Enhanced Epileptic Seizure Detection using Imbalanced Classification

Prabhsimar Kaur, Vishal Bharti, Srabanti Maji

Abstract: ***Epilepsy is the second most persistent neurological condition, endangering the lives of patients. Though there have been many advancements in neurological imaging approaches, the Electroencephalogram (EEG) still remains to be the most effective tool for testing and diagnosing epileptic patients. The visual analytics of EEG signals is a very prolonged process and always open to the subjective judgment of the physicians. The main goal of our study is to build an automatic classifier that can analyze and detect epilepsy from EEG recordings obtained from epileptic and healthy patients, thus helping the neurosurgeons to diagnose epilepsy in a better way. Synthetic minority oversampling technique (SMOTE) has been used for balancing the EEG dataset and the Principal component analysis (PCA) technique is applied further, for reducing the EEG signal dimensionality. For data classification, seven machine learning classifiers have been used and after comparing the results the authors conclude that Artificial Neural Network (ANN), outperforms the other classifiers by providing an accuracy of 97.82%.***

Keywords: ***epilepsy detection, electroencephalogram, oversampling, class imbalance, dimension reduction, optimized parameter.***

I. INTRODUCTION

Epilepsy, also called the seizures are the second most persistent neurological state which is caused by irregular discharges of electricity in the brain. It may affect during any stage of life but is a more common state in infants and elderly people [1]. The worst thing is that epilepsy may effect at any stage of life without giving any prior health alerts, which sometimes may even lead to death. There are mainly two types of epilepsy or seizure which are briefly described as following:

- **Focal Seizures:** Also known as a partial seizure derives its origin from the paroxysmal discharge from the very discrete region of the cortex and then spreads to the neighboring areas. Simple, complex, and focal seizures with secondary generalization are the types of focal seizures. It mainly causes emotional disturbances, spasms, head-turning, muscle rigidity, unusual sensations affecting either the taste, touch or hearing, etc [2].
- **Generalized Seizures:** These type of seizures derive their origin from the diencephalic activation system in the brain which then spreads across to different areas in the brain. Generalized tonic-clonic, absence, atonic, myoclonic, etc. are some types of generalized seizures. Their main symptoms are unconsciousness, jerking movements, muscle stiffness, sporadic, rigidity, etc.

Epilepsy disorders can be managed in some patients using Anti-Epileptic Drug therapy (AED), but others who are untreated are at risk of sudden death at any part of their lives. If the medications are not successful, the patients can undergo brain surgery but it has various adverse effects. Therefore, before recommending for brain surgery, the patients should be analyzed properly through EEG.

The Electroencephalogram (EEG) test is the most widely and helpful tool being used by neurosurgeons for diagnosing neurological disorders. Data acquired by the EEG test helps the neurosurgeons to analyze and understand the current state of the brain. Generally, invasive & non-invasive are two categories of EEG. Invasive EEG recordings are taken by placing the electrodes in the cortex for a long duration of time, thus can damage and lead to infections in the brain. However, non-invasive recordings are taken by placing the electrodes to the scalp surface, thus causing no damage to the brain [3]. In our dataset, non-invasive EEG recordings are being used for the study.

A. Problem Statement

Various challenges for Epilepsy detection have been reported by the researchers [4]-[6]. Out of all the reported challenges, one of the most important challenges is the problem of class imbalance. One can define the problem of class imbalance as an exceedingly imbalanced and highly distorted data distribution [7]. The data used in our study is highly imbalanced as the ratio of Epileptic patients is significantly smaller than the healthy patients. The dataset taken in our work has the prevalence count of 20% i.e. the proportion of the epileptic patients is 20% out of all the population taken in the experiment. Fig 1. represents the class imbalance situation of the dataset taken in our work. The epileptic patient class is represented by the red dots which have very little frequency than the healthy patient class.

Revised Manuscript Received on May 21, 2020.

***** Prabhsimar Kaur

Prabhsimar Kaur*, Department of Computer Science and Engineering, DIT University, Dehardun, India. Email: pskdara594@gmail.com

Vishal Bharti, Department of Computer Science and Engineering, DIT University, Dehardun, India. Email: hod.cse@dituniversity.edu.in

Srabanti Maji, Department of Computer Science and Engineering, DIT University, Dehardun, India. Email: srabanti.maji@dituniversity.edu.in

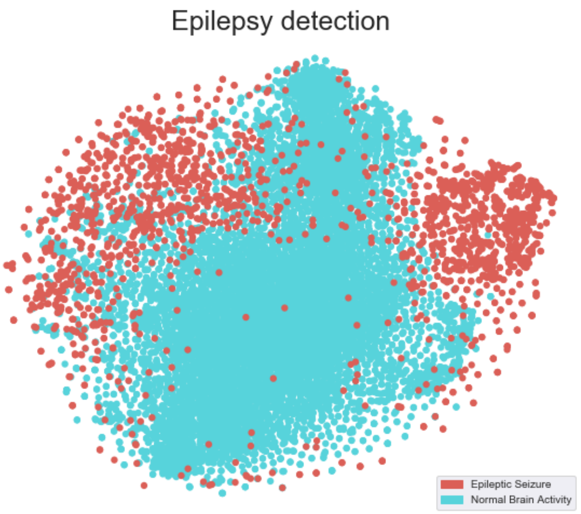

Fig 1. Scatter plot of the epilepsy dataset.

The class imbalance poses an enormous challenge in identifying the features of epileptic patients. The prediction methods used are governed by the majority class, and hence, they misclassify observations of the minority class. The objective is to properly classify the epileptic patients, the classification algorithms should properly detect epileptic patients from healthy patients. If the classification algorithms are not applied properly, the whole purpose shall be defeated as the epileptic patients will be tagged as healthy and vice-versa, thus risking the life of the individual.

