

Epileptic Seizure Prediction Through Machine Learning and Spatio-Temporal Features Based Time Series Analysis of Intracranial Electroencephalogram Data



N. Ilakiyaselvan, Hardik Bhatt, Utkarsh Shukla, A. Nayeemulla khan , A. Shahina

Abstract: Epilepsy is a group of neurological disorders identifiable by infrequent but recurrent seizures. Seizure prediction is widely recognized as a significant problem in the neuroscience domain. Developing a Brain-Computer Interface (BCI) for seizure prediction can provide an alert to the patient, providing a buffer time to get the necessary emergency medication or at least be able to call for help, thus improving the quality of life of the patients. A considerable number of clinical studies presented evidence of symptoms (patterns) before seizure episodes and thus, there is large research on seizure prediction, however, there is very little existing literature that illustrates the use of structured processes in machine learning for predicting seizures. Limited training data and class imbalance (EEG segments corresponding to preictal phase, the duration just before the seizure, to about an hour prior to the episode, are usually in a tiny minority) are a few challenges that need to be addressed when employing machine learning for this task. In this paper we present a comparative study of various machine learning approaches that can be used for classification of EEG signals into preictal and interictal (Interictal is the time between seizures) using the features extracted from the intracranial EEG. Publicly available data has been used for this purpose for both human and canine subjects. After data pre-processing and extensive feature extraction, different models are trained and are effectively used to analyze the temporal dynamics of the brain (interictal and preictal) in affected subjects. We present the improved results for various classification algorithms, with AUROC values of best classification models at 0.99.

Keywords : Epilepsy, Electroencephalogram, Seizure Prediction, Linear Classifier, Ensemble Classifier, Time series analysis.

I. INTRODUCTION

Epilepsy is a chronic neurological condition where abnormal electrical activity in the brain causes seizures.

Revised Manuscript Received on October 30, 2019.

* Correspondence Author

N. Ilakiyaselvan*, School of Computing Science and Engineering, Vellore Institute of Technology, Chennai, India.

Hardik Bhatt, School of Computing Science and Engineering, Vellore Institute of Technology, Chennai, India.

Utkarsh Shukla, School of Computing Science and Engineering, Vellore Institute of Technology, Chennai, India.

A. Nayeemulla Khan, School of Computing Science and Engineering, Vellore Institute of Technology, Chennai, India.

A. Shahina, Department of Information Technology, SSn College of Engineering, Kalavakkam, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](http://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Epileptic seizures are events that can vary from brief and nearly unnoticeable periods to long stretches of vigorous shaking, rhythmic muscle contractions or muscle spasms. Certain studies conclude that 1 in 26 people in the United States develop epilepsy at some point in their life. A cure to epileptic seizure does not exist for most of the patients but it can be controlled by anti-epileptic drugs. The seizure causes an array of problems, especially for patients who have developed drug-resistant epilepsy. The possibility of predicting seizures in advance could be very useful, not only for the patients but also for the medical professionals dealing with such disorders. In current practice, trained neurologists analyze signals from EEG and a synchronized video of the patients for localization and diagnosis. This task tends to be very tedious and slow, as it requires differentiating signals across multiple days. An automated system that accurately analyses patterns and classifies signal segments into different temporal dynamics of the brain would be extremely fast and useful. Due to extensive development in the field of EEG data collection and machine learning [1–4], and upcoming methods in signal acquisition and cleaning, and adequate performance evaluation for efficient seizure intervention systems there is a potential to completely automate the seizure prediction process [5–8]. In this study numerous strategies are employed to clean the raw iEEG data like binning and outlying frequency removal, perform feature extraction to quantize fluctuations or rhythmic behavior, variation or dispersion of data, self-similar unvaried pattern repetition, and energy distribution in a signal, and apply the following four broad categories of machine learning approaches, generalized Regression models, Support Vector Machines, Decision Trees and Tree Ensemble classifiers, to lay out a comparative analysis of the applicability of these approaches to predict the onset of a seizure episode under different constraints.

II. RELATED WORK

Automatic prediction of epileptic seizures and the classification of EEG signal as preictal or interictal have vastly improved primarily due to the development of EEG signal recording technology and exploiting newer machine learning algorithms. The analysis of EEG signals is performed for various research purposes like psychiatric studies, brain-machine interfaces, seizure classification, seizure prediction etc.



Epileptic Seizure Prediction Through Machine Learning and Spatio-Temporal Features Based Time Series Analysis of Intracranial Electroencephalogram Data

In this section, we provide a brief overview of some previous work done in the qualitative analysis of EEG signal data and the corresponding prediction and classification inferences drawn. Brinkmann et al.

