Multi-Feature Learning Model for Epilepsy Classification Supervised by a Highly Robust Heterogeneous Deep Ensemble

Mohammad Asif A Raibag [a], Dr. J Vijay Franklin[b], Dr. Rashal Sarkar[c]

[a]Associate Professor, Department of Computer Science Engineering, PA College of Engineering, Mangalore, India
[b]Professor,Department of Computer Science Engineering,Bannari Institute Technology, Sathyamangalam, India
[c] Professor, Department of Computer Science and Electronic,University of Science and Technology, Meghalaya, India

Abstract: In this present work, we propose a novel heterogeneous deep ensemble-based multi-feature learning environment for epilepsy classification. The proposed model is built to deal with the present prominent issues like data imbalance, low accuracy, and most importantly the need for a reliable classification model. To accomplish this, a multi-level enhancing technique is used to address the problem of class imbalance, which included data sampling with a 95% confidence interval. A variety of sample techniques are used, including random sampling, down - sampling, and the synthetic minority oversampling technique (SMOTE). We have used algorithms like significant predictor test (SPR), Cross-Correlation Analysis (CRA) and Principle Component Analysis (PCA) to select features after retrieving samples data. The main goal of using various feature selection methods was to keep the best features for high accuracy and minimal computation. Using Decision Tree (DT), Logistic Regression (LR), Artificial Neural Network (ANN) with Radial Basis Function (RBF) and Levenberg Marquardt (LM) learning methods and Probabilistic Neural Network (PNN) algorithms as the base classifier, we created a first-of-its-kind heterogeneous deep ensemble model. Maximum Voting Ensemble (MVE) and Best Trained Ensemble (BTE) were used as ensemble decisions for two-class classification, determining if each sample in the dataset is epileptic or not. The proposed system's superiority over main existing techniques was validated by simulation-based performance comparisons in terms of accuracy (93.88%), F-measure (0.91), and AUC (0.94).

Keywords: Epilepsy, Ensemble Learning, Multi-Feature Learning, Artificial Neural Network, Computer Aided Diagnosis.

1. Introduction

Epilepsy is a common brain disorder in which the balance between cerebral excitability and inhibition is tipped toward uncontrolled excitability, resulting in repeated unprovoked seizures. This disease affects about 2% of the population, making it one of the most common severe neurological conditions. Epilepsy is one of the diseases that affects a large population of people in both the developed and developing worlds, according to the WHO. These seizures frequently result in a brief loss of consciousness, putting the person in a life-threatening situation and interfering with day-to-day activities. A seizure can occur at any age, regardless of gender, location, social status, race, or period. Epilepsy does not have a permanent prognosis; however, it is estimated that nearly 70% of people with epilepsy will live seizure-free life if correctly diagnosed and treated. In the treatment of epilepsy, EEG signals are used, these signals are too complex to deal with, the manual diagnosis of these complex signals is a time-consuming task that often necessitates the use of a specialist, and sometimes due to these complexities, decisions are made incorrectly.

In the last few years many interesting work has been carried out to build highly accurate and efficient CAD solutions for epilepsy classification and diagnosis. Considering this as motivation, a novel and robust epilepsy classification model for early and optimal diagnosis decision is proposed. Techniques belonging to machine learning and deep learning are been used strenuously towards epilepsy classification, however, in case of developing a CAD solution, machine learning methods have been found to be more effective. Though major researchers have implemented and proposed machine-learning methods for treating epilepsy, they have failed to address key issues such as low accuracy, class imbalance, data limitations, and ultimately low reliability. As a result, major approaches for epilepsy classification and prediction for real-world applications are hampered. Recalling the fact that the majority of the researches have applied a benchmark dataset named Epileptic Seizure Recognition Dataset very commonly used for epileptic seizure detection or classification, the sample size of this dataset is 11,500 instances, due to such large volume of information there is high probability of data imbalance. Interestingly, this key problem was never addressed in existing researches. On the other hand, lower accuracy too put question on acceptability of the at-hand or existing solutions towards real-world clinical decisions for epilepsy prediction or classification. Therefore, with this intention, a significant number of constructive investigations into Computer-Aided Diagnosis solutions (CAD) have been conducted in recent years, with various methods for health diagnosis and related decision-making systems. CAD systems are one of the most sought-after technologies that provide high precision analysis and decision-making capabilities, which can be developed with an optimistic software design paradigm. In keeping with such a positive design goal, the objective of this paper is to create a reliable epilepsy classification system by using state-of-the-art methods in seizure detection and treatment methods. Researchers hope to use these methods in conjunction with the advancement of artificial intelligence (AI) and machine learning (ML) technologies to improve medical practice in epilepsy diagnosis.

In this paper a novel and robust heterogeneous ensemble learning assisted epileptic classification system is developed. Unlike classical machine learning methods, our proposed system incorporates a multi-phased optimization concept, where at first it addressed the data or class imbalance problem for this we applied different data sampling concepts, including random sampling, down sampling and SMOTE, which were primarily employed to alleviate class-imbalance problem to ensure accurate prediction. However, sampling methods in conjunction with the original epileptic seizure recognition data set enabled significantly large data for training and consequential classification, to avoid unwanted computation; we extracted certain relevant features for epilepsy classification, which is explained in the following sections. Once retrieving the sample data, we performed feature selection using different algorithms like significant predictor test (SPR), Principle Component Analysis (PCA) and Cross-Correlation Analysis (CRA) to retain most suitable features to achieve high accuracy with low computation. Subsequently, once selecting the suitable set of features, we applied a novel heterogeneous ensemble-learning model for multi-class classification.