Many researchers [8],[9] have worked on the class imbalance challenge, proposing various solutions based on machine learning algorithms. But still, the data imbalance issue remains unresolved. The main idea behind our study is to analyze and depict limitations of existing machine learning approaches that are being used in solving the data imbalance problem and to propose a highly efficient solution to resolve this issue.

B. Our Contribution

Through this paper, we have done a thorough analysis and through statistical experimentation compared the performance of various machine learning classifiers that deal with the problem of improper class balancing. We have made an effort in identifying the limitation of each classifier and help develop a solution mechanism that shall help to solve the class imbalance problem in the real world situation. Our specific contributions in the paper have been listed below:

- We have done a comparison of seven machine learning classifiers, detailed in the literature, used for prediction and detection of epilepsy.
- The selection of the most efficient classifier based on five performance measures has been made. We have then compared them with the class imbalance problems applying the same algorithms.
- The comparison being drawn amongst the results obtained from the same classifiers when applied on both balanced and imbalanced set helps us to understand the drawbacks of the classifiers.
- Further, the PCA technique for dimension reduction has been applied on balanced data set and the comparison has been drawn.
- Finally, the findings of our experimentation and drawbacks of the experimented approaches have been reported.

The organization of our work can be summarized into the following sections. Section I gives a summarized introduction along with a problem statement & contribution. Section II provides us with the literature review done on epilepsy detection and class imbalanced problems. Section III provides us with a short introduction to the seven machine learning classifiers which are being used in the detection of epilepsy. Section IV, experimental design along with data set description has been provided. Section V discusses the derived results and their analysis. In the end, the conclusion & future work have been briefly explained.

C. Related Work

The section reports research work revolving around the detection & prediction of epilepsy. In addition, we have concentrated on the study which addresses the issue of class imbalance in the detection of epilepsy.

- Epilepsy Detection and Prediction

Gotman [10] developed and proposed the model for efficient use of Electroencephalogram signals three decades ago by applying different statistical and analytical techniques for prediction and detection of seizure. In addition, many researchers [11],[12] have carried out their research in the field of EEG signal to deliver better results

Data mining is widely used to discover critical patterns from various types of datasets that help solve real-world situations [13]. However, selecting appropriate features in the EEG dataset is one of the challenges as EEG signals are non-stationary, difficult to understand, and very time-consuming. This section deals with the literature work related to various feature extraction techniques & different machine learning classifiers used in prediction & detection of seizure. Various transformation techniques such as continuous wavelet transformation (CWT), time and frequency domain, (DWT) Discrete wavelet transformation and, (FT) Fourier transformation are used for extracting important features though EEG data sets [14],[15]. ‘Line length’ and ‘relative power’ are the two features which have been identified as successful seizure detection performer [16]. The reported output from feature extraction techniques has been evaluated using various performance measures like specificity, precision, recall, receiver operating characteristic (ROC) curve, and F-score. Many researchers conducted the task of seizure detection by extracting and applying only one feature and it is noted that a single feature can be considered as sufficient to provide accurate performance [17]. Authors in [18] have proposed a fully automated system for detection of seizures using a single ‘line length’ feature giving 84.27 % accuracy, 85.70 % specificity and, 84% recall. In [19] authors used an artificial neural network (ANN) with ‘a line length’ feature to build an automated seizure detection model which gave an accuracy of 99.6%. Another single feature used by many researchers is the ‘entropy’ [20], [21]. Authors in [21] used sample entropy features with extreme learning machine (ELM) to detect seizure with good accuracy.

Further, several studies focus on using various machine learning classifiers for prediction and detection of seizure. The two classifiers – SVM and ANN are widely used for seizure detection when dealing with large data sets. Shoeb and Guttag [22] used SVM with vector feature for seizure detection using CHB-MIT (CHB) data set and reported accuracy of 96%. In addition, authors in [15] used the DWT technique to extract relative energy features from the EEG data set and compared the result using four machine learning classifiers- k-nearest neighbor (KNN), SVM, multilayer perceptron (MLP) & Naïve Bayes (NB). Out of four classifiers, SVM gave better results with an accuracy of 98.75%. The authors [23] used RF classifiers on the features extracted using the quadratic feature extraction (QFE) method on the dataset taken from the university of BONN, Germany, and reported accuracy of 97.35%. Another work carried out by authors [24] build the bootstrap samples to apply on four classifiers- SVM, ENN, MLP, and KNN to give an accuracy of 97%.

- Class Imbalance Epilepsy Classification

In the previous section, we have reported the literature review related to two challenges - selection of statistical features and various performance of machine learning classifiers that have been used for prediction & detection of seizure. Though each classifier is having its own set of advantages & disadvantages, various performance measures have been used in order to evaluate the performances of classifiers. Another main challenge seen in clinical data sets is the problem of class imbalance. Class imbalance is defined as having more observations in majority class making it difficult for the classifier in the detection of the minority class. The researchers in [9] have discussed this problem in their work and have provided a suitable solution to solve class imbalance problems while detecting the seizure. In addition, some researchers have reported a suitable threshold value when assigning the heavyweight to a minority class. The two methods used for enhancing the performance of machine learning classifiers over class imbalanced problems are over-sampling and under-sampling [25],[26]. The author's Li and Wen [27] presents the concept of random sampling for extracting the features from EEG signals. They also described the implementation of (LS SVM) least square support vector machine approach, having the potential of classifying the EEG signals. The results of the experiment provide us with the classifying accuracy of the training data set and test data set, which gives an accuracy of 80.31% and 80.05% respectively.

II. CLASSIFICATION ALGORITHMS FOR SEIZURE DETECTION

Different procedures have been deployed for solving the problem of class inequality. The procedures and solutions are being discussed briefly in this part. Before detailing the solutions for class imbalance, we are summarizing the various machine learning classifiers which are usually being used for epilepsy detection. It must be noted that machine learning classifiers are an important component that is being widely being used for tackling the solutions of class imbalance.