[9] proposed the classification of interictal and preictal intracranial EEG data using Power in Band and inter-electrode synchrony intracranial EEG features for naturally occurring canine Epilepsy. They concluded that optimization of feature selection and best fitting algorithms are subject specific. Luigi et al. [10] proposed a technique for categorization of EEG signals in real time by means of Support Vector Machines. They also gave information about feature extraction that requires low computational power which facilitates the use of the algorithm in real time. Lin & Chen et al. [11] proposed the classification of EEG signals into preictal and interictal EEG epochs using artifact-free signals characterized by 216 global feature descriptors. A parameter selective computer-aided diagnosis system for identifying epileptic seizures was proposed by Sood et al. [12]. This technique performs the work by identifying the appropriate features from the data for identifying seizures. Tawfik et al. [13] proposed an epileptic seizure detection technique which uses Support Vector Machines. The idea proposed takes into consideration the fact that weighted permutation entropy-based measures for EEG signals when a person is in an ictal state are lower compared to when the person is in the interictal state. Wang & Lyu et al. [14] proposed a new approach where they use elimination based feature selection method to increase the efficiency of the existing algorithms and diminish the redundant points in the EEG signal. An evolutionary harmony search based algorithm for feature selection on EEG signals was presented by Zainuddin et al. [15] Focusing on the feature selection technique is the main contribution of the authors towards an EEG based seizure detection framework. Zhang et al. [16] proposed using Linear Support Vector Machine classifier for epileptic seizure prediction wherein spectral power and ratios of spectral power are extracted from intracranial EEG signals and processed by a second-order Kalman filter which is then fed as input to the SVM classifier.

Shafiul Alam et al. [17] proposed a method that uses higher order statistical moments of signals calculated in the empirical mode decomposition (EMD) domain for detecting epileptic seizures. Andriy Temko et al. [18] presents a methodology to effectively develop and combine approaches like Gaussian mixture models, Gaussian Super vector, Support Vector Machines and Hybrid Likelihood Ratio for efficient seizure prediction. Subha et al. [19] discusses different signal processing methods to extract the hidden information from the signal. Chen et al. [20] developed a method to decompose EEG data into seven commonly used wavelet families to the level of each mother wavelet and then wavelets and decomposition levels were searched in an exhaustive selection of frequency bands to provide optimal accuracy and low computational costs. Alotaiby et al. [21] devised a patient specific epileptic seizure prediction method relying on the common spatial pattern (CSP) based feature extraction of scalp electroencephalogram (sEEG) signals. Fei et al. [22] proposed a novel method to capture subtle chaotic dynamics for epileptic signals in fractional Fourier transform domain.

III. DATA SET DESCRIPTION

In 2014 the American Epilepsy Society provided data and hosted a competition on Kaggle.com, a platform for predictive modeling and analytics competitions, to get help from contributors around the world to develop algorithms that can compete with human expert encephalographers in terms of epileptic seizure detection. The data comprises of fairly long intracranial Electroencephalography (iEEG) signals, which are recordings of the brain activity quantified by measuring out the extracellular field potentials owing to the neuronal discharges. The data has iEEG recordings of seven different subjects, two human subjects, and five canines. The data has clips of two types, Preictal; A recording between 65 minutes to 5 minutes before the onset of a seizure, Interictal; the normal brain activity of a subject such that the segments are as far from any seizure as can be practically achieved to avoid contamination with preictal or postictal signals. The canine subjects have data gathered from 16 channels while the human subjects have 24 channels. The location of the implant placements may differ within different subjects.

IV. METHODOLOGY

Artifacts in EEG recordings are forms of outliers and are considered as disturbances in a regular brain-signal, not originating from the brain, which usually shows up in the signal as noisy frequency bands. Binning and outlying frequency removal is carried out. Any frequency below 0.5Hz and above 200 Hz is removed as they are assumed to give no significant information gain. The data is then windowed into 30-sec segments which results in 20 non-overlapping windows for every 10 min clip. After artifact removal, the features described in the following paragraph are then extracted for each segment for each of the subjects. Power spectral density and energy at specific frequencies are extracted as they may be used to identify seizures. Spectral entropy is a measure of the spectral power distribution of a signal treating it as a probability distribution in the frequency domain. The spectral entropy is lower when there is information in the signal. This aspect is used for feature extraction in biomedical signals as shown in [23]. We extract signal energy based features like entropies and spectral densities. Statistical quantitative features as moments, particularly skewness, kurtosis, are a measure of the shape of the distribution of a set of points, while standard deviation quantifies the amount of variation or dispersion of data points of the signal. These statistics are applied as features. Further measures like Hjorth parameters which indicate statistical properties used in signal processing in the time domain are also used. The parameters are Activity, which indicates the surface of the power spectrum in the frequency domain, Mobility, which is the proportion of standard deviation of the power spectrum, and Complexity, which represents a measure of similarity between the signal and a pure sine wave are also computed. Fractal Dimension (FD) estimates are obtained from the segment to capture self-similar unvaried repetitive patterns in the EEG signal.

Fractal dimension is shown to characterize the nonlinear behavior and state of many chaotic systems. Fractal dimensions for each segment are computed and used as features. The Haar wavelet is used to derive DWT for each segment. Additional features from spectral bands which measure the signal energy in a specific frequency range, as calculated through Fourier transform, and spectral frequency are also taken in. A flow chart of the methodology is given in Figure.1.

