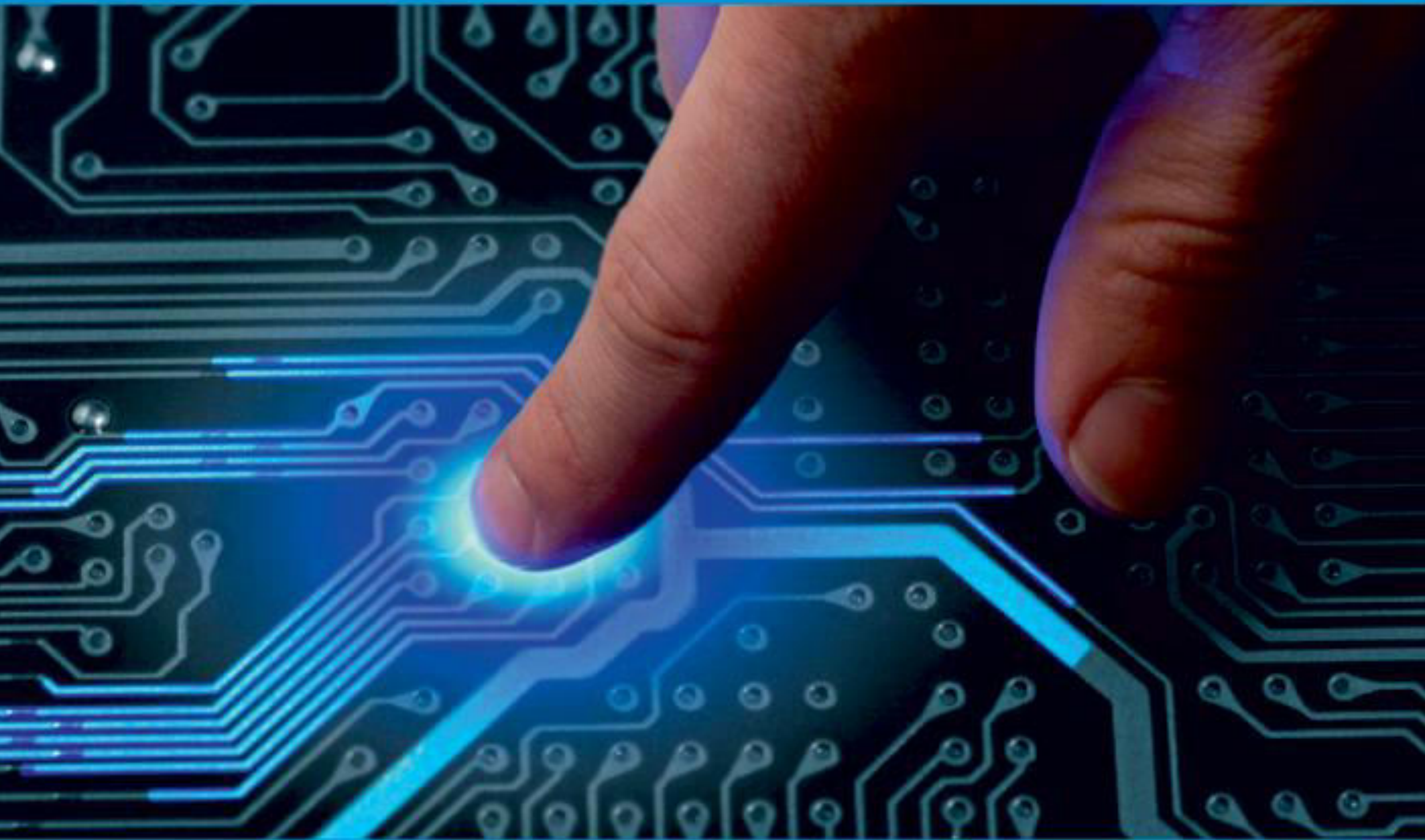
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 7, July 2022

Impact Factor: 8.165

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007097 |

Electroencephalogram Signal Classification using Machine Learning Algorithm

Sivakumar R, Anirudh E R, Sai Viswanth Reddy Y

Department of Sensor and Biomedical Technology, School of Electronics Engineering

Vellore Institute of Technology, Vellore, India.