A. Machine Learning (ML) Classifiers

In this part of the paper, we have detailed the machine learning classifiers which are being used for epilepsy detection, which have been tested & analyzed.

- Naïve Bayes (NB)

For the purpose of classification Naïve Bayes makes use of the Bayes condition rule of probability [28]. This approach involves seeking a class to the new instance which results in an increase of probability when the variable values are given. The main aim of the NB classifier is to find the appropriate value Y which results in the maximum value of $P(Y/X_1, X_2, \ldots\ldots, X_n)$.

Using the Bayes Theorem,

$$P(Y/X_1, X_2, \ldots\ldots, X_n) = \frac{P(X_1, X_2, \ldots\ldots, X_n/Y)P(Y)}{P(X_1, X_2, \ldots\ldots, X_n)} \quad (1)$$

Maximizing $P(Y/X_1, X_2, \ldots\ldots, X_n)$ implies maximizing $P(X_1, X_2, \ldots\ldots, X_n/Y)$.

From the historical data this can be easily calculated, assuming class conditional equality between variables:

$$P\left(X_1, X_2, \ldots\ldots, \frac{X_n}{Y}\right) = P\left(\frac{X_1}{Y}\right) P\left(\frac{X_2}{Y}\right) P\left(\frac{X_3}{Y}\right) \ldots\ldots P\left(\frac{X_n}{Y}\right) \quad (2)$$

The expectation is not always fulfilled. The democratization of the continuous variables is another disadvantage of this process. This means that some knowledge may be lost, or that it is believed that these variables are distributed roughly normally which may not be accurate.

- Support Vector Machine (SVM)

This classifier divides the given dataset by a line called hyper-plane into two different classes. SVM gives better performance when used to solve classification problems. Each data point is represented as a single instance in n-dimensional space where each feature value reflects the coordinates of the plane. The number of exploratory variables is defined by a training vector x_i in R^n, i = 1, 2,……, L, n, and the number of instances in the training set is given by L. The binary classification is achieved by addressing the following issue of optimization.

$$\text{Minimize } \frac{1}{2}\,||w||^2 + C\sum_{i=1}^{L} \epsilon_i \quad (3)$$

Subject to

$$\{y_i * (w^T\varphi(x_i) + b) \geq 1 - \epsilon_i, \epsilon_i \geq 0, i = 1,2, \ldots\ldots, L\} \quad (4)$$

The equation for the hyper-plane is given as $w^T\varphi(x_i) + b$, where $\varphi(x_i)$ maps x_i to higher-dimensional spot and w is the weight vector. Slack variables ϵ_i are added to accommodate certain errors in the event that these points cannot be linearly separated [29]. C is known as the cost parameter and should always be > 0. The goal of minimizing $\frac{1}{2}\,||w||^2$ is to maximize the distance among two margins in order to find out the best hyper-plane which best separates the two different classes.

- Logistic Regression (LR)

LR is a generalized type of linear model that is widely used in statics field. This classifier is being named on a function called the logistic function [30].

The model coefficients are denoted by the vector $\alpha = (\alpha 0, \alpha 1, \dots, \alpha n)$, exploratory variables are denoted by $Z = (1, Z_1, Z_2, \dots, Z_n)$ and model's error is denoted by $\in$. The linear model shall be described as:

$$Y = \alpha_0 + \alpha_1 Z_1 + \alpha_2 Z_2 + \dots + \alpha_n Z_n + \varepsilon \quad (5)$$

A logit function denoted by g is implemented in LR that restricts the linear combination of variables to maintain values between range [0, 1]. $g(p) = X\alpha$, where p denotes the probability of epileptic patients who are dealing with it. The logit function is written as:

$$g(p) = \ln \frac{p}{1-p} \; with \; p = \frac{e^{Z\alpha}}{1+e^{Z\alpha}} \quad (6)$$

- K-Nearest Neighbor (KNN)

This classifier is based on instance-based learning. The instances that are near are known as "neighbor". For each new point, the distance from each case is measured and according to nearest neighbors, they are classified into groups. The classifier consistently computes the value for a target until there is no change in the class of an instance [31]. Normally the Euclidean distance is used to calculate the distance amongst two points says a and b which is given is:

$$d(a, b) = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2} \quad (7)$$

- Decision Tree (DT)

DT is a simpler, easy to implement classifier which requires no domain knowledge. The main advantage is that the DT approach can be extended to enormous data that is to be interpreted. DT uses a tree hierarchy to resolve a problem composed of nodes, arcs, and branches. The arcs are connected from none to a node. The branch consists of various attributes, an internal node has a test for which the attribute is used, and the leaf node consists of the classes predicted from the DT to build a suitable decision [32].

- Random forest (RF)

RF is also called as Random Decision Forest (RDF), used to solve both regression and classification problems. Its structure is like a forest that contains multiple numbers of decision trees and each single decision tree gives a unit value for most popular (x) input class. The decision tree with more value is chosen for the classification [33](Liaw and Wiener, 2002). The RF classifier is not biased, as there are several trees and each tree is trained on a data subset, thus it can work well when we have both numerical and categorical features.

- Artificial neural network (ANN)

ANN is a data processing classifier that had its base as neurons and is used mostly for predicting diseases. Its structure consists of multiple numbers of tiny processors to help with doing data processing and to tackle any issue. ANN is general capacity approximations that are why they can be connected to any machine learning issue. ANN mainly comprises three different layers namely – input, output, and hidden layer. The input layer includes the explanatory variables representing input nodes. The inputs are then being multiplied with a common weight and further passed on to each node of the hidden layer where certain bias is applied to them. Further after summation, an activation function is implemented for generating the output of neuron which is then passed on to the next layer. Lastly, the final result is generated by an output layer [34].