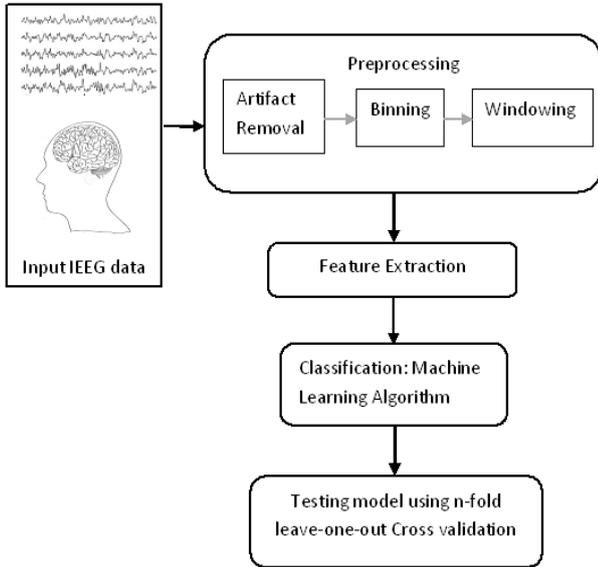


Figure 1: Flow of methodology

A total of 56 features are extracted for every 30-second non-overlapping segment of the EEG signals. For every 10 minutes recording, the 56 features derived from the 30-sec segment are concatenated to derive a 1220 dimensional feature vector representing the recording. Features extracted differ in number among subjects because the features also quantify the correlation between channels, and hence depend on the number of channels for each subject. For canine subjects, dog 1 to dog 4, each with 16 channel iEEG data, feature vectors of length 768 are extracted while for dog 5, with 15 channels, a feature vector of length 705 is extracted. For Human subjects, each with 24 channels, feature vectors of length 1334 are extracted. The dataset is shuffled to eliminate data grouping which may have a bearing on the training algorithm. After shuffling, the data is partitioned into randomized fivefold training and a validation sets.

After testing a range of machine learning algorithms suited for classification of the signal into interictal or preictal classes, we use three algorithms for this task and in addition to these, four ensemble models are explored. We lay out below a comparative analysis of the applicability and results for each of these. The three shortlisted algorithms are listed below:

A. Logistic Regression

Logistic Regression is primarily used when the response variable is categorical and we would like to predict the probability of the particular output given our input x as

$$p(x) = \frac{1}{1 + e^{-(\alpha + \beta x)}} \tag{1}$$

where α is the intercept while β denotes the regression

coefficient.

Adding more independent variables to the model will increase the variance. However this can result in overfitting and reduce generalization, hence the regularisation terms are added. L1 regularisation takes into account the absolute difference of the predicted and computed value as a penalty whereas L2 takes into account the squared difference as the penalty. The main aspect where they differ is that L1 reduces the causal effect of the less important features to zero and removes them completely. In our approach we use L2 regularisation over L1 because L2 generally leads to smaller coefficients while L1 results in sparse coefficient vectors with just a few higher value coefficients which increases the variance which corresponds to overfitting.

B. Support Vector Machines (SVM)

SVM determines non-linear class demarcations boundaries by cleverly using linear models. The linear classification model build in the new space serves as a non-linear decision boundary for the input space. SVMs works as follows: For a training data of the form $\{X_i, y_i\}$, where each y_i is either 1 or -1 denoting the class of the n dimensional input X_i , the aim is to compute a maximum-margin. Hyper plane that classifies the vector X_i into one of the two groups, and also making sure that the distance of the nearest point X_{ij} is maximum from the plane. A hyperplane is defined by a set of points $\{X_i\}$ satisfying

$$w \cdot x - b = 0 \tag{2}$$

The bounds $p(w,b)$ of the calculated hyperplane $H(w,b)$ is the distance from the hyperplane to the support vectors, i.e,

$$\rho(w, b) = \min_i \frac{|w \cdot x_i + b|}{\|w\|} \tag{3}$$

In this work, we define some kernels which enable us to work in higher dimensions (with respect to the input vector) without computing the coordinates of the data in the higher dimension. This approach is generally computationally cheaper than computing the coordinates of the data in the newly defined space. A standard SVM seeks to find a margin that separates all positive and negative examples. However, this can lead to poorly fit models if any examples are mislabeled or extremely unusual. To account for this, in 1995 Cortes and Vapnik et al. [24] proposed the idea of a “soft margin” SVM that allows some examples to be “ignored” or placed on the wrong side of the margin; this innovation often leads to a better overall fit.

C. Decision Trees

Decision Tree is a supervised learning algorithm which revolves around the idea of formulating rules and performing decision-based splitting based on different attributes to construct a tree structure. The splitting and decision rules learned by the algorithm is based on information gain, defined as the effective change in entropy after a decision rule has been extracted based on an artifact a . Information gain after splitting can be defined as

$$IG(w, b) = H(s) - H(s, a) \tag{4}$$

Epileptic Seizure Prediction Through Machine Learning and Spatio-Temporal Features Based Time Series Analysis of Intracranial Electroencephalogram Data

where $H(s,a)$ is given by

$$H(s, a) = \sum_{i=0}^n P(x) H(x) \quad (5)$$

Here $P(x)$ is the probability of event x . Decision tree is very sensitive to the depth that is chosen. Greater the depth of the tree, more it tends to over fit (high variance). On the other hand if the tree is too short the results are very generalized and many false positives are encountered.