Noticeably, in the proposed heterogeneous ensemble model, we applied base classifiers from the different categories including Probabilistic Neural Network (PNN), Logistic Regression (LR), and Decision Tree (DT), Artificial Neural Network (ANN) with Gradient Descent (GD), Radial Basis Function (RBF) and Levenberg Marquardt (LM) learning methods. Noticeably, hypothesizing that the inclusion of deep features can enable higher accuracy, we applied ANN-RBF and ANN-LM with different number of hidden layers (here, 1, 2 and 3 hidden layers). Thus, our proposed heterogeneous ensemble model applied a total of 12 base classifiers, which performed epilepsy classification for each sample distinctly. Eventually, as ensemble decision, we applied two well-known techniques Maximum Voting Ensemble (MVE) and Best Trained Ensemble (BTE) to finally perform the multi-class epilepsy classification. The performance assessment in terms of accuracy, F-measure and area under curve (AUC) exhibits superior performance over other existing methods. The rest of the contents in this manuscript are divided as follows. Section II discusses the related work, followed by Section III that discusses the overall proposed system and its implementation, while the simulation results and allied inferences are given in Section IV. Overall research conclusion is given in Section V, and the references used in this research are presented at the last part of the manuscript.

2. Related Work

Kostas et.al. [1] in this paper, a Long Short-Term Memory (LSTM) network with deep learning was proposed for predicting epilepsy. After testing several modules to evaluate seizure prediction performance using various preictal window lengths, a two-layer LSTM network was chosen. When compared to other traditional machine learning techniques in the literature, the proposed work significantly improved seizure prediction performance. The study of Satapathy et al. [2] was based on a large EEG dataset that was used to test the two methods, SVM and Neural networks, for seizure detection using different kernel methods. The majority voting system was used to evaluate the performance of each classifier, and it was discovered that the SVM was more efficient than other neural networks. Bidirectional long short-term memory (BiLSTM) model integrated with attention was proposed as a novel deep learning-based approach. The proposed model included the following components: an attention mechanism for capturing spatial features, a BiLSTM for extracting more temporal features from the input, a time-distributed fully-connected layer for extracting features from each time step, a pooling layer for extracting global features from each sample, and finally a fully connected layer for extracting further features to reduce the last dimension of input matrix into several classes [3]. The Softmax layer uses the fully connected layer's computed results to calculate probabilities that each sample belongs to a class. The model yielded good results, which are considered superior to many existing approaches. Lacunarity and Bayesian linear discriminant analysis (BLDA) were combined to create a seizure detection algorithm [4]. The wavelet coefficients at scales 3, 4, and 5 were chosen after wavelet decomposition on EEGs were performed. From the selected scales, features such as lacunarity and fluctuation index were extracted and fed into the BLDA for training and classification. The average accuracy obtained in this work was 96%.

Ansari et al. [5] wanted to optimize feature selection for seizure detection automatically. For this, they used deep CNN to extract optimal features, which were then fed to a random forest for classification. Accuracy of 77% was reported. In this work, a multi-view deep learning model for capturing brain abnormalities associated with seizures using multi-channel scalp EEG signals was proposed [6]. An autoencoder-based model to learn inter and intra correlations of EEG channels was designed. These correlations were combined with features to detect seizures, which were extracted in supervised learning via spectrogram representation. Accuracy of 94% was obtained in this system. In [7] to detect seizures in EEG signals, a ChannelAtt model was proposed, which was an end-to-end multi-view deep learning model with a channel-aware attention mechanism. The model used global-based attention, which can dynamically store channel contributions and capture relationships between channels. The CHB-MIT EEG dataset was used as experimental data in the model's evaluation, and the proposed system achieved a 96% accuracy rate.

Amin and Kamboh [8] created the RUSBoost algorithm to process imbalanced seizure/non-seizure data, and they used RUSBoost and the CHB-MIT data set to conduct patient-specific experiments. The method was simple to learn and performed well, with a 97% seizure detection accuracy. Hunyadi et al. [10] this work presents a novel patient-specific seizure detection algorithm. The focus was to extract meaningful information from the dataset and providing a learning algorithm to exploit it. They used the spatial distribution of the ictal pattern, which is characteristic of a patient's seizures. To convey structural information of the channel-feature matrices extracted from the EEG, the proposed algorithm used nuclear norm regularization. This method was compared to two existing approaches, and it was discovered that the proposed algorithm outperformed the other two methods significantly. Truong proposed a seizure detection method based on intracranial EEG data. The frequency and time domain features were extracted, and the data were classified using the Random Forest classifier. The goal of this proposed method was to reduce the number of channels and to improve computational efficiency by selecting the channels that contribute the most to seizure identification [12].

3.Proposed Work

The following figure shows the flowchart for the epilepsy model and the detailed explanation of these steps are discussed below.

Figure-1. Proposed Classification Model

3.1) Dataset

The Epileptic Seizure Dataset used in this paper is from the UCI-ML repository, which is open to the public. The EEG dataset contains 11,500 samples, each with 178 features labelled as X1, X2, X3….X178. The samples are categorized into five different classes Y = {1, 2, 3, 4, 5} in the last column based on the criteria shown in the table below. All EEG signals recorded here used the same 128-channel amplifier system, digitized with a sampling rate of 173.61 Hz and 12-bit A/D resolution. Each sample had 178 features indicating the brainwave measurement per second for the different mentioned cases. Fig. 1 shows a sample view of the epileptic seizure dataset.