B. Imbalanced Classification

There are two methods used in imbalanced approaches to classification. The primary technique includes under-sampling, oversampling, etc. is used on data as an initial processing step for balancing classes. The secondary technique includes cost-sensitive learning which is used within the machine learning classification algorithms.

- Sampling

There are two ways in which sampling can be done: over-sampling and under-sampling [35]. The class value, which is in the majority, is shrunk in the under-sampling process. It means that a group of records is randomly omitted from the training dataset. The drawback of the under-sampling process is that while balancing the dataset, it removes some of the significant records. Easy ensemble and random under-sampling are the two techniques commonly used to balance the data set using the under-sampling method.

However, in the over-sampling process, the minority value of the class is increased to be in consonance with the majority value of the class. The increments of the records are done randomly without replacement. The drawback of the over-sampling process is an increase in computational cost because of redundancy and inconsistency in results. Random oversampling and SMOTE are the two techniques commonly used for balancing the data set through the over-sampling approach.

In our work, we have used the SMOTE technique to balance our data set. This technique is helpful in solving the over-fitting problem which is caused by using the random over-sampling method. This approach selects random samples from the minor categories and produces new random samples amongst selected samples & neighboring samples from similar categories. This results in the generation of new samples between samples of the minor categories but without overlapping original samples.

- Cost-sensitive learning

It is an important and widely used technique of generating models for highly imbalanced data sets. This approach is based on a confusion matrix and the main idea behind this approach is assigning a higher weight to the minority class. The cost-method which is related to the confusion matrix is used to adjust the weights. The cost of the matrix depends on four measures which are-

- True Positive
- True Negative
- False Positive
- False Negative

C. Dimension Reduction

The hyper-spectral representation space can be significantly reduced using the approach of feature selection and feature extraction. The goal behind both methods is the reduction in the number of features without much information loss. The technique of selection is such that the features representing the subset are picked up from amongst the original data set by first analyzing their capacity for differentiation, based upon a statistical distance measurement between classifiers.

The method of feature extraction helps in solving the problem of dimension reduction through data projection from amongst the original feature space to a low-dimensional subspace which consists mostly of the original information [36]. The Principal Component Analysis (PCA) is the most widely known approach for feature extraction which aims in the reduction of data dimension by finding certain orthogonal linear combinations of the original variables having the greatest variance. PCA is a versatile data analytical tool. The advantage of PCA is that once the trend of data is found, then it becomes very easy to reduce the data by reduction of the number of dimensions without having much information loss.

III. EXPERIMENTAL APPROACH

In this part, we have reported our experimental analysis carried out through using seven machine learning classifiers, followed by approaches for solving class imbalance problem and dimension reduction.

A. Design of Experiment

This part deals with the experimental workflow. The model that we propose for our work in given in Fig 2.

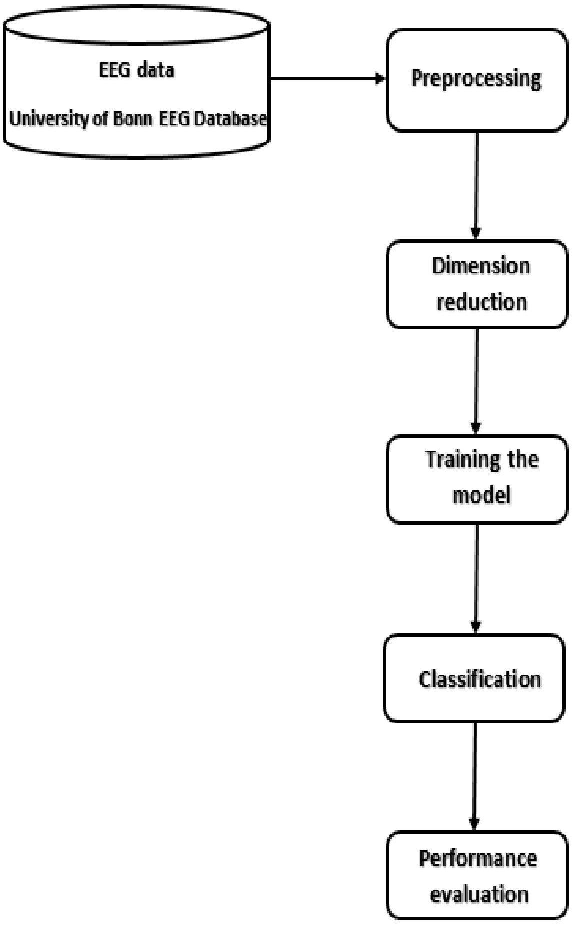

Fig. 2. Proposed model for epileptic seizure detection

- Workflow of experiments

Python language has been used to do experimental work. The experimental work is segmented into three phases. In the first phase, the comparison amongst seven machine learning classifiers has been drawn using the original data set. The comparison has been made in relation to five performance measures: accuracy, recall, precision F1-score, and AUC. The comparison leads to the selection of the most appropriate classifier.

In the second phase, the SMOTE technique is used for solving the class imbalance problem. Again the comparison has been drawn using the same seven classifiers but with balanced data set.

In the third phase, we used PCA for dimension reduction on a balanced dataset, and finally the comparison has been drawn using the same seven classifiers.