D. Sequential Ensemble Models (Adaboost)

AdaBoost is an ensemble supervised learning algorithm that combines a set of weak classifiers to boost their performance. Each sample in the dataset is weighted and the weak learner is trained on a subset of the data. The misclassified samples of the weak learner are assigned higher weights. This enhances the probability of the sample being used in the training of the next classifier. The algorithms aim to create a classifier that focuses on examples misclassified in the previous steps. Each classifier is also assigned a weight which depends on the accuracy achieved. The equation for classification can be represented as:

$$F(x) = \text{sign}\left(\sum_{m=1}^M \theta_m f_m(x)\right) \quad (6)$$

Where f_m stands for the m^{th} weak classifier and θ_m is the corresponding weight. Below we describe briefly how an ensemble model works:

For a dataset of n points where X_i belongs to R^d and y belongs to $\{1, -1\}$ denotes the two classes, initial weights of the points are set as $w_i(X_i, y_i) = 1/n, i=1,2,\dots,n$. Each of the M classifiers are trained on the dataset. We select the one with the lowest weighted Classification error. The error weights of the selected classifier are calculated by the equations

$$\epsilon_m = E_{w_m} [1_{y \neq f(x)}] \quad (7)$$

$$\theta_m = \frac{1}{2} \ln\left(\frac{1-\epsilon_m}{\epsilon_m}\right) \quad (8)$$

The modified weights θ_m depend on the error rate of the classifiers ϵ_m which is defined as the ratio of the number of misclassification over the training set size.

E. Sequential Ensemble Models (Gradient Boost)

Gradient boosting algorithm is a supervised model which consists of an ensemble of weak prediction classifiers, predominantly decision trees. The classifiers are trained sequentially over the training set where the weak models learn from the misclassification of the previous models. Each model contributes towards reducing the loss function and minimizing the error rate to provide a more accurate estimate of the response variable. The ensemble model due to the use of the boosting exhibits high bias and low variance.

F. Parallel Ensemble Models (Random Forest)

Random forest is a supervised learning algorithm. Random Forests consist of multiple decision trees initialized with

different hyper-parameters ensembled using bagging technique. Bagging is a mechanism in which the predictions from multiple base models are used together for training. The main advantage of Random Forest over the decision trees is its ability to prevent over fitting as it randomizes the feature subset and builds smaller trees for classification as opposed to a single deep tree. The prediction from the model is achieved by a voting process wherein votes are received from each tree. The important hyper-parameters are the number of trees it uses for final prediction and the number of features in the subset that each tree uses.

G. Parallel Ensemble Models (Extra Trees)

Extra Trees is a supervised learning algorithm and is a modification of Random Forest. In Extra Trees, the process of feature selection is completely randomized for different trees as opposed to the random forest where feature selection is based on specific rules. The splitting threshold of the nodes in Extra Trees is also randomized whereas it is fixed in a Random Forest. The reason that Extra Trees perform better than the Random Forest in some cases is that it makes the decision boundaries smooth and does not use the bagging mechanism which is computationally expensive when the volume of the data is considerably high. Extra Trees is more generalized and tolerant against overfitting since the hyper-parameters of each tree is different and hence the performance and prediction of each tree has a minimal correlation.

IV. RESULTS AND DISCUSSION

After artifact removal and feature extraction from the signal, the features are scaled so that the coefficients are penalized based on their predictive power and not their amplitude. These features are used to train the following four broad categories of models; Generalized Regression models (Logistic Regression), Support Vector Machines (Linear and RBF Kernel Support Vector Machine Classifiers), Decision Trees (Decision Tree Classifier) and Tree Ensemble models (Parallel ensemble models: Random Forests, Extra Trees Classifiers and Sequential ensemble models: Adaboost and Gradient Boosting Classifiers). The training dataset is skewed towards the interictal class. Hence instead of using accuracy as a measure, we use the area under the receiver operating characteristic curve as a measure. This also allows us to penalize for false negatives. The performance of the model for each of the subject in terms of Area under ROC curve is given in Table 1. The corresponding ROC curves for the subjects for Linear SVM, Logistic Regression and AdaBoost is shown in Figure 3.

The results in Table 1 show that simplicity rules over common models as Logistic Regression and Linear kernel SVM tend to classify the test segments for five out of seven subjects almost perfectly, as seen in Columns 1 and 2 of Table 1. One expects kernel SVM to outperform Linear SVM, for non-linear iEEG signals with high dimensional feature representation used. However, the contrary appears to be true for our experiments. We can infer that the linear kernel performs better than nonlinear kernels such

Table 1: Average Area Under ROC curve values for all the subjects for the four categories of models

Subject Details	SVM	Generalized Regression	Decision Tree	Tree Ensemble Models			
	Linear SVM	Logistic Regression	Decision Tree	Parallel Ensemble Models		Sequential Ensemble models	
				Random Forests	Extra Trees	Gradient Boosting	AdaBoost
Dog 1	0.86	0.88	0.56	0.87	0.89	0.92	0.94
Dog 2	0.99	0.99	0.76	0.98	0.97	0.99	0.97
Dog 3	0.96	0.96	0.68	0.9	0.92	0.95	0.95
Dog 4	0.96	0.97	0.68	0.9	0.88	0.94	0.93
Dog 5	0.99	0.99	0.78	0.99	0.99	0.99	0.97
Patient 1	0.97	0.95	0.78	0.97	0.98	0.97	0.92
Patient 2	0.96	0.93	0.90	0.93	0.97	0.98	0.96

