and interpretation also for function learning and teaching and training (Jabbar et al., 2013; Al-Janabi et al., 2020, 2018). This is because KNN is a simple and straightforward algorithm for pattern recognition. Other than SVM and KNN, Artificial Neural Network (ANN) is also one of the most used classifier methods in the signal dataset. ANN is meant to be the simulation of the thinking process in the human brain

in which the artificial neurons are interconnected (Maksimenko et al., 2018; Nguyen et al., 2019; Xiong et al., 2020). It is also called ANN as an extraordinary black box that is trained to achieve the expected intelligent process against the input and output information stream. As ANN is considered deep learning due to the layers of ANN, ANN is self-teaching, learning as it goes by filtering information through multiple hidden layers in a similar way to humans. The application of Artificial Neural Networks (ANN) in the classification of neural network signals may help even non-experts interpret the results (Dhanapal and Bhanu, 2020). There are other more advanced variants of ANN. However, for simplicity and due to only one rat is experimenting, ANN is quite a powerful algorithm where more advance deep learning can be used when the volume of data is much higher, and many rats need to experiment.

Before any of the classification methods were applied to the EEG dataset, we split the dataset into training

and test datasets by using 70 to 30 ratios. The 70:30 ratio for splitting data was chosen because it is often used in data mining, and however, it may vary depending on the experiment to suit the requirement.

The first classification method that will be used is the KNN algorithm. KNN in data mining and predictive modelling refers to a memory-based (or instance-based) algorithm for classification and regression problems ([Imandoust & Bolandraftar, 2013](#)). However, in this experiment, we are also experimenting with various values for k, which varies from 1 to 50. This is because we can also investigate the performance of KNN by observing each possible k value to find the optimal performance. Value k was chosen up to 50 was adopted from a study done by Mustafa and Lokman ([2014](#)). We have also tried to use a k value less than 50, but it gave an error due to the data frame element of parameter training, and the test dataset is not suitable for plotting value k less than 50.

After that, we proceed with SVM and ANN classification. For SVM classification, the library used in R is "e1071", which is for miscellaneous functions of the Department of Statistics, Probability Theory Group. The kernel is a function work by transforming the input state space to a higher-dimensional space, where the data can be linearly separated ([Andre et al., 2013](#)). However, before the SVM function was called, we tested the various cost values from 1 to 8 using the tune function. The result showed that the cost of tune function of value 7 gives high accuracy. Hence, value 7 is chosen. The 'tune' function from the 'e1071' package in R is to tune the hyperparameters of SVM using a grid search algorithm where it depends on the application.

The last classification, ANN, uses library "nnet" for feed-forward Neural Networks and Multinomial Log-Liner Models. The minimum number of neurons in the hidden layer was chosen to be eight because we need to classify eight bands of the EEG signal. Overall the layer of ANN implementation consists of input, eight hidden layers, and one output layer that will classify the band. **Table 1** summarised the parameter set used in this experiment

2.4 Phase 3 – Final Phase

There are various methods in evaluating the quality of classification models ([Novakovic et al., 2017](#)). In this paper, the performance of each classification method was evaluated to see how well the method classified the bands accordingly. Therefore, we use accuracy as the evaluation metric. The formula for accuracy

evaluation was adopted from Novakovic et al. ([2017](#)), which is described as the number of correctly classified examples divided by the total number of cases.

Table 1. Parameter settings.

Algorithm	Parameter	Value
FFT	Frequency	Min=0 Max=500
	Window Size	512
KNN	k value	1 till 50
SVM	Cost	7
	Kernel	Radial
ANN	Hidden Layer	8
	Decay	5e-4
	Maxit	200

3.0 RESULTS AND DISCUSSION

3.1 Result of FFT + KNN

In the FFT + KNN classification experiment, the results show that the highest accuracy for KNN is 29%, with k value is 1. As the k value increases, accuracy is depleted to zero value, and thus, the increasing k value is not valid. The statement above aligns with Xie ([2012](#)), where a smaller number of neighbours will produce more accurate but higher variance prediction and vice versa for a larger value. According to Kim et al. ([2012](#)), the main advantage of the KNN algorithm is that it performs well with multi-modal classes because the basis of its decision is based on a small neighbourhood of similar objects that still lead to good accuracy. However, in this experiment, KNN is not a good choice to perform eight bands classification because the data point signal value is too close to each other.