B. Dataset Used

The dataset taken for our work is an EEG database recorded in the Epileptology Department from Bonn University, Germany [37]. Original dataset consists of five different folders, each folder consists of one hundred files, where each file represents a single patient. Each data file is recorded for 23.6 seconds with a 128-channel 12-bit EEG recording system. Corresponding time series is being sampled into 4097 data points. All the data points are recorded at different point of time having specific EEG values. The 4097 data points is divided into 23 chunks where each chunk consists of 178 data points for 1 second and the last column contains the feature y. The possible values of y lie between 1-5 where the number 1 represents the seizure activity, the number 2 represents the area where the tumor was detected, number 3 indicates the region where the tumor is present, also it gives the EEG recording of the unhealthy brain, number 4 gives the EEG recordings when the patients eyes are closed and number 5 represents the EEG signals where patient's eyes are open. The patients in classes 2-5 did not suffer from epileptic seizure whereas the number 1 class patients only suffer from seizure activity. Thus we use binary classification for programming part i.e. classes 2-5 are taken as 0.

In order to use this dataset in our work, we have restructured it by dropping the first column of patient id as it has no relevance for detecting epilepsy and transforming the target feature y to 'OUTPUT_LABEL' feature with binary values 0 or 1 using binary classification, where 0 represents a healthy patient and 1 represents an epileptic patient.

In the dataset, 2300 (20%) are epileptic patients and 9200 (80%) are healthy individuals. These data show the extreme problem of class imbalance with 20% epileptic patients. The representation of the dataset representing healthy and epileptic class is shown is Fig. 3. To build the models, data into a training set (70%) and a test set (30%) to analyze their performance. Apart from training/testing, we have made use of 10-fold cross-validation to find out the accuracy which is the most common way to validate the data by breaking the training data into 10 miniature train/test splits.

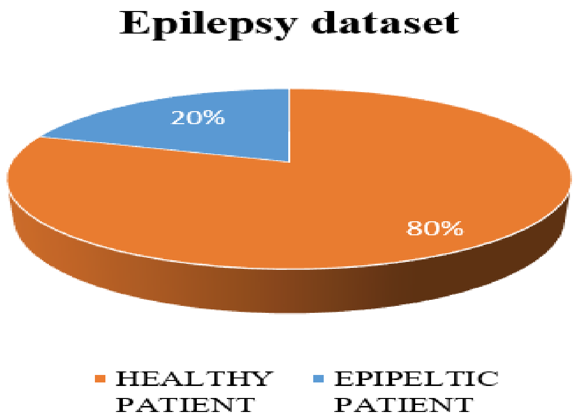

Fig. 3. Illustration of a dataset representing a healthy and epileptic class

C. Performance Measures

In the classification problem, "accuracy" is usually the most commonly used parameter in order to measure the output. In our experimentation accuracy has not been considered to be an important parameter, as we are more focused on tackling the class imbalance issue. Accuracy considered alone as a performance measuring parameter can distort the outcome of epilepsy detection. Hence, we have selected a few other performance measures also for evaluating our classifiers. The class percentage in our data is 20% of epileptic class and 80% of the healthy class, which means any classifier giving an accuracy of less than 94% has not been accepted. Epilepsy cannot be detected properly by attaining a higher rate of accuracy. Sensitivity or recall, which is correctly categorized minority class findings, is most important for accessing the output to best of our knowledge. Thus, along with accuracy we have also selected sensitivity in order to compute the performance. Except for these two performance measures, precision and F1-score measure are also important in evaluating the classifiers. A confusion matrix shown in Table-I is evaluated using the test set to calculate those measures.

Table-I: Confusion matrix

Predicted	**Actual**	
	Epileptic (1)	**Healthy (0)**
Epileptic (1)	True Negative (TN)	False Positive (FP)
Healthy (0)	False Negative (FN)	True Positive (TP)

Accuracy is given as the ratio of predictions that are correct i.e. accurate to the total number of predictions made.

$$Accuracy = \frac{TN+TP}{total\ number\ of\ test\ observations} \quad (8)$$

Recall or sensitivity, also known as the ratio of positive data points which are actually positive to the sum of all positive data points. It is also referred to as TPR that stands for True Positive Rate.

$$Recall = \frac{TP}{total\ number\ of\ actual\ observations\ concerning\ epilepsy} \quad (9)$$

Precision is known as the ratio of predictions correct to the sum of all predictions positive.

$$Precision = \frac{TP}{total\ number\ of\ observations\ concerning\ epilepsy} \quad (10)$$

F1-score, also known as F measure. It is defined as the harmonic mean between precision and recall.

$$F1 - score = 2 * \frac{(Precision * Recall)}{(Precision + Recall)} \quad (11)$$

Another commonly used performance measure for classifiers is the Receiver Operating Characteristic (ROC) curve. It is a graphical plotting, depicting the diagnostically potential of the binary classifier's model as the threshold limit for discrimination is varied. The curve is built by plotting the False Positive Rate (FPR) over the TPR. The False Positive Rate (FPR) is defined as follows:

$$False\ Positive\ Rate = \frac{FP}{total\ number\ of\ actual\ healthy\ patients} \quad (12)$$

In the ROC plot, the model performance or quality is measured with Area within the ROC Curve (AUC). It should be remembered that each of the above-mentioned parameters is interpreted differently. Furthermore, no single performance measuring parameter can be singly used to validate the competitive efficiency of the algorithms which we have experimented within our work.

IV. RESULT AND DISCUSSION

In this paper, we have performed experimentation on seven machine learning classifiers mentioned in Section III. The results of our experimentation have been depicted in a tabular form. Table-II denotes the outcome of the classifiers implemented upon the original dataset. Table-III denotes the outcome of the classifiers implemented upon the balanced data set. Table-IV denotes the outcome of the classifiers with the dimension reduction technique implemented upon the balanced data set.

According to the results obtained, we observe that leaving aside the Logistic Regression classifier, all other classifiers provide an accuracy greater than 90%. Through our analysis, we conclude that when compared with other classifiers, SVM, ANN, and RF provides the most appropriate results. The reason for such results are two-fold (a) when compared with other classifiers SVM, ANN and RF provide better outcomes in all the five performance parameters. (accuracy, precision, recall, F1-score, and AUC) (b) the three algorithms mentioned are not the constraint of any statistical or mathematical assumption.