Table 2: Average Area Under ROC Curve values for all the subjects for Tree Ensemble approaches, before and after Regularisation

Subject Details	Tree Ensemble Models - Before Regularisation				Tree Ensemble Models - After Regularisation			
	Parallel Ensemble Models		Sequential Ensemble Models		Parallel Ensemble Models		Sequential Ensemble Models	
	Random Forests	Extra Trees	Gradient Boosting	Ada Boost	Random Forests	Extra Trees	Gradient Boosting	Ada Boost
Dog 1	0.78	0.77	0.84	0.87	0.87	0.89	0.92	0.94
Dog 2	0.96	0.96	0.98	0.96	0.98	0.97	0.99	0.97
Dog 3	0.87	0.89	0.92	0.92	0.90	0.92	0.95	0.95
Dog 4	0.86	0.87	0.91	0.91	0.90	0.88	0.94	0.93
Dog 5	0.98	0.97	0.97	0.98	0.99	0.99	0.99	0.97
Patient 1	0.91	0.93	0.87	0.90	0.97	0.98	0.97	0.92
Patient 2	0.91	0.99	0.99	0.93	0.93	0.97	0.96	0.96

as RBF are not necessarily better at classification than the linear one. Since the distribution of classes in our data is imbalanced and the data is limited, SVM with non-linear kernel tend to overfit on the training data due to their complexity, that is, there is not enough data to generalize over a complicated non-linear decision surface.

Logistic Regression performs at par or better as seen in Column 3 of Table 1 for four of the subjects when compared to the best performing Gradient Boosted Classifier. When the probability estimate is not of concern, we can consider Logistic Regression as similar to a linear SVM. Where speed is of essence, a Logistic Regression classifier may serve the purpose.

The dimensionality of the data bears on the results of decision trees as seen in Column 3 of Table 1. The huge number of features slow down the task and gives comparatively poorer results (Average area under ROC curve values is at 0.73 across all seven subjects). Decision trees have high variability, which is mainly caused by their acquisitive approach. Each decision split in the first level nodes shape the tree differently. Even a single additional data point is enough in many cases to get a totally different tree, especially if the sample is small or if the data is noisy.

The interesting observation, however, is that using Tree Ensemble models, the classification performance improves considerably over those subjects for which the results were comparatively less accurate using the other models (even Linear SVM and Logistic Regression). The average performance increases considerably for dog 1 and patient 2,

proving the tree ensemble models generalize better. However, they do not improve the performance of those subjects for which other models already give good results. We see that ensemble models prove to be more robust at the classification task. Ensemble models are prone to overfitting on noisy datasets and in high dimensional spaces (“the curse of dimensionality”). For noisy data, boosting tries to enforce a hard margin giving higher weight to outliers, this gives rise to the dilemma of non-smooth fits and overfitting. AdaBoost, in particular, can also suffer in that regard, as it is simply a linear combination of classifiers which themselves suffer from the problem. For data with large dimensionality, all the features do not contribute towards the classification and can be considered as noise. Boosting can suffer from overfitting on the training data. If the noisy features contribute to poor classification, Adaboost may compound the problem as the later classifiers maybe emphasizing the noisy samples and leading to poorer results. A classifier’s prediction for an instance x (having noisy

Epileptic Seizure Prediction Through Machine Learning and Spatio-Temporal Features Based Time Series Analysis of Intracranial Electroencephalogram Data