ABSTRACT - Recent Studies and research advancements on brain wave signals have helped us to detect various brain disorders at early stages even before it starts to show symptoms. Many brain diseases get unidentified by MRI or CT images since they produce the image of the brain which does not include the functionality of the brain and related abnormalities could not be identified precisely. The EEG signals could directly interpret the functions of the brain with ease and with lesser complications. Although the results from these EEG signals are very promising, the classification of the EEG signals depends on many associated factors. The work includes breaking down the EEG signal into fixed samples and calculate features from the signal. We classify the EEG signal into two major Gait classes and using Machine learning algorithm we predict the corresponding gait pattern.The results concluded that with multiple features we were able to conserve more information of an EEG signal. This helped in obtaining model with more accuracy. Our results help us to conclude that the methodology which we used for the gait pattern identification and classification is comparatively found to be accurate with the salient sequential process for classification and identification process from the pre existing works on similar fields.

KEYWORDS - Electroencephalogram(EEG), Support Vector Machine(SVM), Independent Component Analysis(ICA).

I. INTRODUCTION

In recent years, it has been a trend to design and develop systems that has been one of the most fascinating research interests in neuroscience, engineering and signal processing [1]. These systems have claimed to make striking advances in usability and information transfer in which the machine learning and signal processing techniques have acquired great achievements in recent years[2]-[7].Current works available can identify the abnormalities and diseased conditions according to the electroencephalogram (EEG) by innovative methods of feature extraction and machine learning classification algorithms[4]. This proposed work can also be used to design automated assistive healthcare devices for monitoring patients with abnormal walking patterns or early fall detections. It is necessary to collect the effective signals to identify the current state of operation. It is very difficult to collect the EMG from the patients with amyotrophic lateral sclerosis or strike so we collect these EEG signals instead for better accuracy and to finish the classification and analysis tasks. In this paper, we discuss a novel machine learning approach for classification of EEG signals which corresponds to two major gait patterns namely normal walk and medio lateral movement which indicates any lateral shift in the subject. The investigation of the gait patterns and associated EEG signals has also been collected. The acquired EEG signal has artifacts and noise associated with the signal and thus thereis a requirement to do pre-processing. Here we do the pre-processing with the help of the MATLAB. The proposed work has included some of the pre processing techniques such as computing averaging of the signal
, filtering the signal using Butterworth filters , then proceeding with the ICA(Independent component analysis).The ICA method has proven to be efficient for the removal of motion artifacts .After preprocessing we extract features from the EEG signal which is free from noise and artifacts.The features carries the information which are suffice the requirement to train a model and to identify and classify the machine learning model.This extracted features are used to train Machine learning model.Here we have used the support vector machine(SVM) and further we performed hyperparameter tuning for enhancing the performance of the model.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007097 |

II. LITERATURE SURVEY

An automated fuzzy support vector machine is being implemented for predicting occurrence of a seizure in an EEG signal[1]. The proposed SVM is different from an conventional SVM as it takes consideration of all the instances for drawing boundary between the classes. With this model the highest precision achieved is around 80% as no feature extraction is present. Similarly the EEG signal using Discrete Wavelet Transform(DWT) is split into 5 sub-signals[2] and from these sub-signal they have extracted temporal features such as Mean AbsoluteValue, Standard Deviation and Average Power.

A KNN machine learning model is trained with the help of these features to predict the epilepsy presence inthe EEG signal. The average results obtained are about 94% with KNN classifier.The unique element of thiswork[3] is that it has used slantlet transform and spare coding to greatly reduce the instances of false alarmsand also to increase the speed of detection. With the help of slantlet transform , all the salient information has been conserved and mapped to a sparse space**.** The performance in this paper is greatly affected by the window length and the machine learning classifier. The highest performance is achieved with window length equal to1 sec and using sparse representation classifier.