Table-II: Comparisons of classifiers when applied to the original data set.

Classifier	Accuracy	Precision	Recall	F1-score	AUC
KNN	90.03	88.50	57.5	66.7	0.712
NB	91.54	89.22	80.71	77.94	0.752
RF	94.45	96.91	77.21	85.99	0.806
DT	93.51	86.95	71.24	70.75	0.694
LR	76.96	78.52	34.02	59.02	0.502
ANN	**95.32**	**87.63**	**88.71**	**90.27**	**0.894**
SVM	94.80	88.81	86.43	87.72	0.890

Table-III: Comparison of classifiers when applied to the balanced data set.

Classifier	Accuracy	Precision	Recall	F1-score	AUC
KNN	92.09	90.50	62.51	76.79	0.850
NB	93.54	92.22	88.71	89.94	0.831
RF	94.12	88.23	95.51	91.07	0.920
DT	93.54	89.90	74.24	77.75	0.706
LR	80.33	68.56	39.78	62.90	0.512
ANN	**96.02**	**86.20**	**94.64**	**90.67**	**0.937**
SVM	95.43	85.18	92.07	88.61	0.846

Table-IV: Comparison of classifiers with PCA technique when applied to the balanced data set.

Classifier	Accuracy	Precision	Recall	F1-score	AUC
KNN	96.02	97.92	81.91	89.20	0.901
NB	93.95	88.32	93.05	90.62	0.862
RF	96.31	89.76	96.56	92.77	0.988
DT	94.20	90.21	80.78	82.78	0.868
LR	81.96	75.45	47.61	79.68	0.532
ANN	**97.82**	**97.27**	**96.96**	**95.91**	**0.996**
SVM	97.68	92.25	96.52	94.34	0.996

The information about our experimentation and selection of algorithms are detailed below. LR classifier did not provide us with good results as this classifier is based upon the assumption & condition of variables. While performing experimentation with KNN, data is first standardized and normalized using the minimum-maximum scale. Normalization is of the utmost importance when the variables are in different ranges and scales. If the algorithms are not normalized then the bias shall shift towards the larger scales. To quote an example, in our case although it might not be of utmost importance, still the bias shall shift towards the balance variable. Three nearest neighbors are taken into consideration and Euclidean distance is used in the classification. This method's simplicity doesn't allow for effective results as shown in Tables – II, III and IV. DT and NB give a good value of recall, yet this recall is accomplished by a rising number of false epileptic cases. In fact, DT leads to over-fitting and NB has its presumption of conditional independence (CI) are the other big drawbacks for not choosing these as best models.

RF classifier has an inbuilt method named as weighted random forest (WRF), for balancing the data set. It also has the ability to estimate missing data and can also estimate the importance of variables. Thus, RF is considered as one of the better models in our work.

While experimenting with SVM, the number of false negatives was reduced as we applied classifiers on balanced data set along with PCA, this resulted in an increase in accuracy and other performance measures. SVM requires an enormous amount of data for finding out the best separating hyper-plane, which was available in our case. Thus SVM is selected as one of the best models in our work. For kernel selection, we have made a comparison among four kernels- rbf, ploy, sigmoid, and linear. Out of these, 'rbf' kernel gave the best result thus it has been chosen for performing the classification.

In the case of ANN, the network is built up of seven nodes in the input layer, there are three nodes in the hidden layer and an output layer. A sigmoid activation function is being used in the experimentation. Resilient Back Propagation (Rprop) algorithm is used for adjusting weights. ANN provided us with good results and has the ability to solve classification problems and can easily learn from complex relationships. Confusion matrices for ANN are given in Table-V, Table-VI, and Table-VII.

Table-V: For an original data set

Actual	Predicted	
	0	1
0	2735	24
1	78	613

Table-VI: For a balanced data set

Actual	Predicted	
	0	1
0	2647	112
1	25	666

Table-VII: For a balanced dataset with PCA

Actual	Predicted	
	0	1
0	2705	54
1	21	670

The ROC curve is plotted between TPR (x-axis) and FPR (y-axis). ROC graph for ANN classifier using original data set, balanced data set, and using PCA on the balanced dataset is given in Fig. 4 to Fig. 6.