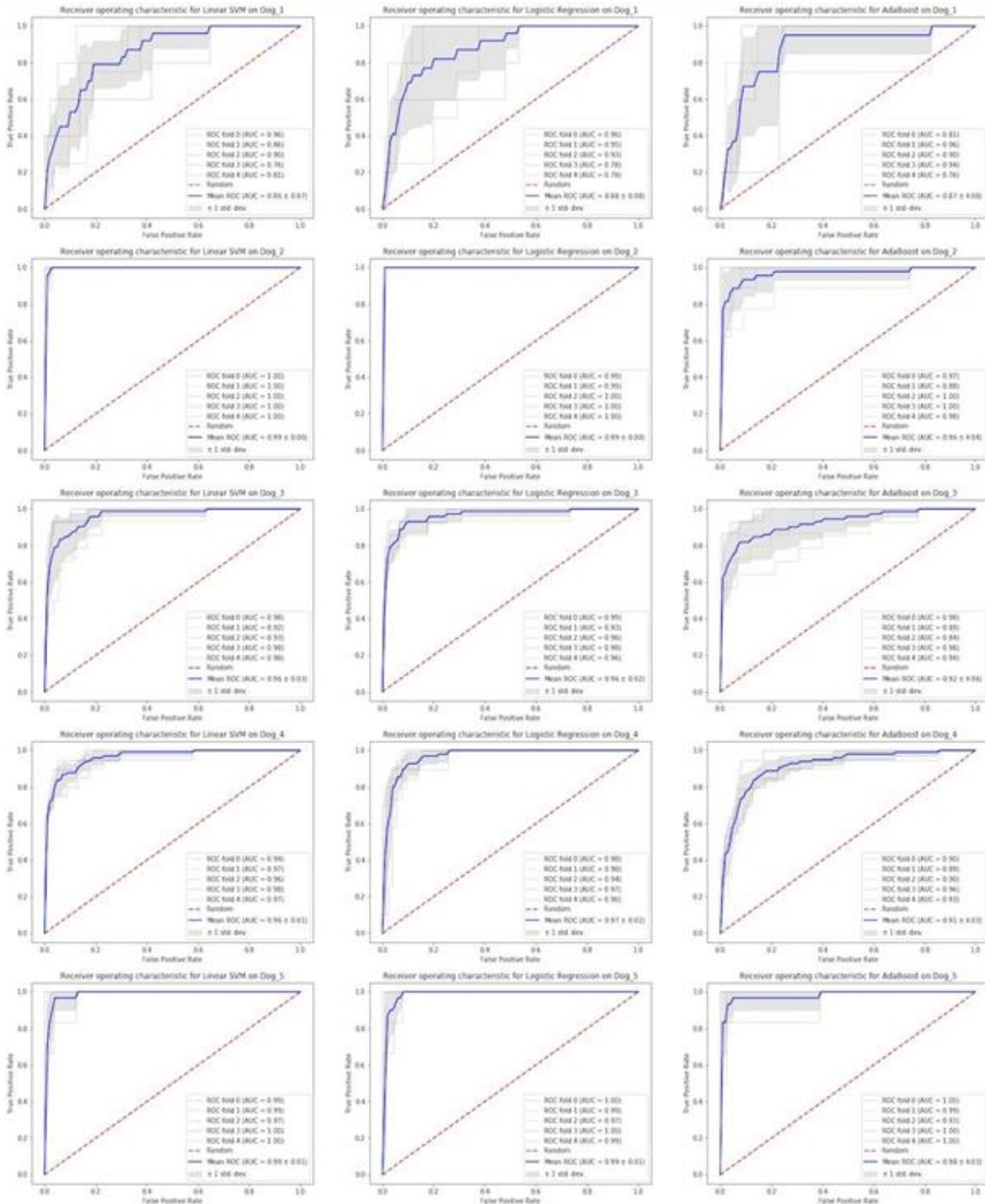


Figure 2: Receiver Operating Characteristic Curves for Canines Subjects on Linear SVM, LogisticRegression and AdaBoost

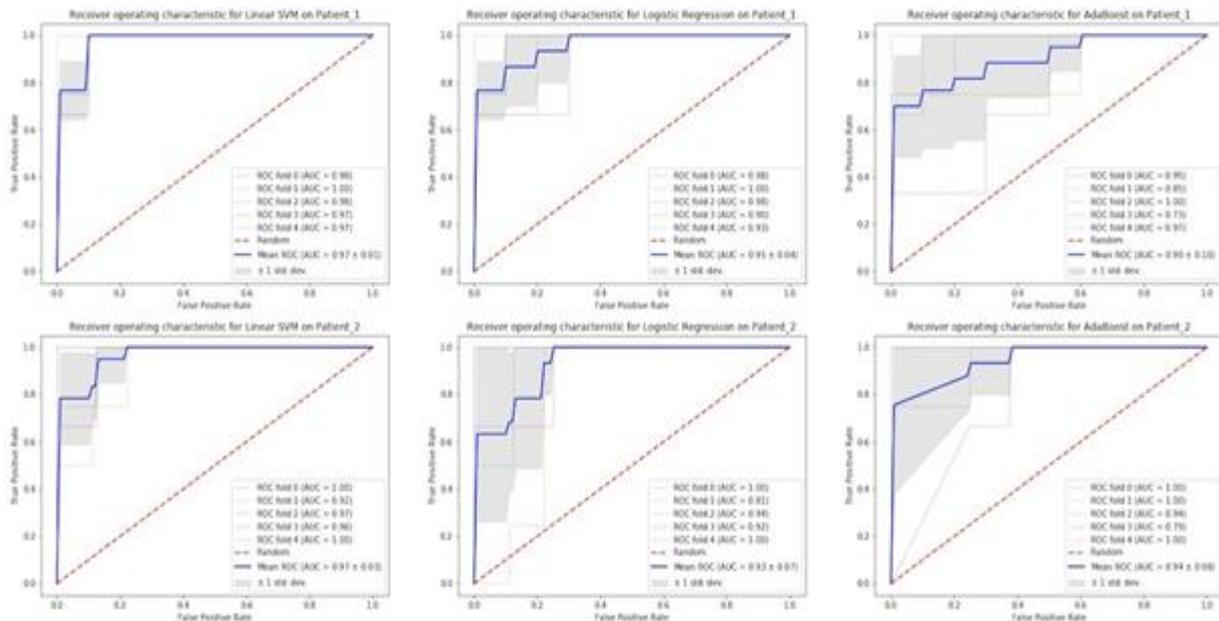


Figure 3: Receiver Operating Characteristic Curves for Human Subjects on Linear SVM, Logistic Regression and AdaBoost