3.2 Result of FFT + SVM

The goal of SVM is to determine classes of observations and build boundaries to predict which future class observations belong to based on the measurements. Thus, the parameter used in SVM consists of two, which are using the kernel and the cost of misclassifications. The tune command was used to find the best-fitting value cost and calculate the accuracy for each value. We have set the tuning value from 1 to 8. The result turned out that the value cost 7 gave higher accuracy, which is 100%. Hence, we used the SVM command in training and test datasets with cost value 7 to classify the eight bands. Both give the same trend of 8 frequency bands result. **Figure 7** and **Figure 8** showed the result of FFT + SVM successfully classified accordingly the respective bands which are Very low (0 – 1.24 Hz), Delta (1.25– 4.5 Hz), theta (4.75

- 6.75 Hz), alpha-1 (7 – 9.5 Hz), alpha-2 (9.75 – 12.5), beta-1 (12.75 – 18.5 Hz), beta-2 (18.75 – 35 Hz) and Gamma (35.25 – 45 Hz). To compute the graph below, we calculated the mean value of each classified band. Thus, the value showed in mV^2 .

The result in the training dataset shows the Frontal Cortex gives a higher mean for Delta, while in the test dataset, Alpha-1 gave a higher mean. The Parietal Cortex and Right Hippocampus regions for both datasets show Delta band shows the highest mean value and the Gamma band is the lowest. Both showed almost the same trend. However, the trend changed in the Left Hippocampus region for both datasets. The top highest for the training dataset is Delta, Theta, and Beta-2. The top highest is Alpha-1, Theta, and Delta in the test dataset. The graph trend is expected to be different due to the random split data we set before implementing the classification step.

3.3 Result of FFT + ANN

The result of the EEG dataset applied with the FFT + ANN classification is shown in **Figure 9**. ANN consists of some neurons in hidden layers, which in this case, we use a value of eight because the purpose of classification must be in eight different bands. The black lines show the connections with weights. As mentioned previously, the ANN net is essentially a black box, so there is not much explanation on the fitting, the weights, and the model ([Olden & Jackson, 2002](#)).

The results showed 49.57%, which is almost 50%. Based on this, we conclude that a network of linear neurons did not exhibit significant performance with an accuracy of less than 50% for both datasets. The reason for combining different classification methods is to investigate its effect on the EEG dataset and at the same time automate the eight bands classification. **Table 2** is the result that summarises all three classification methods that have been applied.

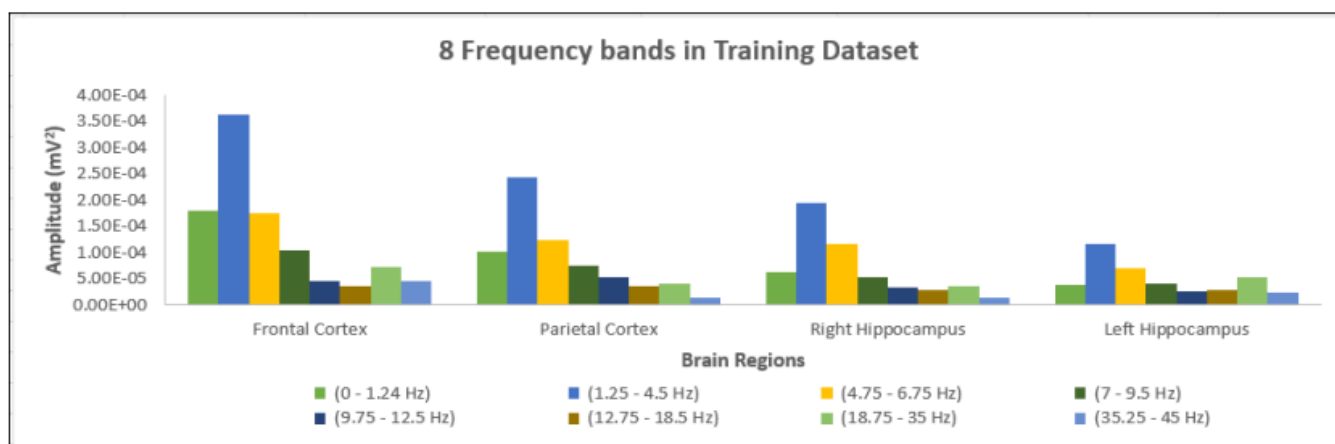


Figure 7. Result of different frequency bands using SVM in the training dataset.