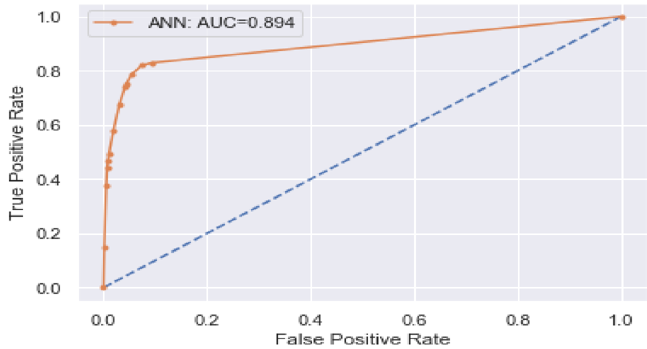

Fig. 4. AUC for ANN using the original data set

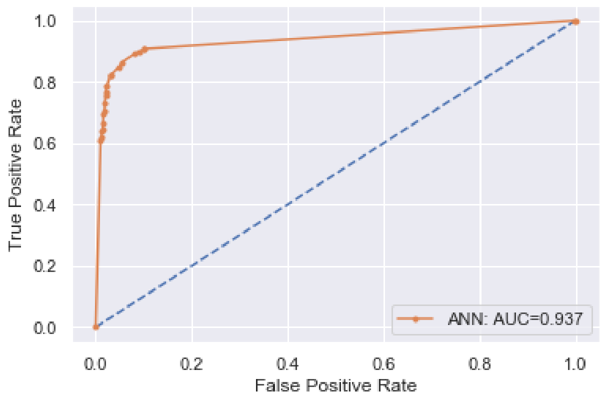

Fig. 5. AUC for ANN using the balanced data set

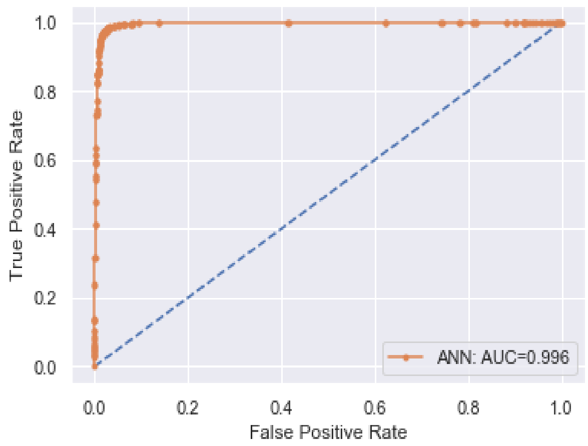

Fig. 6. AUC for ANN using PCA technique on the balanced data set

V. CONCLUSION

Through this paper, we have focused on the class imbalance problem being faced while detecting epilepsy in patients. Our experimentation work has been phased out into three phases, in the first phase, the comparison amongst seven machine learning classifiers has been drawn using the original data set. The comparison has been made drawing inferences from five performance parameters: accuracy, precision, recall, F1-score, and AUC. In the second phase, the SMOTE technique has been used for balancing the dataset. In the third phase, the PCA technique has been applied for dimensionality reduction of the signal. The derived result shows that the ANN classifiers provide better results than other classifiers when the 5 performance parameters are taken into consideration. The ANN gives an accuracy of 97.82%, precision of 97.27, recall of 96.96, F1-score of 95.91, and AUC of 0.996.

In the future, we are proposing to implement the ensemble learning technique for the prediction of epileptic disorder using a different dataset of patients.

REFERENCES

1. R. Fisher, W. Boas, W. Blume, C. Elger, P. Genton, P. Lee, and J. Engel Jr., “Epileptic seizures and epilepsy: Definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE),” Epilepsia, vol. 46, no. 4, pp. 470–472, 2005.
2. J. Engel, “ILAE classification of epilepsy syndromes,” Epileps Research, vol. 70S, pp. 5–10, 2006.
3. T. Ball, M. Kern, I. Mutschler, A. Aertsen, and A. Schulze-Bonhage, “Signal quality of simultaneously recorded invasive and non-invasive EEG,” Neuroimage, Mar. 2009.
4. M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, “A review on ensembles for the class imbalance problem: Bagging-, Boosting- , and hybrid-based approaches,” IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 42, no. 4, pp. 463–484, Jul. 2012.
5. Q. Yuan, W. Zhou, L. Zhang, F. Zhang, F. Xu, Y. Leng, et al., "Epileptic seizure detection based on imbalanced classification and wavelet packet transform," Seizure, vol. 50, p. 99, 2017.
6. C. E. Elger and C. Hoppe, ‘‘Diagnostic challenges in epilepsy: Seizure under-reporting and seizure detection,’’ Lancet Neurol., vol. 17, no. 3, pp. 279–288, 2018.
7. N. V. Chawla, “Data mining for imbalanced datasets: An overview,” in Data Mining and Knowledge Discovery Handbook, 2010, pp. 875–886.
8. Birjandtalab, J., Pouyan, M. B., Nourani, M., “ Unsupervised EEG analysis for automated epileptic seizure detection,” First International Workshop on Pattern Recognition, International Society for Optics and Photonics, pp. 100110M-100110M-5, 2016.
9. N. D. Truong et al., “Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram,” Neural Networks, pp. 104–111, 2018.
10. J. Gotman and L. Wang, “State-dependent spike detection: Concepts and preliminary results,” Electroencephalogr. Clin. Neurophysiol., vol. 79, pp. 11–19, 1991.
11. A. Dorai, K. Ponnambalam, "Automated epileptic seizure onset detection," in Proc. the International Conference on Autonomous and Intelligent Systems, pp. 1-4, 2010.
12. A.T. Tzallas, M.G. Tsipouras and D.I. Fotiadis “Automatic seizure detection based on time-frequency analysis and artificial neural networks”, Comput. Intell. Neurosci., vol. 2007, pp.80510-1 -80510-13, 2007
13. D. Delen et al., Practical Text Mining and Statistical Analysis for NonStructured Text Data Applications. Waltham,MA, USA: Academic, 2012.
14. Logesparan, L. “Assessing the impact of signal normalization: preliminary results on epileptic seizure detection”. 33rd Annual International Conference of the IEEE EMBS (pp. 1439-1442). Boston: IEEE, 2011.
15. H. U. Amin, A. S. Malik, R. F. Ahmad, N. Badruddin, N. Kamel, M. Hussain, et al., "Feature extraction and classification for EEG signals using wavelet transform and machine learning techniques," Australasian Physical & Engineering Sciences in Medicine, pp. 1-11, 2015.
16. C. Guerrero-Mosquera, A. Malanda Trigueros, J. Iriarte Franco, and A. Navia-Vazquez, “New feature extraction approach for epileptic EEG signal detection using time-frequency distributions,” Med. Biol. Eng. Comput., vol. 48, no. 4, pp. 321–330, Apr. 2010.
17. R. Esteller, J. Echauz, T. Cheng, B. Litt, and B. Pless, “Line length: An efficient feature for seizure onset detection,” Proc 23rd Int. Conf. IEEE Engineering Medicine Biology Soc., vol. 2, pp. 1707–1710, 2001.
18. N. Koolen, “Line length as a robust method to detect high-activity events: Automated burst detection in premature EEG recordings,” Clin Neurophysiol., vol. 125, no. 10, pp. 1985-94, Feb. 2014.
19. L. Guo, D. Rivero, J. Dorado , J.R. Rabuñal and A. Pazos “Automatic epileptic seizure detection in EEGs based on line length feature and artificial neural networks”, J. Neurosci. Methods, vol. 191, pp.101 -109, 2010
20. U. R. Acharya, et al., "Automated diagnosis of epileptic EEG using entropies", Biomedical Signal Processing and Control, vol. 7, no. 4, pp. 401-408, 2012.
21. Y. Zhang, Y. Zhang, J. Wang and X. Zheng, "Comparison of classification methods on EEG signals based on wavelet packet decomposition", Neural Computing and Applications, Vol. 26, July 2015, pp. 1217-1225.
22. A. Shoeb and J. Guttag, “Application of machine learning to epileptic seizure detection,” in Proc. Int. Conf. Mach. Learn., Jun. 2010.
23. T. Zhang, W.Chen, M. Li, “AR based quadratic feature extraction in the VMD domain for the automated seizure detection of EEG using random forest classifier”, Biomedical Signal Processing and Control, Volume 31, pp 550-559, Elsevier, January 2017.
24. M. P. Hosseini, D. Pompili, K. Elisevich, and H. Soltanian-Zadeh,”Random Ensemble Learning For EEG Classification,” Artificial intelligence in medicine, vol. 84, pp. 146 158, 2018.
25. G.M. Weiss, “Mining with Rarity: A Unifying Framework,” ACM SIGKDD Explorations Newsletter, vol. 6, no. 1, pp. 7-19, 2004.
26. Z. H. Zhou and X. Y. Liu, “Training Cost-Sensitive Neural Networks with Methods Addressing the Class Imbalance Problem,” IEEE Transactions on Knowledge and Data Engineering, vol. 18, no. 1, pp. 63-77, 2006.
27. Siuly, Y. Li and P. Wen, "Classification of EEG signals using sampling techniques and least square support vector machines," RSKT 2009, LNCS 5589, pp. 375-382, 2009.
28. I. Rish, “An empirical study of the naive Bayes classifier,” in Proc. IJCAI Workshop Empirical Methods AI, 2001, pp. 41–46.
29. W. Noble, “What is a support vector machine?” Nature Biotechnology, vol. 24(12), p. 15641567, 2006.
30. C.-Y.J. Peng, K.L. Lee, and G.M. Ingersoll, “An Introduction to Logistic Regression Analysis and Reporting,” The J. Educational Research, vol. 96, no. 1, pp. 3-14, Sept. 2002.