There are works existing for classifying the epileptic seizure using EEG signals; they have tremendously implemented nonlinear approaches[4]. Before feature extraction removed all the artifacts in the EEG signal. Most of the work they have focussed on obtaining a matrix of multi feature dimension including brownian motion , mean absolute value and Root mean square. Mainly used three machine learning models for classification out of which the ensemble model outperformed others by achieving 92% accuracy. The most novel strategy on achieving greater accuracy for classification of Epilepsy seizure in EEG signals using a single channel electrode(FT10-T8)[5].The whole classification process comprises mainly of three steps - preprocessing , feature extraction from wavelet transform and classification using Extreme Learning Machine. This method has obtained a classification accuracy around 95 percent.

The time-frequency analytical algorithm called local mean decomposition(LMD) , LMD is used to decompose the EEG signal into product functions[6]. Then five classifiers are used out of which KNN and SVM showed promising results were another set of recent studies on this domain.Recent studies of rhythmic activity of the EEG signals has found that in lower frequency bands such as delta , alpha and beta sub- bands[7]. The proposed methodology here includes implementation in the different datasets - CHB MIT scalpEEG database and TUH Seizure Corpus. The results were quite amusing as this technique achieved an average sensitivity in the range of 82 percentage to 92 percentage.

Another novel method to classify epileptic seizures and normal EEG data is to utilize the Intrinsic Time-scale Decomposition (ITD)-based features[8]. After decomposing the EEG signal into Proper Rotation Components(PRC) then feature extraction methods are applied to the first five PRCs. With the help of these features five different machine learning models are trained upon - KNN, Linear Discriminant Analysis , Naive Bayes , Support Vector Machine and Logistic Regression classifiers. KNN classifier predicted with the highest accuracy among these 5 models with a score of 95%. For the feature extraction from the EEG signal wavelet transform have been used wherein for the Machine learning model Support Vector Machine(SVM) is being trained[9] Instead of focussing on the whole EEG signal the proposed method concentrated on extracting features from only alpha sub-band. The classification accuracy attained with the SVM classifier is almost 95%.The epileptic seizure activities of the EEG signal have been detected with the advanced technology of deep learning[10].

The method proposed for the one-dimensional EEG signal classification for seizure detection involves the simplest 1-D Convolutional Neural Network(CNN). A unique technique based on Singular Value decomposition for the classification of Seizure Epileptic EEG signals[11]. The SVD is applied sequentially on a sliding window of one second width of EEG data and the r singular values are obtained and used to indicate sudden changes in the signals. Through observations the singular values deviated notably upward from the baseline at the start of epileptic seizure. This shows that the singular values are sensitive to the amplitude changes in the EEG signal and therefore it is a reliable method for detection of an epileptic seizure.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165
|| Volume 10, Issue 7, July 2022 ||
| DOI: 10.15680/IJIRCCE.2022.1007097 |

In this study ,the EEG signal mostly lies between 0.5 Hz - 70 Hz , as muscle and motion artifacts are outside this range, the most important EEG sub-bands rhythmic signals present in this range. Changes in the frequency distribution and in the time domain will help in classifying the gait classes. A powerful machine learning model is used to segment the two classes with higher accuracy.

III. METHODOLOGY

The Raw EEG data extracted from the subject is noted to have lot of noise and motion related artifacts and undergoes the filtering and artifact removal process.The filtered EEG is then used for the feature extraction and for the identification and classification process of our proposed methodology.