Out[29]:

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	...	X169	X170	X171	X172	X173	X174	X175	X176	X177	X178
0	135	190	229	223	192	125	55	-9	-33	-38	...	8	-17	-15	-31	-77	-103	-127	-116	-83	-51
1	386	382	356	331	320	315	307	272	244	232	...	168	164	150	146	152	157	156	154	143	129
2	-32	-39	-47	-37	-32	-36	-57	-73	-85	-94	...	29	57	64	48	19	-12	-30	-35	-35	-36
3	-105	-101	-96	-92	-89	-95	-102	-100	-87	-79	...	-80	-82	-81	-80	-77	-85	-77	-72	-69	-65
4	-9	-65	-98	-102	-78	-48	-16	0	-21	-59	...	10	4	2	-12	-32	-41	-65	-83	-89	-73

5 rows × 178 columns

Fig. 1. Sample Epileptic Dataset

[https://archive.ics.uci.edu/ml/datasets/Epileptic+Seizure+Recognition] [15].

3.2)Data Pre-processing and Sampling

We discovered a class imbalance in the dataset after reviewing the data and its element-by-element information. In other words, the class-priors were significantly skewed towards the Non-Epileptic class in our benchmark data. In this case, training a model could lead to any prediction output being labeled as positive, resulting in a high likelihood of false alarm. When any machine learning classifier is applied to such class-imbalanced data, it may result in inaccurate detection and diagnosis decisions, which can be disastrous in the real world. Regrettably, no significant prior research had addressed these limitations. As a result, we used data sub-sampling techniques as pre-processing methods in this paper. These sub-sampling techniques can alleviate the problem of class imbalance, which can prevent machine-learning models from learning efficiently and providing higher accuracy. As a result, we used three different sampling methods in this paper: random sampling, down sampling, and SMOTE. We were able to obtain random sampled data, down-sampled data, SMOTE data, and the original data to evaluate the efficiency of sub-sampling based on relative performance. Obtaining these four-sample data, we performed feature extraction and selection, which is discussed in the sub-sequent sections.

3.3) Feature Extraction

Numerous features from the time domain, frequency domain, and nonlinear dynamics have been discovered in the literature. A set of features that were potentially useful for epilepsy classification and had computationally reasonable requirements for real-time implementation were chosen in this study. Expertise, observations, and our understanding of EEG signal characteristics are used to extract and select features. The following features were extracted from the raw data using the Fast Fourier Transform (FFT) Method: Amplitude, Phase, Energy, Mean/Average, Standard Deviation, Entropy, and Median.

3.4) Feature Selection

There is a high probability of insignificant data elements being present in the dataset that have no bearing on the final prediction and also degrade prediction accuracy. Therefore, only the most important features that have a significant impact on epilepsy classification are retained for this we applied the following feature selection methods. The Significant Predictor Method (SPM) uses the correlation concept between attributes to determine the level of significance of a variable for classification. It investigates the relationship between various data attributes and their importance in epilepsy classification. To achieve this, we used the rank-sum test method in our proposed method, which measures the correlation between users and their respective feature values in terms of classification. Each attribute was designated as an independent variable, with the output variable, or the class level, as the dependent variable. As a result, we identified elements with higher significance by estimating the level of significance for each attribute across the data. Because we used p=0.05 as the level of significance, only those data instances or elements with a p-value greater than that were kept, while the rest were discarded for further analysis.

Principle Component Analysis (PCA), here for each feature set and corresponding instance, we retrieved the principle component and Eigenvalues of the covariance or correlation. We estimated the distance of each feature instance of an element from the average principal component (0.5) in this case. Thus, those instances having higher distance were dropped, while the instances with low-eigen distance were retained, hypothesizing them to be significant towards classification.

Cross-correlation analysis (CRA) is a statistical method for displaying the degree to which two variables are related. Additionally, it states or signifies the strengths as well as the direction of the association. The link or association's strength can typically be measured in a range of plus or minus, 1 to 0. A value that is closer to 1 indicates a stronger connection or relationship. We used Pearson correlation-based CRA assessment in our proposed model. Because of the above-mentioned feature selection methods, we were able to obtain a variety of feature sets that could be used for classification. Specifically, we concentrated our efforts in this study on determining the most appropriate or optimal computing environment for epilepsy detection and classification. With this thought, we evaluated each feature as obtained above towards their suitability to yield higher accuracy. Once the optimal feature sets were obtained, we performed multi-class classification using our proposed heterogeneous deep ensemble-assisted multi-feature learning model for epilepsy classification. A detailed discussion of the proposed ensemble learning models comprising different machine learning algorithms is discussed in the subsequent sections.

3.6) Heterogeneous deep ensemble-based learning for two-class classification

In major at hand approaches authors have applied single machine learning algorithm as standalone classifier for epilepsy classification, however, different algorithms perform differently or their respective performance over the same data differs, signifying diversity of performance. In such case, generalizing a distinct approach as optimal solution is unfair. To avoid such issues, ensemble learning concept can be of great significance. In this paper we applied classifiers of the different categories that constitutes a HEL model, the overall base classifiers used in our proposed model are given as follows. In synch with above mentioned base classifiers, we designed two different ensemble learning concepts named Maximum Voting Ensemble (MVE) and Best Trained Ensemble (BTE) for classification of epilepsy. A snippet of these base classifiers and eventual HEL formation is given in the subsequent sections.