attributes/outlier instance) may result in a very high positive value, but because the actual label is negative, this prediction would suffer a large loss/penalty since the penalty is exponentiated. This suggests that the classifier arrived at using the algorithm might not be the ideal one since it would seek to minimize the total exponential loss so this one outlier/noisy point ends up having a very strong influence on the final model learned. We see from the last column of Table 1 that though the le approach, its performance is not at par with the other classifiers considered. Under Ensemble approaches, Regularisation is a good aid. Since we have trees as base estimators for boosting ensemble models, they suffer from high variance and hence regularisation helps decrease that. Often this is done in cases when the problem is ill-posed, like ours where the number of parameters is greater than the number of samples. Regularisation is performed using model hyper-parameter tuning on the loss function, sub sample size (for Gradient Boosting), learning Rate (for both AdaBoost and Gradient Boosting), i.e. regularisation via shrinkage and stochastic gradient descent is carried out. Regularisation via shrinkage for gradient boosting would be, setting the learning rate below 1 and increasing the number of estimators which improves performance considerably. The learning rate parameter in Boosting shrinks the contribution of each new base estimator that is added in the series. In combination with shrinkage, stochastic gradient boosting produces more accurate models by reducing the variance via Bagging (Bagging reduces variance by averaging). From Table 2, we see the application of regularisation improves the performance of each of the classifiers for each of the subjects. The effect of regularisation is significant. For subjects dog 1 and patient 1, analyzing the results in Table 2, we can infer that regularisation has been least useful for Adaboost and considerably beneficial for ExtraTrees and Gradient Boosting.

V. CONCLUSION

From the aforementioned results, we can conclude the performance of Gradient Boosting Classifier seems to be the best amongst various specified algorithms. Though Gradient Boosting outperforms Logistic Regression and SVM, the latter two still provide very promising results. The training time and complexity is less than an Ensemble model and hence can be used when quick results are needed and time is a constraint. Deciding a benchmark procedure for seizure prediction is not easy as different approaches exist amongst researchers. One theory focuses on classification of the clips while training the models in a patient-specific manner since the variation between the signal of the patients may have a big range. The other supports a generalized training manner where the signals from all the patients should be taken. But since our vision was to look for approaches that will work well across different subjects without a considerable drop in performance, we find Boosting methods to work best, particularly Gradient Boosting Classifier.

In future work, as we discuss in the results section that Ensemble methods when regularized out-perform other approaches, the follow-up work should constitute better regularisation approaches on Ensemble models, using other regularized forms of Adaboost (RegBoost, AdaBoostReg, LPBoost, QPBoost) should help. Greater insights can be gained by using Convolutional Neural Networks as a feature extractor over the spectrogram of the EEG signals. Also using auto encoders over the raw EEG data for noise removal or on extracted features for removing redundant features could be highly productive.

Epileptic Seizure Prediction Through Machine Learning and Spatio-Temporal Features Based Time Series Analysis of Intracranial Electroencephalogram Data

ACKNOWLEDGEMENT

The authors would like to thank Mr. Arvind Babu, Alpha Learning Ltd., for his valuable contribution to this article and the Brain Computing Interface (BCI) laboratory set up at Vellore Institute of Technology, Chennai.