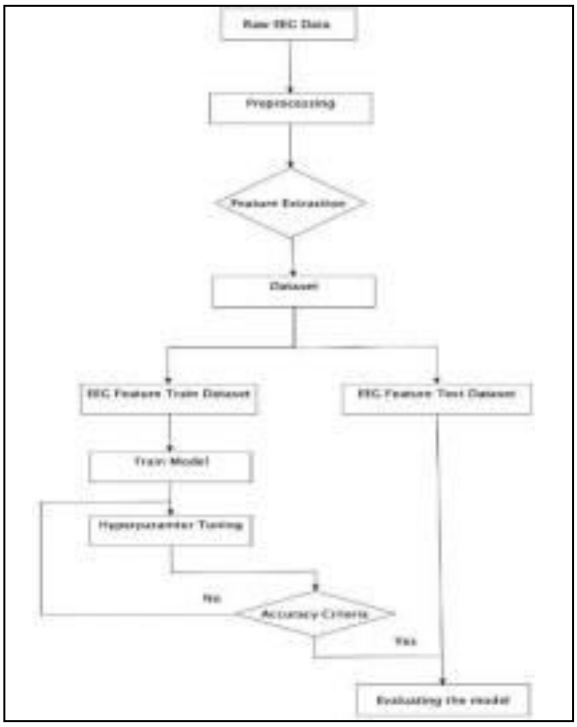


Fig. 1. The figure illustrates the process followed at a glance.

A. Pre-processing

The Raw EEG signal consists of a lot of noise and artifacts hence it is mandatory to remove these unwanted signals from the original signal.In Order to make the signal noise free we have implemented some of the below described techniques in sequential manner.An average EEG signal ranges from 0.5Hz to 80 Hz.Thus we remove the frequency components which lies outside this frequency range.The type of filter used here is a Butterworth filter and it is implemented in using MATLAB mainly for processing of these EEG signals. A band stop filter is applied to remove 60 Hz power line interference noise present in the EEG signal. The bandstop filters are specialized in having a very narrow band which rejects the frequencies.

After the noise removal using filtering we focus on removing the artifacts using the ICA process.Independent Component Analysis (ICA) is used here to identify the hidden factors in a signal.ICA analysis helps us to identify the artifacts present in the signal. Environmental disturbances and even the blink of an eye can be easily separated using ICA. The main purpose of ICA is for a random vector that includes searching for a linear transformation which minimizes the statistical dependence between the components involved in the signal.ICA analysis helped us to identify the different types of artifact present in each channel electrode and also muscle artifacts was found to be more in our EEG signal. The artifact probability was calculated for each channel with the help of a neural

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165
|| Volume 10, Issue 7, July 2022 ||
| DOI: 10.15680/IJIRCCE.2022.1007097 |

network.The channel electrodes which had the artifact removed from the EEG signal thus made the EEG signal free from noise.

B. Feature Extraction

EEG signals are in pure analog form and would not get useful information for further processing so in order to get significant oversights of these obtained EEG signals we proceed with the feature extraction process. For all the feature extraction processes the obtained EEG signal is divided into small window sizes to make the windowed signal stationary in the considered window region.The features used here are statistical features like variance,kurtosis and skewness Hjorth parameters(activity,mobility,complexity),Root Mean square and Mean absolute Value.

The segment EEG signal depicts the deviation from the mean. If the variance is more, then the data is spread out.The variance of the EEG signal is calculated by taking the differences between each number in the data set and the mean, then squaring the differences to make them positive, and finally dividing the sum of the squares by the number of values in the data set.Indicates the measure of the combined weight of our feature data distribution's tails relative to its centre of distribution. When the normal feature data is plotted via a histogram, it shows a bell peak and most of these feature data within three standard deviations (plus or minus) of the mean.Hjorth parameters are indicators of statistical properties used to conserve spatial information of EEG signal. The parameters which are included here in our considered feature dataset are namely activity , mobility and complexity.Activity parameter depicts the signal power that is the variance of a time function. This can indicate the surface of the power spectrum in the frequency domain. Mobility parameter represents the proportion of standard deviation of the power spectrum.Complexity parameter and compares the signal's similarity to a pure sine wave, where the value results to 1 if the signal is more similar. It thus represents the alterations in the frequency in the EEG signal samples.RMS velocity is the value of the square root of the sum of the squares of the stacking velocity values divided by the number of values.RMS is a time-domain feature, which is widely used for analyzing EEG Signals. Mean Absolute Value of a data set is the average distance between each feature dataset value and its mean value. Thus it is a way to describe variation in a data set. Mean Absolute Value is defined as mean absolute value of EEG Samples taken over a given period of time. After various temporal and spectral features are extracted from the processed EEG signal , a machine learning classifier is used to differentiate between the two classes - Walk , lateral perturbations on the basis of the features. Support Vector Machine Model is trained on the extracted feature dataset which contains features such as Mean Absolute Value
, Root mean Square Entropy. The performance for this model has been evaluated based on the f1-score,accuracy and roc curves. To utilize the hyperparameters of the support vector machine efficiently we have done the hyperparameter tuning.