Logistic Regression (LR) is a regression approach often applied for data mining and text classification. In our proposed epilepsy classification problem, LOGR performs regression over the different independent variables (i.e., the amplitude, skew, energy, kurtosis and so on) and the dependent variable (i.e., subject class label). In this manner, the regression output classifies each subject as Epileptic or Non-Epileptic. In our proposed LOGR method, we applied (1) to perform linear regression over input features.

$$logit[\pi(x)] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \dots . + \beta_m X_m \qquad (1)$$

We have applied logit function, logit [π(x)] signifying the dependent variable while the independent variable be xi. The above derived mathematical model converts the dichotomous outputs using logit function and hence π(x) vary in between 0 to 1 to -∞ to +∞. In synch with the input epileptic dataset, in equation (2) m presents the total number of independent variables. On the other hand, the probability of a subject to be epileptic is given by π. Thus, the likelihood of the dependent variable i.e., subject as "Epileptic" or Non-Epileptic' is obtained as per (2).

$$\pi(x) = \frac{e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots\dots+\beta_m X_m}}{1+e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots\dots+\beta_m X_m}} \qquad (2)$$

Decision Tree (DT, C4.5) is one of the most applied data mining concepts in data-science. Its robustness and ease of implementation makes it suitable for numerous data mining, pattern and text classification purposes. Based on its applications suitability, different DT variants were proposed such as IDE, CART and C4.5, however, amongst them, C4.5 has been applied in major data science applications because of its ease of implementation and computationally-efficient rule mining ability. In our proposed model, it performs association rule mining amongst the different subject's features and corresponding subject class label. To achieve it, it performed association rule mining in between the split-condition the input epileptic features of each subject was divided into multiple branches at each node of the tree. Thus, applying Information Gain Ratio (IGR) information as per the different association rule mining, it classifies each subject as epileptic or Non-epileptic.

Deep Neuro-Computing Environment is one of the most used machine learning and artificial intelligence (AI) algorithms across academia-industries. Its learning efficiency over non-linear patterns and ease of implementation make it suitable for major classification problems. Functionally, ANN possess three layers; input layer, hidden layer and output layer (Fig. 2), where the data to be classified is fed as input to the input layer. Learning the data or input patterns, ANN often classifies data as per expected target categories at the output layer. During the course of learning and prediction, the efficiency of ANN primarily depends on the learning method used. Thus, to enhance learning method, which decides how efficient the neuro-computing model is capable of learning, while maintaining precision, different algorithms have been developed. Thus, based on the learning mechanism ANN evolved from ANN-steepest descent (ANN-SD) to ANN-Levenberg Marquardt (ANN-LM), and now Extreme Learning Machine (ELM). Fig. 2 presents a simple presentation of neural network with three layers; input, hidden and output layer. A probabilistic neural network (PNN) is a type of feedforward neural network that is commonly used to solve classification problems. With the input layer, hidden layer, and output layers, we used a three-layered PNN architecture in our proposed work. The overall model was built to perform two-class classification (K=2), but it can also be built to perform multi-class classification (Fig. 3).

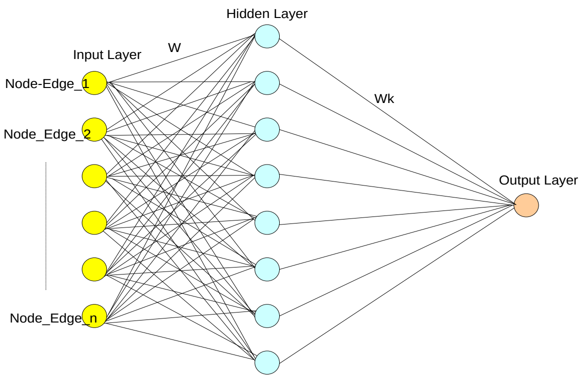

Fig. 2. ANN architecture with single hidden layer

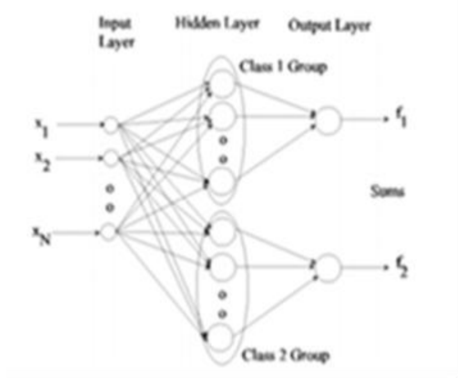

Fig. 3. PNN architecture

3.6) Heterogeneous Ensemble Learning Model

Maximum Voting Ensemble (MVE): To create a novel heterogeneous ensemble learning (HEL) model, we used the 12 machine learning models discussed above as the base classifier. All classifiers are run on the same dataset to create an ensemble structure, classifying each subject as "Epileptic" or "Non-Epileptic" and labeling them as "1" or "0." We were able to obtain the highest possible voting score by gaining the class output label of each subject. Finally, based on consensus, the subject with the highest score (either 1 or 0) is classified as Epileptic or Non-Epileptic. Best Trained Ensemble (BTE): Unlike MVE, BTE finds the best performing base classifier first. BTE was designed in such a way that it obtains the respective performance of each base classifier and, after identifying the classifiers with the highest accuracy, considers their average prediction as to the final classification output. To put it another way, we found a total of 5 base classifiers with higher accuracy and F-Measure values among all 12 base classifiers. We considered the cumulative (consensus) prediction as to the final subject-class after retrieving the best performing classifiers (with high accuracy and F-Measure).