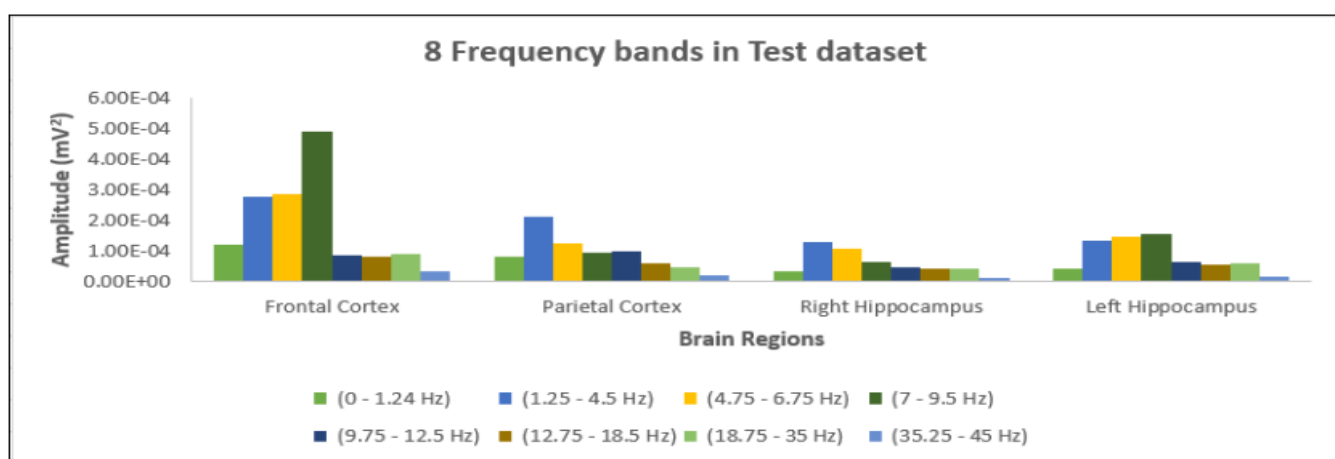


Figure 8. Result of different frequency bands using SVM in the test dataset.

Table 2. Summary of results.

Classification Technique	Accuracy Results
FFT + KNN	29%
FFT + SVM	100%
FFT + ANN	49.57%

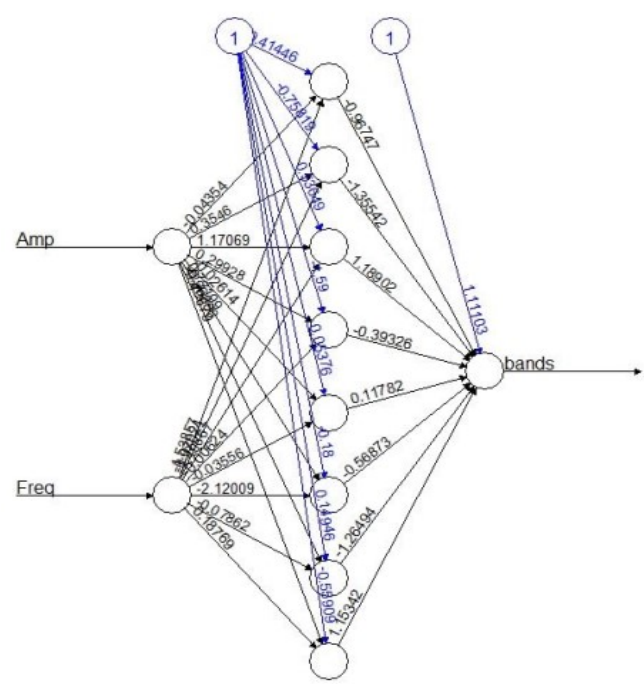


Figure 9. The ANN's results.