C. Classification

Support Vector Machine (SVM) is a supervised machine learning algorithm that can be used for both classification or regression challenges. However, it is mostly used in classification problems. . In the SVM algorithm the data points are plotted in n-dimensional spacing depending on the number of features. The classification of the data points depends on the support vectors for each target class. Based on these support vectors a plane is drawn to segment these target classes. This plane is called a hyperplane as it is segmenting the classes in n-dimension. It requires a lot of dataset and more spatial features to train the model as it helps the model to learn correctly and fit into the dataset. A similarity function has been set in the SVM classifier;this similarity function tackles nonlinear classification problems by adding features to resemble a particular landmark.

This similarity function is set to Gaussian Radial Basis function(RBF) , but it is very computationally expensive.The kernel trick has been used in the svm to reduce the time to add additional features; its basic principle is to transform non-linear data and project into a higher dimensional space where the classes could be linearly separable. F1-score, Accuracy and ROC curves are calculated as performance parameters for the SVM classifier.For improving the performance of our models we use hyper parameter tuning which improves the accuracy as well as the speed of the model.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165
|| Volume 10, Issue 7, July 2022 ||
| DOI: 10.15680/IJIRCCE.2022.1007097 |

D.HyperParameter Tuning

The hyper-parameters are crucial in deciding the adaptiveness of the model.These parameter values influence the learning process directly and help in determining the values of model parameters to attain high performance. Optimization of hyper paramters is an exhausting process as it takes time to evaluate the model on each combination of hyper-parameters values. To overcome this , we have used RandomSearchCv Optimization

technique to narrow down the parameters which is best suited to the model.The result of a hyperparameter optimization is a single set of well performing hyperparameters that we can use to configure the model. RandomizedSearchCV Optimization algorithm tries to evaluate a model in a given number of random combinations by selecting a random value for each hyperparameter at every iteration. This algorithm is quite efficient when the hyperparameter search space is too large. We have chosen a few hyperparameters to tune which affects the model in a positive direction.

Hyperparameter	**Value**
kernel	rbf
degree	2
C	10

TABLE I. Hyperparameter values for the SVM model

Kernel - it transforms the data of a nonlinear system which is impeccable to separate to a linear decision boundary of higher order dimension. By using the kernel , the model is able to train the transformed data much faster.

Degree - it is defined as the degree of the polynomial equation which signifies the complexity of the model ,if the degree is higher that the model complexity is more.

C - a regularization term which regulates the model to correctly fit the data.

V. RESULTS

The spectral and temporal features extracted from the EEG signal are highly significant in affecting the model performance.Few features which seem to conserve more information of EEG signal and provide good accuracy are Entropies, Statistical Features , Root Mean Square(RMS) and Mean Absolute Value(MAV).The performance metrics used to evaluate the model are precision,recall,f1 score,accuracy and ROC curves. Continuous improvement in the model is based on these values to know if the model is correctly fitting the EEG dataset.

Gait type	**Exclusion of cross validation**			**Inclusion of cross validation**		
	Precision	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-Score**
Medio lateral movement	0.85	0.66	0.74	0.83	0.63	0.72
Walk	0.77	0.91	0.83	0.75	0.89	0.82

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007097 |

TABLE II. performance metrics before hyperparameter tuning Train Dataset

Gait type	Exclusion of cross validation			Inclusion of cross validation		
	Precision	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-Score**
Medio lateral movement	0.82	0.64	0.72	0.78	0.63	0.70
Walk	0.75	0.88	0.81	0.73	0.85	0.79