REFERENCES

1. U. R. Acharya, F. Molinari, S. V. Sree, S. Chattopadhyay, K.-H. Ng, J. S. Suri, Automated diagnosis of epileptic eeg using entropies, *Biomedical Signal Processing and Control* 7 (4) (2012) 401–408.
2. A. M. Murro, D. W. King, J. R. Smith, B. B. Gallagher, H. F. Flanigin, K. Meador, Computerized seizure detection of complex partial seizures, *Electroencephalography and Clinical Neurophysiology* 79 (4) (1991) 330–333.
3. B. Pushpa, D. Najumissa, Classification of epileptic seizure eeg signals using emd and anfis, in: *Informatics, Electronics & Vision (ICIEV), 2014 International Conference on, IEEE, 2014*, pp. 1–5.
4. R. Uthayakumar, Fractal dimension in epileptic eeg signal analysis, in: *Applications of Chaos and Nonlinear Dynamics in Science and Engineering-Vol. 3*, Springer, 2013, pp. 103–157.
5. K. Gadhouri, J. Kais, Seizure prediction for therapeutic devices: A review., *Journal of Neuroscience Methods* 260 (2016) 270–282. D. R. Freestone, P. J. Karoly, M. J. Cook, A forward-looking review of seizure prediction., *Current Opinion in Neurology* 30.2 (2017) 167–173.
6. D.R. Freestone, P.J.Karoly, and M. J. Cook, “ A forward-looking review of seizure prediction”, *Current Opinion in Neurology*, vol. 30.2, pp. 167-173, 2017.
7. B. H. Brinkmann, J. Wagenaar, D. Abbot, P. Adkins, S. C. Bosshard, M. Chen, Q. M. Tieng, J. He, F. Munoz-Almaraz, P. Botella-Rocamora, Crowdsourcing reproducible seizure forecasting in human and canine epilepsy., *Brain* 139.6 (2016) 1713–1722.
8. E. B. Assi, D. K. Nguyen, S. Rihana, M. Sawan, Towards accurate prediction of epileptic seizures: A review, *Biomedical Signal Processing and Control* 34 (2017) 144–157.
9. B. H. Brinkmann, E. E. Patterson, C. Vite, V. M. Vasoli, D. Crepeau, M. Stead, J. J. Howbert, V. Cherkassky, J. B. Wagenaar, B. Litt, et al., Forecasting seizures using intracranial EEG measures and SVM in naturally occurring canine epilepsy, *PloS one* 10 (8) (2015) 1–12.
10. L. Chisci, A. Mavino, G. Perferi, M. Sciandrone, C. Anile, G. Colicchio, F. Fuggetta, Real-time epileptic seizure prediction using AR models and support vector machines, *IEEE Transactions on Biomedical Engineering* 57 (5) (2010) 1124–1132.
11. L.-C. Lin, S. C.-J. Chen, C.-T. Chiang, H.-C. Wu, R.-C. Yang, C.-S. Ouyang, Classification preictal and interictal stages via integrating interchannel and time-domain analysis of EEG features, *Clinical EEG and neuroscience* 48 (2) (2017) 139–145.
12. M. Sood, S. V. Bhooshan, Parameter-selective based CAD system for epileptic seizure classification, *International Journal of Applied Engineering Research* 10 (10) (2015) 25389–25408.
13. N. S. Tawfik, S. M. Youssef, M. Kholief, A hybrid automated detection of epileptic seizures in EEG records, *Computers & Electrical Engineering* 53 (2016) 177–190.
14. N. Wang, M. R. Lyu, Extracting and selecting distinctive EEG features for efficient epileptic seizure prediction, *IEEE Journal of Biomedical and Health Informatics* 19 (5) (2015) 1648–1659.
15. Z. Zainuddin, K. H. Lai, P. Ong, An enhanced harmony search based algorithm for feature selection: Applications in epileptic seizure detection and prediction, *Computers & Electrical Engineering* 53 (2016) 143–162.
16. Z. Zhang, K. K. Parhi, Low-complexity seizure prediction from iEEG/sEEG using spectral power and ratios of spectral power, *IEEE transactions on biomedical circuits and systems* 10 (3) (2016) 693–706.
17. S. S. Alam, M. I. H. Bhuiyan, Detection of seizure and epilepsy using higher order statistics in the emd domain, *IEEE journal of biomedical and health informatics* 17 (2) (2013) 312–318.
18. A. Temko, A. Sarkar, G. Lightbody, Detection of seizures in intracranial eeg: Upenn and mayo clinic’s seizure detection challenge, in: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2015*, pp. 6582–6585.
19. D. P. Subha, P. K. Joseph, R. Acharya, C. M. Lim, Eeg signal analysis: a survey, *Journal of medical systems* 34 (2) (2010) 195–212.
20. D. Chen, S. Wan, J. Xiang, F. S. Bao, A high-performance seizure detection algorithm based on discrete wavelet transform (dwt) and eeg, *PloS one* 12 (3) (2017) e0171318.

21. T. N. Alotaiby, S. A. Alshebeili, F. M. Alotaiby, S. R. Alrshoud, Epileptic seizure prediction using csp and lda for scalp eeg signals, *Computational intelligence and neuroscience* 2017 (2017).
22. K. Fei, W. Wang, Q. Yang, S. Tang, Chaos feature study in fractional fourier domain for preictal prediction of epileptic seizure, *Neurocomputing* 249 (2017) 290–298.
23. M. Borowska, Entropy-based algorithms in the analysis of biomedical signals, *Studies in Logic, Grammar and Rhetoric* 43 (1) (2015) 21–32.
24. C. Cortes, V. Vapnik, Support-vector networks, *Machine Learning* 20 (3) (1995) 273–297.

AUTHORS PROFILE



N. Ilakiyavelvan, is currently working as Assistant Professor (Senior) in the School of Computing Science and Engineering at Vellore Institute of Technology, Chennai. He completed his Bachelors’s degree in Information Technology and Master’s degree at Anna University. He has 9 years of teaching experience and research interests in the field of Software Engineering, Natural language Processing, Biomedical signal analysis.



Hardik Bhatt, presently a Computer Science undergrad from Vellore Institute of Technology, Chennai. He is a Machine Learning Engineer at Mesh and worked at Alpha Learning, Euro Exim Bank UK. Computer Science major with a zeal for Machine Learning Projects.



Utkarsh Shukla, presently a Computer Science undergrad from Vellore Institute of Technology, Chennai. He has research interest in the area of Brain Computing Interface, Biomedical Signal Analysis etc. He is a motivated technologist, bringing in innovation in the way things work in the world of science



A. Nayeemulla Khan is a professor in the School of Computing Science and Engineering at Vellore Institute of Technology, Chennai. He has more than 20 years of teaching and industry experience, with over 6 years of research in the field of Speech Processing. He has research interests in the areas of Speech Processing & Speech Applications, Biomedical Signal Analysis, Pattern Recognition, etc.



A. Shahina is a professor in the department of Information Technology at SSN College of Engineering. She has 14 years of teaching and research experience, with over 5 years of research exclusively in the field of Speech Processing, one of the widely growing and popular research areas. She aims to develop speech based clinical applications, and technologies for viable biometric person authentication systems through sustained research.