The main highlight in this experiment is the result achieved by the FFT+SVM method. This method successfully classifies the bands into eight bands with an accuracy of 100%. This is because SVM is regarded as a useful tool for effectively complementing the information gained. According to Auria and Moro (2011), SVM can produce accurate and robust classification results and conveniently evaluate relevant information. Hence, we can conclude that the problem of the article was successfully achieved by using the FFT+SVM method. From this result, we can help CDR USM analyse their EEG dataset in a shorter time by using the script we used in this experiment.

References

Al-Fahoum, A. S., & Al-Fraihat, A. A. (2014). Methods of eeg signal features extraction using linear analysis in frequency and time-frequency domains. *International Scholarly Research Notices* 2014, 1–7. <https://doi.org/10.1155/2014/730218>

Al-Janabi, O. M., Malim, N. H. A. H., & Cheah, Y. N. (2018). Analogy-based common-sense knowledge for opinion-target identification and aggregation. *2018 2nd International Conference on Imaging, Signal Processing and Communication, ICISPC 2018*, 1–5. <https://doi.org/10.1109/ICISPC44900.2018.9006687>

Al-Janabi, O. M., Malim, N. H. A. H., & Cheah, Y. N. (2020). Aspect categorisation using domain-trained word embedding and

The approximate time they can complete the experiment is reduced to one month because they are still required to collect the EEG data from the rats for the drug habituation experiment but classifying the bands into eight only takes a few minutes to complete.

4.0 CONCLUSIONS

Brain studies have helped in many experiments research fields such as medical, business, entertainment, education, and technology. This is why for centuries, scientists and philosophers have been fascinated by the brain, but until recently, they viewed the brain as nearly incomprehensible. As CDR USM takes approximately two months to analyse the EEG data, it is impractical and time-consuming. Therefore, we present translational research in which we solve the problem faced by CDR USM by proposing a framework to automate the process of classifying the eight bands that consists of two main objectives that we have successfully achieved. The first objective is to hybrid the FFT and classification method. From the experiment that has been done, method FFT+SVM generated accuracy with 100% and successfully classified the EEG signals into eight bands. Besides that, the results of different accuracy reading from different classification methods proved an effect in different classification methods. Hence, from approximately two months of the process manually being done before CDR USM, the time is reduced to one month. CDR USM can use the script we implemented in this experiment to classify the bands accordingly. The script only takes a few minutes to complete. Moreover, the difference t this research portrayed compared to other past work is that this paper automatically classified the bands from a raw EEG signal. Other past works classify only certain labels they need to produce, and the data has been labelled accordingly. Lastly, the future work that can be done is to process more EEG datasets and experiment with more rats instead of one and observe any special features.

Conflicts of Interest: The author declares no conflict of interest.