4. Results and Discussion

Considering the up-surging significance of epilepsy classification, in this paper we made effort to design a multi-level optimization concept by improving data selection and pre-processing, feature selection and enhanced machine learning based classification. Here, our key motive was to identify a most suitable computing environment that would yield higher accuracy, reliability and computational efficiency, which are must for an efficient CAD solution. We obtained True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) values and derived performance parameters like accuracy, precision, recall and F-Measure, as per Table 1.

Parameter	Expression	Definition
Accuracy	$\frac{(TN+TP)}{(TN+FN+FP+TP)}$	Signifies the proportion of predicted fault prone modules that are inspected out of all modules.
Precision	$\frac{TP}{(TP+FP)}$	States the degree to which the repeated measurements under unchanged conditions show the same results.
Recall	$TP/((TP+FN))$	It indicates how many of the relevant items are to be identified.
F-measure	$2(Recall.Precision)/(Recall+Precision)$	It combines the precision and recall numeric value to give a single score, which is defined as the harmonic mean of the recall and precision.

Table 1. Performance Parameters

Observing results (Fig. 4 to Fig. 6), the performance of various classifiers with the type of feature selection employed and the sampling technique used is shown. The original data when used with the SPM method gave 66% accuracy for LR classifier. The other classifiers showed very dismal performance using SPM feature selection method. In contrast the accuracy of ANN-LM with three hidden layer was 79% using the PCA method. The other classifiers too showed decent performance when used PCA as the feature selection approach. The other feature selection method used was CRA and with this method the LR classifier showed 70% of accuracy. Next we used down-sampling technique and the three feature selection methods respectively and we had the following observations. SPM feature selection method gave accuracy of 80% for ANN-RBF classifier with two hidden layer. The PCA and CRA gave 81% and 87 % accuracy for ANN-RBF classifier with two hidden layer and ANN-LM classifier with two hidden layer respectively. Interestingly with SMOTE data, the classifiers performance improved drastically with all the three feature selection techniques. An accuracy of 91%, 91% and 93% for LR classifier using SPM, ANN-LM classifier with one hidden layer with PCA and ANN-LM classifier with three hidden layer with CRA method respectively was obtained.

Similarly, F-Measure with original data using SPM achieved 0.69 for ANN-LM classifier with three hidden layer, followed by PCA with 0.78 for ANN-RBF with two hidden layer and CRA with 0.85 for ANN-LM classifier with two hidden layer, F-Measure with down-sampled data was 0.79 for ANN-LM classifier with two hidden layer (SPM), 0.88 for ANN-RBF with one hidden layer (PCA), 0.87 for LR (CRA) respectively. Interestingly, the AUC performance by the classifiers on SMOTE data was found superior over other sampled data i.e., Original sample, up-sample and down-sample data. Based on above results, the superiority of SMOTE is justified over other sampled data. Observing Fig. 5, it can easily be found that amongst the all base classifiers as well as ensemble learning structures, BTE ensemble model exhibits the highest prediction accuracy of 83.7%.

Considering ensemble learning and their respective performance we found that BTE outperforms MVE, where the BTE achieves 82% of accuracy while the later could retain the highest accuracy of 78%. Though, the highest random performance accuracy by BTE was found

84.5%. Fig. 4 presents the F-Measure performance by the different machine learning classifiers, including BTE and MVE heterogeneous ensembles. Observing the results, we can easily find that BTE ensemble model achieves the maximum F-measure of 0.83, followed by LOGR (0.82), ANN-RBF (with two hidden layers) (0.815). Interestingly, Fig. 6 reveals that almost all classifiers possess similar AUC performance. The detailed simulation results with different sampling methods, feature selection and machine learning classifiers are given in Table 2 to Table 4.

Accuracy (%) with Original Data														
Feature Selection	**LOGR**	**DT**	**AGD1H**	**AGD2H**	**AGD3H**	**ANF1H**	**ANF2H**	**ANF3H**	**ALM1H**	**ALM2H**	**ALM3H**	**PNN**	**BTE**	**MVE**
SPM	66.73	56.41	63.53	66.67	63.37	66.87	57.11	66.57	59.22	65.51	65.22	66.02	68.46	69.72
PCA	69.87	70.12	64.66	73.37	74.81	69.78	71.27	78.11	71.87	72.44	79.31	70.77	71.51	74.22
CRA	70.61	66.74	62.47	66.55	65.38	71.16	66.66	69.35	68.62	68.62	69.97	66.35	69.72	78.71
F-measure with Original Data														
SPM	0.68	0.68	0.62	0.62	0.61	0.61	0.53	0.49	0.57	0.62	0.69	0.31	0.51	0.68
PCA	0.68	0.71	0.71	0.68	0.60	0.75	0.78	0.65	0.56	0.71	0.68	0.72	0.77	0.78
CRA	0.85	0.77	0.78	0.81	0.81	0.85	0.85	0.84	0.84	0.85	0.84	0.81	0.87	0.84
AUC with Original Data														
SPM	0.57	0.61	0.61	0.51	0.58	0.61	0.59	0.59	0.63	0.63	0.59	0.61	0.55	0.63
PCA	0.45	0.51	0.54	0.55	0.49	0.55	0.51	0.49	0.55	0.50	0.54	0.51	0.51	0.57
CRA	0.56	0.59	0.55	0.56	0.46	0.59	0.56	0.58	0.60	0.55	0.56	0.56	0.58	0.61

Table 2. Performance measures for Original Data

The performance analyses for the original data by the classifiers is given in Table 2. The results illustrated the overall accuracies of the classifiers. The (DT+SPM) has the lowest accuracy of 56%, followed by (AGD1H+PCA) and (AGD1H+CRA) with 64% and 62% respectively. Similarly, for F-Measure the lowest performance is of (ANF3H+SPM) with 0.49, followed by (ALM1H+PCA) and (DT+CRA) with 0.56 and 0.77 respectively. Finally (SPM+AGD2H), (LR+ PCA) and (AGD3H+CRA) have 0.51, 0.45and 0.46 respectively are the lowest AUC values for the original dataset. The best performance for this particular dataset is 78% (CRA+MVE) for accuracy, 0.87 (CRA+BTE) for F-Measure and 0.61 (CRA+MVE) for AUC.