31. T. Denœux, "A k-nearest neighbor classification rule based on Dempster–Shafer theory," IEEE Trans. Syst., Man, Cybern., vol. 25, no. 5, pp. 804–813, May 1995.
32. C. Olaru and L. Wehenkel, "A Complete Fuzzy Decision Tree Technique," Fuzzy Sets and Systems, vol. 138, no. 2, pp. 221-254, 2003.
33. A. Liaw and A. Wiener. (2002) Classification and regression by random forest. R Newsletter [Online]. Available: http://www.r-project.org
34. K. A-S. Qeethara, "Artificial Neural Networks in Medical Diagnosis, " International Journal of Computer Science Issues, vol. 8(2), 2011.
35. C. X. Ling and V. S. Sheng, "Cost-Sensitive Learning and the Class Imbalance Problem," Encyclopedia of Machine Learning, springer, 2008
36. P. H. Hsu, Y. H. Tseng, and P. Gong, "Dimension reduction of hyperspectral images for classification applications," Ann. GIS, vol. 8, no. 1, pp. 1–8, Jun. 2002.
37. R. G. Andrzejak, K. Lehnertz, C. Rieke, F. Mormann, P. David, and C. E. Elger, "Indications of non-linear deterministic and finite dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," Phys. Rev. E, vol. 64, no. 6, pp. (061907)1–8, 2001.

AUTHORS PROFILE

Prabhsimar Kaur is pursuing her M.Tech from Department of Computer Science and Engineering at DIT University, Dehardun, India. She has completed her B.Tech in Electronics and Computer Science Engineering from Rayat Bahra University, Kharar, India. She has more than two years of working expereience as Project Engineer at WIPRO LTD and has worked on ETL Tools- Abinitio, Datastage, Hadoop, and Node.js. She also has experience of working on client projects for companies like Airtel and Purple Intellect. She has attended three international conferences and has written four book chapters on both National and International domain like Springer, De-Gruyter and IEEE.

Vishal Bharti is working as Professor and Head in Department of CSE in DIT University, Dehradun. Previously he worked as Professor and Head in AIT-CSE at Chandigarh University, Mohali, Punjab, Associate Prof. & Head, Department of CSE at Dronacharya College of Engineering Gurgaon. He completed his Ph.D. in 2016. He did his M.Tech.and B.E. from Birla Institute of Technology, Mesra. He also holds Doctorate in Management Studies in addition to MBA in IT and Executive-MBA in S&M. He is having a mixed bag of experience of 13+ Years in both Academia and IT Industry He has five filed patents, twenty-seven copyrights, and 13 Govt. Grants(SERB, DST, NSTEDB, EDI etc. to his credit. He published fifty-six research papers at both National & International domain like IEEE, Springer, Taylor and Francis, and was awarded Best Faculty Award.

Srabanti Maji is working as an Assistant Professor at the Department of CSE in DIT University, Dehradun. Previously she has worked at SGHCMT Patiala under Punjabi University in Punjab. She has done Ph.D.(CSE) in the year 2013 from Thapar University, Patiala. She has of more than 5 year's of research experience and 13 years of teaching experience. She has published around 4 SCI Journal publications in the field of Machine Learning and has many International Journal publications. She also has attended 5 conferences and has published 2 book chapter publications. She is also having a ACM membership.