- topic modelling. *Lecture Notes in Electrical Engineering*, 619, 191–198. https://doi.org/10.1007/978-981-15-1289-6_18
- Almustafa, K. M. (2020). Prediction of heart disease and classifiers' sensitivity analysis. *BMC Bioinformatics*, 21(1), 1–18. <https://doi.org/10.1186/s12859-020-03626-y>
- Andre, A. B., Beltrame, E., & Wainer, J. (2013). A combination of support vector machine and k-nearest neighbors for machine fault detection. *Applied Artificial Intelligence*, 27(1), 36–49. <https://doi.org/10.1080/08839514.2013.747370>
- Anusha, K. S., Mathews, M. T., & Puthankattil, S. D. (2012). Classification of normal and epileptic EEG signal using time amp; Frequency domain features through artificial neural network. *Proceedings - 2012 International Conference on Advances in Computing and Communications, ICACC 2012*, 98–101. <https://doi.org/10.1109/ICACC.2012.21>
- Anwar, H., Khan, Q. U., Nadeem, N., Pervaiz, I., Ali, M., & Cheema, F. F. (2020). Epileptic seizures. *Discoveries Journal*, 8(2), 1–19. <https://doi.org/10.15190/d.2020.7>
- Asadur Rahman, M., Foisal Hossain, M., Hossain, M., & Ahmmed, R. (2020). Employing PCA and t-statistical approach for feature extraction and classification of emotion from multichannel EEG signal. *Egyptian Informatics Journal*, 21(1), 23–35. <https://doi.org/10.1016/j.eij.2019.10.002>
- Auria, L., & Moro, R. A. (2011). Support Vector Machines (SVM) as a Technique for Solvency Analysis. *DIW Berlin Discussion Paper No. 811*. <https://doi.org/10.2139/ssrn.1424949>
- Balaji, A., Haldar, A., Patil, K., Ruthvik, T. S., Ca, V., Jartarkar, M., & Baths, V. (2017). EEG-based classification of bilingual unspoken speech using ANN. *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* 1022–1025. <https://doi.org/10.1109/EMBC.2017.8037000>
- Barranco-Gutiérrez, A. I. (2020). Machine learning for brain images classification of two language speakers. *Computational Intelligence and Neuroscience*, 2020. <https://doi.org/10.1155/2020/9045456>
- Bastos, N. S., Marques, B. P., Adamatti, D. F., & Billa, C. Z. (2020). Analysing EEG signals using decision trees: a study of modulation of amplitude. *Computational Intelligence and Neuroscience*, 2020. 1-11. <https://doi.org/10.1155/2020/3598416>
- Berbić, J., Ocvirk, E., Carević, D., & Lončar, G. (2017). Application of neural networks and support vector machine for significant wave height prediction. *Oceanologia*, 59(3), 331–349. <https://doi.org/10.1016/j.oceano.2017.03.007>
- Chan, A., Early, C. E., Subedi, S., Li, Y., & Lin, H. (2015). Systematic analysis of machine learning algorithms on EEG data for brain state intelligence. *Proceedings - 2015 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2015*, 793–799. <https://doi.org/10.1109/BIBM.2015.7359788>
- Chao, H., Dong, L., Liu, Y., & Lu, B. (2020). Improved deep feature learning by synchronisation measurements for multi-channel eeg emotion recognition. *Complexity*, 2020, 1-15. <https://doi.org/10.1155/2020/6816502>
- Chriskos, P., Kaitalidou, D. S., Karakasis, G., Frantzidis, C., Gkivogkli, P. T., Bamidis, P., & Kourtidou-Papadeli, C. (2017). Automatic sleep stage classification applying machine learning algorithms on eeg recordings. *Proceedings - IEEE Symposium on Computer-Based Medical Systems, 2017-June*, 435–439. <https://doi.org/10.1109/CBMS.2017.83>
- Chu, N. N. (2017). IEEE and CTA's work toward neurotech standards for EEG data. *IEEE Consumer Electronics Magazine*, 6(4), 137–139. <https://doi.org/10.1109/MCE.2017.2714298>
- Dhanapal, R. and Bhanu, D. (2020). Electroencephalogram classification using various artificial neural networks, *Journal of Critical Reviews*, 7(4), 891–894. <https://doi.org/10.31838/jcr.07.04.170>
- Duffy, F. H. (1981). Brain electrical activity mapping (BEAM): Computerised access to complex brain function. *International Journal of Neuroscience*, 13(1), 55–65. <https://doi.org/10.3109/00207458108991795>
- Gárate-Escamila, A. K., El Hassani, A.H., & Andrés, E. (2020). Classification models for heart disease prediction using feature selection and PCA. *Informatics in Medicine Unlocked*, 19, 100330. <https://doi.org/10.1016/j.imu.2020.100330>
- Giannakaki, K., Giannakakis, G., Farmaki, C., & Sakkalis, V. (2017). Emotional state recognition using advanced machine learning techniques on EEG data. *Proceedings - IEEE Symposium on Computer-Based Medical Systems, 2017-June(July)*, 337–342. <https://doi.org/10.1109/CBMS.2017.156>
- Heideman, M. T., Johnson, D. H., & Burrus, C. S. (1984). Gauss and the history of the fast fourier transform. *IEEE ASSP Magazine*, 1(4), 14–21. <https://doi.org/10.1109/MASSP.1984.1162257>
- Imandoust, S. B., & Bolandraftar, M. (2013). Application of K-Nearest Neighbor (KNN) approach for predicting economic events : theoretical background. *International Journal of Engineering Research and Applications*, 3(5), 605–610.
- Jabbar, M. A., Deekshatulu, B. L., & Chandra, P. (2013). Classification of heart disease using K- nearest neighbor and genetic algorithm. *Procedia Technology*, 10, 85–94. <https://doi.org/10.1016/j.protcy.2013.12.340>
- Javeed, A., Rizvi, S. S., Zhou, S., Riaz, R., Khan, S. U., & Kwon, S. J. (2020). Heart risk failure prediction using a novel feature selection method for feature refinement and neural network for classification. *Mobile Information Systems*, 2020, 1-11. <https://doi.org/10.1155/2020/8843115>
- Kim, D. K., Jeong, K. S., McKay, R. I. B., Chon, T. S., & Joo, G. J. (2012). Machine learning for predictive management: Short and long term prediction of phytoplankton biomass using genetic algorithm based recurrent neural networks. *International Journal of Environmental Research*, 6(1), 95–108. <https://doi.org/10.22059/ijer.2011.476>
- Kondo, S., Najm, I., Kunieda, T., Perryman, S., Yacubova, K., & Lüders, H. O. (2001). Electroencephalographic characterisation of an adult rat model of radiation-induced cortical dysplasia. *Epilepsia*, 42(10), 1221–1227.