Accuracy (%) with Down Sampled Data														
Feature Selection	**LOGR**	**DT**	**AGD1H**	**AGD2H**	**AGD3H**	**ANF1H**	**ANF2H**	**ANF3H**	**ALM1H**	**ALM2H**	**ALM3H**	**PNN**	**BTE**	**MVE**
SPM	74.15	77.88	77.35	68.35	68.11	80.5	80.55	78.11	78.55	80.25	80.95	79	81.66	84.97
PCA	78.45	78.61	68.97	74.25	80.25	79.88	81.55	80.75	78.26	79.11	80.77	81.45	85.55	88.53
CRA	84.35	85.5	77.35	76.22	79.52	86.66	81.78	79.55	86.9	84.39	81.24	84.71	86.15	89.85
F-measure with Down Sampled Data														
SPM	0.73	0.64	0.71	0.54	0.64	0.75	0.77	0.74	0.69	0.79	0.75	0.68	0.77	0.72
PCA	0.83	0.77	0.74	0.64	0.81	0.88	0.85	0.86	0.85	0.81	0.87	0.85	0.81	0.83
CRA	0.87	0.77	0.82	0.83	0.75	0.83	0.83	0.82	0.81	0.84	0.85	0.79	0.85	0.88
AUC with Down Sampled Data														
SPM	0.74	0.74	0.67	0.74	0.67	0.64	0.74	0.66	0.68	0.74	0.69	0.66	0.69	0.74
PCA	0.80	0.81	0.81	0.83	0.83	0.84	0.84	0.79	0.84	0.81	0.78	0.84	0.81	0.84
CRA	0.84	0.81	0.81	0.84	0.87	0.87	0.87	0.86	0.87	0.87	0.87	0.84	0.87	0.89

Table 3. Performance measures for Down Sampled Data

The performance analyses for the Down sampled data by the classifiers is as shown in Table 3. The results illustrated the overall accuracies of the classifiers. The (AGD3H+SPM) has the lowest accuracy of 68%, followed by (AGD1H+PCA) and (AGD3H+CRA) with 68% and 76% respectively. Similarly, for F-Measure and AUC we have the performance values displayed in table 3. The best performance for this particular dataset is 89% (CRA+MVE) for accuracy, 0.88 (CRA+MVE) for F-Measure and 0.89 (CRA+MVE) for AUC.

Accuracy (%) with SMOTE Data														
Feature Selection	**LOGR**	**DT**	**AGD1H**	**AGD2H**	**AGD3H**	**ANF1H**	**ANF2H**	**ANF3H**	**ALM1H**	**ALM2H**	**ALM3H**	**PNN**	**BTE**	**MVE**
SPM	84.15	80.69	78.81	88.54	88.18	90.36	88.03	89.32	88.09	86.92	90.18	83.9	89.2	89.18
PCA	89.78	84.06	86.18	78.48	80.86	88.14	89.05	84.24	86.44	89.14	91.06	80.51	90.14	86.11
CRA	91.17	89.68	88.58	90.54	90.59	92.77	90.74	91.66	90.89	91.47	93.22	89.9	**93.88**	90.65
F-measure with SMOTE Data														
SPM	0.79	0.68	0.75	0.79	0.78	0.77	0.77	0.79	0.77	0.77	0.78	0.76	0.79	0.75
PCA	0.75	0.85	0.56	0.48	0.78	0.76	0.74	0.76	0.76	0.74	0.81	0.76	0.83	0.88
CRA	0.84	0.79	0.74	0.77	0.77	0.88	0.82	0.83	0.85	0.81	0.86	0.87	0.91	0.89
AUC with SMOTE Data														
SPM	0.66	0.79	0.62	0.79	0.74	0.72	0.79	0.80	0.85	0.83	0.89	0.82	0.52	0.79
PCA	0.83	0.91	0.89	0.91	0.91	0.90	0.89	0.90	0.91	0.88	0.86	0.83	0.86	0.88
CRA	0.91	0.89	0.92	0.92	0.90	0.91	0.91	0.91	0.90	0.91	0.93	0.89	0.94	0.91

Table 4. Performance measures for SMOTE Data

The performance analyses for the SMOTE data is as shown in Table 4. The results illustrated the overall accuracies of the classifiers. The best performance for this particular dataset is 94% (CRA+BTE) for accuracy, 0.91 (CRA+BTE) for F-Measure and 0.94 (CRA+BTE) for AUC.