TABLE III. performance metrics before hyperparameter tuning Test Dataset

From the above tables II and III , the performance scores are tabulated for the train and test data respectively. Both the dataset have been trained with and without K-fold cross validation technique. It is applied to overcome overfitting or underfitting of the model by validating the model in each iteration.This will reducecomplexity of the model. For the train dataset the classifier is predicting the walk gait class with high recall and f1 score of 91% and 83% respectively whereas the Medio lateral movement is predicted with great precision of about 85%. But for both the gait classes the f1- score is low where f1 score depends both on precision and recall. Now there is a need to increase the performance of the model in all metrics. TABLE
IV. performance metrics after hyperparameter tuning Train Dataset

Gait type	Exclusion of cross validation			Inclusion of cross validation		
	Precision	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-Score**
Medio lateral movement	0.88	0.88	0.88	0.82	0.84	0.83
Walk	0.90	0.90	0.90	0.87	0.85	0.86

TABLE IV. performance metrics after hyperparameter tuning Train Dataset

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165
|| Volume 10, Issue 7, July 2022 ||
| DOI: 10.15680/IJIRCCE.2022.1007097 |

Gait type	**Exclusion of cross validation**			**Inclusion of cross validation**		
	Precision	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-Score**
Medio lateral Movement	0.81	0.83	0.82	0.85	0.83	0.84
Walk	0.86	0.84	0.85	0.86	0.87	0.87

TABLE V. Performance metrics after hyperparameter tuning Test Dataset

The tables IV and V represent the scores of the model after implementing hyperparameter tuning. Various hyper parameters values were used for the SVM class and we have selected the best hyperparameter combination of sets to improve the model's predictions. The scores for each metric has been increased almostby 10% for Walk and Medio lateral movement class. The Walk target class is achieving an average of 90% and Medio lateral movement class attained almost 87%.

Data set	**Exclusion of cross validation**	**Inclusion of cross validation**
	Accuracy	**Accuracy**
TRAIN	0.80	0.78
TEST	0.77	0.75

TABLE VI. Accuracy for SVM before hyperparameter Tuning

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165
|| Volume 10, Issue 7, July 2022 ||
| DOI: 10.15680/IJIRCCE.2022.1007097 |

Data set	Exclusion of cross validation	Inclusion of cross validation
	Accuracy	**Accuracy**
TRAIN	0.89	0.85
TEST	0.85	0.83

TABLE VII. Accuracy for SVM after hyperparameter Tuning

The above tables have been tabulated the accuracy for the train and test dataset before and after hyperparameter tuning. Tremendous improvement in the accuracy has been attained with the concept of hyperparameter tuning. For the test dataset the accuracy is 85% with hyperparameter tuning and withoutit is around 77% so there is almost 8% increase.

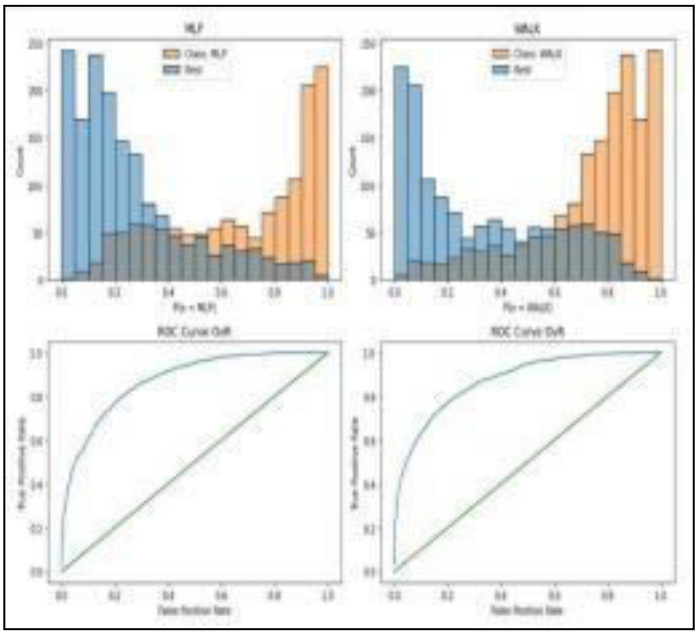


Fig. 2. ROC curves and histogram bins split for random forest classifier

In Fig 2 , the ROC curve plots are plotted for each class and also histogram is calculated keeping one class as positive and other as negative. ROC curve is the graph which measures the performance of the model by utilizing two parameters - True Positive Rate(TPR) and False Positive Rate. The model is performing well when the area under the roc curve is more or the roc curve is near to the top left corner of the plot.We used the One vs One strategy because here we have only two classes and calculated the ROC curve plots.This process is completed when each target class is assigned the positive class. The ROCcurve plots for the target classes are similar and it is towards the left which signifies the model is classifying the classes with almost equal and high accuracy.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007097 |