- <https://doi.org/10.1046/j.1528-1157.2001.38300.x>
- Kortelainen, J., Väyrynen, E., Jia, X., Seppänen, T., & Thakor, N. (2012). EEG-based detection of awakening from isoflurane anesthesia in rats. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, August*, 4279–4282. <https://doi.org/10.1109/EMBC.2012.6346912>
- Kumar, G., & Bhatia, P. K. (2014). A detailed review of feature extraction in image processing systems. *International Conference on Advanced Computing and Communication Technologies, ACCT*, 5–12. <https://doi.org/10.1109/ACCT.2014.74>
- Kumari, C. K., Premsai, K., Gowri, M. S., Kumar, K. S., & Murthy, K. N. (2019). Human Emotions Classification using EEG Signals in Virtual Instrumentation Platform. *International Journal of Recent Technology and Engineering (IJRTE)*, November, 8(4), 12104–12107. <https://doi.org/10.35940/ijrte.D8598.118419>
- Lan, Z., Sourina, O., Wang, L., Scherer, R., & Müller-Putz, G. (2017). Unsupervised feature learning for EEG-based emotion recognition. *Proceedings - 2017 International Conference on Cyberworlds, CW 2017 - in Cooperation with: Eurographics Association International Federation for Information Processing ACM SIGGRAPH, 2017-January*, 182–185. <https://doi.org/10.1109/CW.2017.19>
- Lemm, S., Blankertz, B., Dickhaus, T., & Müller, K. R. (2011). Introduction to machine learning for brain imaging. *NeuroImage*, 56(2), 387–399. <https://doi.org/10.1016/j.neuroimage.2010.11.004>
- Li, M., & Liu, K. (2020). Probabilistic prediction of significant wave height using dynamic Bayesian network and information flow. *Water*, 12(8), 2075. <https://doi.org/10.3390/W12082075>
- Maksimenko, V. A., Kurkin, S. A., Pitsik, E. N., Musatov, V. Y., Runnova, A. E., Efremova, T. Y., Hramov, A. E., & Pisarchik, A. N. (2018). Artificial neural network classification of motor-related EEG: An increase in classification accuracy by reducing signal complexity. *Complexity*, 2018, 10. <https://doi.org/10.1155/2018/9385947>
- Masruroh, A. H., Imah, E. M., & Rahmawati, E. (2019). Classification of emotional state based on EEG signal using AMGLVQ. *Procedia Computer Science*, 157, 552–559. <https://doi.org/10.1016/j.procs.2019.09.013>
- Michel, C. M., Vuilleumier, P., & Van De Ville, D. (2019). Current opinions in brain imaging methods and applications. *Brain Topography*, 32(6), 923–925. <https://doi.org/10.1007/s10548-019-00747-3>
- Mohammed, M., Khan, M. B., & Bashie, E. B. M. (2016). *Machine Learning: Algorithms and Applications* (1st ed.). CRC Press. <https://doi.org/10.1201/9781315371658>
- Murugappan, M., Murugappan, S., Balaganapathy, B., & Gerard, C. (2014). Wireless EEG signals based Neuromarketing system using Fast Fourier Transform (FFT). *Proceedings - 2014 IEEE 10th International Colloquium on Signal Processing and Its Applications, CSPA 2014*, 25–30. <https://doi.org/10.1109/CSPA.2014.6805714>
- Murugappan, Murugappan, Murugappan, S., & Zheng, B. S. (2013). Frequency band analysis of electrocardiogram (ECG) signals for human emotional state classification using discrete wavelet transform (DWT). *Journal of Physical Therapy Science*, 25(7), 753–759. <https://doi.org/10.1589/jpts.25.753>