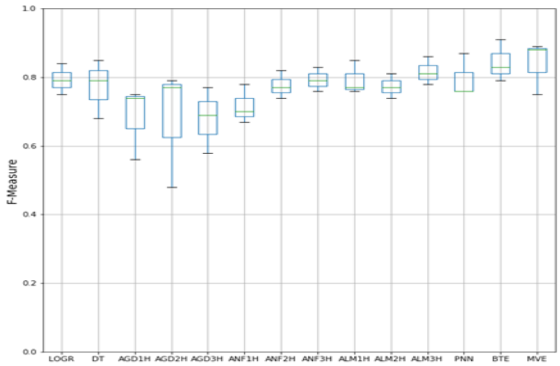

Fig. 4. F-Measure of different ML algorithms

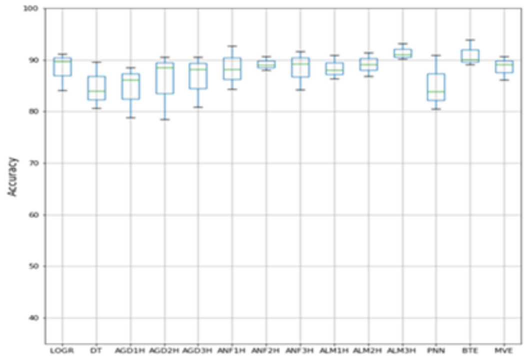

Fig. 5. Accuracy (%) of different ML algorithm

The performance comparisons between the different ML algorithms are displayed in box plots in Fig. 4 to Fig. 6, respectively. A box-and-whisker plot is a visual representation of a set of data on an interval scale. In exploratory data analysis, it is frequently utilised. It's a form of graph that depicts the shape of a distribution, as well as its central value and variability. The vertical scale's range is from the selected column's minimum to maximum value, or from the highest to lowest of the displayed reference points.

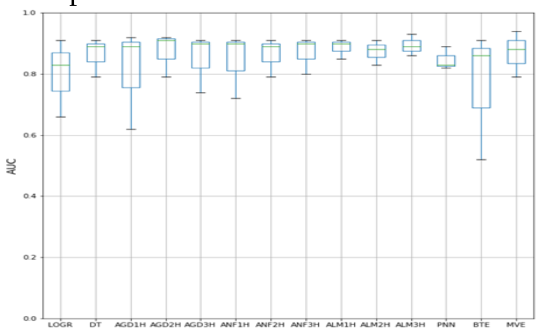

Fig. 6. AUC of different ML algorithms

5. Conclusion

Epilepsy is a severe cerebral disorder that can strike at any time and has severe consequences for the patient's ability to carry out day-to-day activities. For many experts, predicting and classifying this disease has been a difficult research challenge. Making decisions in the healthcare industry can be difficult at times. Medical decision-making classification systems allow medical data to be analyzed in less time and in more detail. As a result, in this study, we present the first-of-its-kind multi-level enhancement framework for epilepsy classification systems. To address the data imbalance problem, we used methods such as random sampling, down-sampling, and SMOTE. For selecting important features from the dataset, SPR, PCA, and CRA were adopted. After collecting the suitable set of features from the various feature selection methods, a novel and first-of-its-kind heterogeneous ensemble model was developed integrating LR, DT, ANN-(RBF), ANN-(LM), and PNN as the basic classifier.

To summarize, BTE and MVE ensembles were created using a strategic cluster of a total of 12 base classifiers. The results show that employing the SMOTE approach to optimize the dataset is essential; else, the model's performance would be low. When used in conjunction with other classifiers, the CRA and PCA algorithms perform better. The SMOTE sampled data with CRA features in conjunction with BTE ensemble learning obtains the maximum performance with 93.88% accuracy, 0.91 F-Measure, and 0.94 AUC, according to the overall performance assessment. In future we will compare the proposed model to the Extreme Learning Machine (ELM) and Least Square SVM (LSSVM) algorithms, as well as create an automated system that will aid physicians in making better and more accurate decisions for their patients.

Acknowledgment

We are grateful for the support given to this research by Visvesvaraya Technological University, Jnana Sangama, Belagavi. We'd also like to thank the PA College of Engineering in Mangalore for providing all of the resources we needed to complete this task.

References

[1]. K. M. Tsiouris, V. C. Pezoulas, M. Zervakis, S. Konitsiotis, D. D. Koutsouris, and D. I. Fotiadis, “A Long Short-Term Memory deep learning network for the prediction of epileptic seizures using EEG signals”, Computers in Biology and Medicine, pp. 24-37, 2018.

[2]. S. K. Satapathy, A. K. Jagadev, and S. Dehuri, “Weighted Majority Voting Based Ensemble of Classifiers Using Different Machine Learning Techniques for Classification of EEG Signal to Detect Epileptic Seizure”, Informatica, pp. 99–110, 2017.

[3]. X. Yao, X. Li, Q. Ye, Y. Huang, Q. Cheng, and G-Q Zhang, “A robust deep learning approach for automatic classification of seizures against non-seizures”, Biomedical Signal Processing and Control, 2020.

[4]. W. Zhou, Y. Liu, Q. Yuan, and X. Li, “Epileptic seizure detection using lacunarity and Bayesian linear discriminant analysis in intracranial EEG”, IEEE Transaction on Biomedical Engineering, pp. 3375–3381, 2013.

[5]. A.H. Ansari, P.J. Cherian, A. Caicedo, G. Naulaers, M. Vos, and S.V. Huffel, “Neonatal seizure detection using deep convolutional neural networks”, International Journal of Neural Systems, Vol. 28, 2018.

[6]. Y. Yuan, G. Xun, F. Ma, Q. Suo, H. Xue, K. Jia, and A. Zhang, “A novel channel-aware attention framework for multi-channel EEG seizure detection via multi-view deep learning”, IEEE Proceedings of the International Conference on Biomedical & Health Informatics, pp.206–209, 2018.