V. CONCLUSION

The results obtained concluded that EEG signals corresponding to the Walk and Medio lateral movement and the features extracted from these were sufficient and efficient in training a model and predicting and identifying the gait accurately. In this work, the gait class identification and its classification process is implemented using a support vector machine. The results obtained post hyperparameter tuning is observed to have

improved accuracy approximately 9% compared to model performance before hyperparameter tuning. The results of this work also suggest that Support vector machines with hyperparameter tuning have performed well in prediction and classification of the target class. Hyperparameter tuning techniques have further improved the performance of these models by enhancing the speed too. The features extracted from the EEG signal which was used to train the model were suffice and contained all the necessary information about the signal which includes its spatial and temporal aspects as well.This serves to add the novelty of our approach to the existing works on the similar domain.The proposed method can be extended to fields of developing assistive health care devices where in gait pattern identification and analysis is required for determining the progression of recovery especially in the case of physiotherapeutic application.

REFERENCES

[1] P. Suguna, B. Kirubagari and R. Umamaheswari, "An Effective Fuzzy Support Vector Machine based Seizure Detection Model using EEG Signals," 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2020, pp. 58-62.

[2] B. Harender and R. K. Sharma, "DWT based epileptic seizure detection from EEG signal using k-NN classifier," 2017 International Conference on Trends in Electronics and Informatics (ICEI), 2017, pp. 762-765.

[3] A. K. Ibrahim, H. Zhuang, N. Erdol and A. M. Ali, "EEG seizure detection by integrating slantlet transform with sparse coding," 2017 25th European Signal Processing Conference (EUSIPCO), 2017, pp. 459-462.

[4] T. Das, A. Ghosh, S. Guha and P. Basak, "Classification of EEG Signals for Prediction of Seizure using Multi-Feature Extraction," 2017 1st International Conference on Electronics, Materials Engineering and Nano-Technology (IEMENTech), 2017, pp. 1-4.

[5] S. Ammar and M. Senouci, "Seizure detection with single channel EEG using Extreme Learning Machine," 2016 17th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA), 2016, pp. 776-779.

[6] T. Zhang and W. Chen, "LMD Based Features for the Automatic Seizure Detection of EEG Signals Using SVM," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 25, no. 8, pp. 1100-1108.

[7] K. M. Tsiouris, S. Konitsiotis, D. D. Koutsouris and D. I. Fotiadis, "Unsupervised Seizure Detection basedon Rhythmical Activity and Spike Detection in EEG Signals," 2019 IE EMBS International Conference on Biomedical Health Informatics (BHI), 2019, pp, 1-4.

[8] M.DEGIRMENCI and A.AKAN, "EEG based Epileptic Seizures Detection using Intrinsic Time-Scale Decomposition," 2020 Medical Technologies Congress, 2020, pp.

[9] S.Kannadhasan and R,Nagarajan, Performance Improvement of H-Shaped Antenna With Zener Diode for Textile Applications, The Journal of the Textile Institute, Taylor & Francis Group, DOI: 10.1080/00405000.2021.1944523

[10] D.Selvathi and V.K.Meera, "Realization of epileptic seizure detection in EEG signal using wavelet transform and SVM classifier," 2017 International Conference on Signal Processing and Communication (ICPSC), 2017, pp.18-22.

[11] A.Shahid, N.Kamel, A.S. Malik and M.A. Jatoi "Epileptic seizure detection using the singular values of EEG signals," 2013 ICME International Conference on Complex Medical Engineering, 2013, pp 652-655