- Mustafa, M., & Lokman, N. A. A. (2014). Classification of Spike-wave Discharge with STFT Approach. *Journal of Engineering and Technology (JET)*, 5(1), 11–20.
- Nguyen, V. T. N., Huynh, V. T., & Nguyen, T. H. H. (2019). Electroencephalography analysis using neural network. *NICS 2018 - Proceedings of 2018 5th NAFOSTED Conference on Information and Computer Science*, 144–147. <https://doi.org/10.1109/NICS.2018.8606797>
- Novakovic, J., Veljovi, A., Ilic, S., Papic, Z., & Tomovic, M. (2017). Evaluation of classification models in machine learning. *Theory and Applications of Mathematics & Computer Science*, 7(1), 39–46. Retrieved from <https://www.uav.ro/applications/se/journal/index.php/TAMCS/article/view/158>
- Olden, J. D., & Jackson, D. A. (2002). Illuminating the "black box": Understanding variable contributions in artificial neural networks. *Ecological Modelling*, 154, 135–150. [https://doi.org/10.1016/S0304-3800\(02\)00064-9](https://doi.org/10.1016/S0304-3800(02)00064-9)
- Petrovska, B., Zdravevski, E., Lameski, P., Corizzo, R., Štajduhar, I., & Lerga, J. (2020). Deep learning for feature extraction in remote sensing: A case-study of aerial scene classification. *Sensors*, 20(14), 1–22. <https://doi.org/10.3390/s20143906>
- Sarmast, S. T., Abdullahi, A. M., & Jahan, N. (2020). Current classification of seizures and epilepsies: scope, limitations and recommendations for future action. *Cureus*, 12(9), e10549. <https://doi.org/10.7759/cureus.10549>
- Selvaraj, J., Murugappan, M., Wan, K., & Yaacob, S. (2013). Classification of emotional states from electrocardiogram signals: A non-linear approach based on hurst. *BioMedical Engineering Online*, 12(1), 44. <https://doi.org/10.1186/1475-925X-12-44>
- Shu, L., Yu, Y., Chen, W., Hua, H., Li, Q., Jin, J., & Xu, X. (2020). Wearable emotion recognition using heart rate data from a smart bracelet. *Sensors (Switzerland)*, 20(3), 1–19. <https://doi.org/10.3390/s20030718>
- Si, Y. (2020). Machine learning applications for electroencephalograph signals in epilepsy: a quick review. *Acta Epileptologica*, 2(5), 5–11. <https://doi.org/10.1186/s42494-020-00014-0>
- Siuly, S., Li, Y., & Zhang, Y. (2016). Significance of EEG signals in medical and health research. In: *EEG Signal Analysis and Classification* (pp. 23–41). Springer. https://doi.org/10.1007/978-3-319-47653-7_2
- Tzallas, A. T., Giannakeas, N., Zoulis, K. N., Tsipouras, M. G., Glavas, E., Tzamourta, K. D., Astrakas, L. G., & Konitsiotis, S. (2017). EEG classification and short-term epilepsy prognosis using brain computer interface software. *Proceedings - IEEE*

- Symposium on Computer-Based Medical Systems, 2017-June*, 349–353. <https://doi.org/10.1109/CBMS.2017.97>
- Xie, L. (2012). *KNN Classification and Regression using SAS*.
- Xiong, Q., Zhang, X., Wang, W. F., & Gu, Y. (2020). A parallel algorithm framework for feature extraction of EEG signals on MPI. *Computational and Mathematical Methods in Medicine*, 2020, 9812019. <https://doi.org/10.1155/2020/9812019>