[7]. Y. Yuan, G. Xun, K. Jia, A. Zhang, “A multi-view deep learning framework for EEG seizure detection”, IEEE Journal on Biomedical and Health Informatics, pp.83–94, 2018.

[8] S. Amin, and A.M. Kamboh, “A robust approach towards epileptic seizure detection”, Proceedings of 26th International IEEE Workshop on Machine Learning for Signal Processing pp. 1–6, 2016.

[10] B. Hunyadi, M. Signoretto, W.V. Paesschen, J.A.K. Suykens, S.V. Huffel, and M.D. Vos, “Incorporating structural information from the multichannel EEG improves patient specific seizure detection”, Clinical Neurophysiology, pp.2352–2361, 2012.
[12] N.D. Truong, L. Kuhlmann, M.R. Bonyadi, J. Yang, “Supervised learning in automatic channel selection for epileptic seizure detection”, Expert Systems with Applications. pp. 199–207, 2017.
[13]. A. O’Shea, G. Lightbody, G. Boylan, and A. Temko, “Neonatal seizure detection using convolutional neural networks”, IEEE International Workshop on Machine Learning for Signal Processing, 2017.
[14]. A. Antoniades, L. Spyrou, C.C. Took, and S. Sanei, “Deep learning for epileptic intracranial EEG data”. 26th IEEE International Workshop on Machine Learning for Signal Processing, pp. 1-6, 2016.
[15] RG. Andrzejak, K. Lehnertz, C. Rieke, F. Mormann, P. David and C.E. Elger, “Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state”, Phys. Rev. E, 64, 061907, 2001.
[16] F. Duman, N. Ozdemir, and E. Yildirim, “Patient specific seizure prediction algorithm using Hilbert- Huang Transform”, IEEE- EMBS International Conference in Biomedical and Health Informatics, pp. 705–708, 2012.
[17] IT. Jolliffe, J. Cadima, “Principal component analysis: a review and recent developments”, Philosophical Transactions, The Royal Soceity Publishing, 2016.
[18] P. Ghaderyan, A. Abbasi and M. Sedaaghi, "An efficient seizure prediction method using KNN-based undersampling and linear frequency measures," Journal of Neuroscience Methods, 2014.
[19] A. Subasi and E. Elebi, “Classification of EEG signals using neural network and logistic regression,” Computer Methods and Programs in Biomedicine, pp. 87–99, 2005.
[20] P. Swami, T. K. Gandhi, B. K. Panigrahi, M. Tripathi, and S. Anand, “A novel robust diagnostic model to detect seizures in electroencephalography,” Expert Systems with Applications, vol. 56, pp. 116–130, 2016.
[21] J. H. Kang, Y. G. Chung, and S. P. Kim, “An efficient detection of epileptic seizure by differentiation and spectral analysis of electroencephalograms”, Computational Biology, and Health Informatics 66:352–356, 2015.
[22] Y. Nesibe, T. Gulay, and K. Cihan, “Epilepsy diagnosis using artificial neural network learned by PSO”. Turkish Journal of Electrical Engineering and Computer Science, pp. 421–432, 2015.
[23] Seyyed Abed Hosseini, “Nonlinear Analysis of EEG Dynamics in Different Epilepsy States Using Lagged PoincarE Maps ", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.10, No.8, pp. 61-67, 2018.DOI: 10.5815/ijigsp.2018.08.07.
[24] Anup Kumar Keshri, Shamba Chatterjee, Barda Nand Das, Rakesh Kumar Sinha, "WEBspike: A New Proposition of Deterministic Finite Automata and Parallel Algorithm Based Web Application for EEG Spike Recognition", International Journal of Information Technology and Computer Science(IJITCS), vol.5, no.11, pp.93-102, 2013. DOI: 10.5815/ijitcs.2013.11.10.
[25] Sevda G, Muhammed K. U, G Cetinel, Erhan B, Mehmet R. B., "Automated Pre-Seizure Detection for Epileptic Patients Using Machine Learning Methods", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.9, No.7, pp.1-9, 2017.DOI: 10.5815/ijigsp.2017.07.01.

Biographies of Authors

Mohammad Asif A Raibag received the B.E (CSE) degree from TCE-Gadag and M.Tech (CSE) degree from BVBCET-Hubli in 2004 and 2006 respectively. He has 16 years of teaching experience at college level. He has worked as faculty in the Department of CSE at REC-Hulkoti, BVBCET-Hubli, HMSIT-Tumkur and GAT-Bangalore. Currently he is serving in PA College of Engineering in CSE department as Associate Professor and pursuing research at VTU, Belagavi.

Dr. Vijay Franklin received the B.E(CSE) degree from BIT-Sathy, M.Tech(CSE) degree from BVBCET-Hubli and PhD from Anna University Chennai in 2004, 2006 and 2014 respectively. He has teaching experience of more than 16 years at college level. He has worked at different levels at BIT-Sathy. Currently he is Professor CSE,BIT-Sathy. He has more than 39 publications in reputed intenational journals and conferences. His expertise is in Cognitive Engineering, Soft Computing, Optimization and 4G Networks.

Dr. Rashal Sarkar received the B.E(CSE) degree from B.P.U.T, Orissa, M.Tech(CSE) degree from R.V College of Engineering, Bangalore and PhD from Himalayan Garhwal University, Uttarakhand. He has teaching & administrative experience of more than 15 years at college & University level. Currently he is Associate Professor in department of Computer Science & Electronics, University of Science and Technology, Meghalaya. He has more than 10 publications in reputed intenational journals and conferences. His expertise is in Data Mining, Data Science, Machine Learning, and Artificial Intelligent.