Journal of Investigations on Engineering & Technology
Volume 4, Issue 2, 47-60, 2021
ISSN: 2687-3052

Research Paper

Classification of Epileptic Seizure Dataset Using Different Machine Learning Algorithms and PCA Feature Reduction Technique

Shamriz NAHZAT[1], Mete YAĞANOĞLU [2]

(Alınış / Received: 01.10.2021, Kabul / Accepted: 23.12.2021, Online Yayınlanma / Published Online: 31.12.2021)

Keywords
Epileptic Seizures
K-Nearest Neighbors (KNN)
Machine Learning (ML)
Principal Components Analysis (PCA)
Random Forest (RF

Abstract: Epileptic seizures are currently one of the leading reasons for morbidity and mortality in the world. With the rise of epileptic seizures around the world and their effect on people's lives, it's more important than ever to get an accurate and timely diagnosis. These days, machine learning techniques are utilized to forecast or diagnose various life-threatening diseases such as epilepsy, cancer, diabetes, heart disease, thyroid, and so on. Early detection and treatment of diseases such as epilepsy will save a person's life. The fundamental goal of this work is to find the best classification algorithm for epileptic seizures by applying the Principal Components Analysis feature reduction technique in the dataset. In this paper, we applied K-Nearest Neighbors, Random Forest, Support Vector Machine, Artificial Neural Network, and Decision Tree algorithms by using the Principal Components Analysis feature reduction technique in the dataset to predict epilepsy, and the performance of classifiers are analyzed with using Principal Components Analysis and without using the Principal Components Analysis technique. The models used in this analysis have various degrees of accuracy. This study indicates that the used model can accurately predict epilepsy. Our findings indicate that using Principal Components Analysis feature reduction in the dataset, the Random Forest, classifier with 97 % accuracy and low computational times (training and testing time) produces the best results. Also, the K-Nearest Neighbors and Random Forest with 99 % accuracy without using Principal Components Analysis feature reduction in the dataset shows the best result compared to other machine learning techniques.

1. Introduction

Seizure, which is, exactly known as an epileptic seizure, is a transient neurological disorder of the brain that can be triggered by a sudden over-activity of nerve cells in the brain [1]. It is a widespread neurological condition that affects people of all ages [2]. One percent of people around the world suffer from this disease [3].

There are many causes of epilepsy, including vascular, brain infections, brain tumors, nutritional deficiencies, pyridoxine deficiency, and calcium metabolism disorders. To accurately diagnose epilepsy, research is needed to properly understand the mechanisms that cause epileptic disorders.

Magnetic resonance imaging (MRI), computed tomography (CT) scan, positron emission tomography (PET), ultrasound, and electroencephalogram (EEG) diagnostic tools are available. However, MRI, CT scan, and Ultrasound are expensive and can't be used for long-term detection. On the other hand, EEG, is a low-cost test, that can be used for long-term detection. As a result, EEG is the most effective method for diagnosing epilepsy [4].

[1] Atatürk University, Department of Computer Engineering, 25240, Erzurum, Türkiye
[2] Atatürk University, Department of Computer Engineering, 25240, Erzurum, Türkiye

The EEG provides a wealth of physiological and pathological data that is useful in the treatment of epilepsy cases, such as assessing the epileptogenic zone for presurgical assessments [5]. Currently, EEG diagnosis focuses on neurologists physically inspecting EEG recordings. The visual scoring of long-term EEG is time-consuming and boring. As a result, the automatic recognition technology is helpful to neurologists when analyzing EEG records or informations.

Machine learning (ML), a sub-branch and cornerstone of artificial intelligence (AI), has made great strides over the past two decades. ML uses mathematical and computer science concepts as well as algorithms to reveal the underlying features of data and intrinsic connections. It is currently widely used in the field of disease diagnosis. These days, machine learning techniques are utilized to forecast or diagnose various dangerous sicknesses such as epilepsy, cancer, diabetes, heart disease, thyroid, and so on. Early detection and treatment of diseases such as epilepsy will save a person's life.

But diagnosing possible seizures in advance is not an easy job. The majority seizures occur unexpectedly and finding ways to diagnose feasible seizures before they occur, is a challenge for many researchers. The method utilized in this article, will aid in determining whether or not somebody is having a seizure.

The fundamental goal of this work is to find the best classification algorithm for epileptic seizures by applying the Principal Components Analysis (PCA) feature reduction technique in the dataset. In this paper, we applied K-Nearest Neighbors (KNN), Random Forest (RF), Support Vector Machine (SVM), Artificial Neural Network (ANN), and Decision Tree (DT) algorithms by using the PCA feature reduction technique in the dataset to predict epilepsy, and the performance of classifiers are analyzed with using PCA and without using the PCA technique.

2. Related Works

This part of the paper is allocated to several scholars who have discussed issues relating to epileptic seizures and have used machine learning methods to predict epileptic seizures. We discuss a number of recent studies on epileptic seizure diagnosis using EEG signals.

In 2020 Almustafa has utilized various Machine learning techniques like RF, DT, K-NN, Naïve Bayes, Logistic Regression, Random Tree, J48 and Stochastic Gradient Descent (S.G.D.) to the classification of the Epileptic Seizure dataset and achieved 97.08% accuracy by using the Random Forest classifier [5].

In 2019 Nandy et al. used SVM classifier for classification Epileptic Seizure and for optimization of hyper-parameters of SVM has used a Bayesian optimization algorithm. Furthermore, they used Linear Discriminant Analysis (LDA) and Quadratic Linear Discriminant Analysis (QLDA) for comparison, in their paper, the SVM classifier shows 97.05% accuracy [4].

In 2019 Usman et al. have used principal component analysis (PCA) for feature extraction and support vector machines classifier to the classification of the Epileptic Seizure, the proposed model shows an average sensitivity of 93.1% [6].

Hamad et al. have utilized the discrete wavelet transform (DWT) approach to extract features, then, these features are employed to train the SVM with radial basis function (RBF) kernel function. In order to achieve an effective EEG classification, the grey wolf optimizer (GWO) was employed to choose the important feature subset and the appropriate SVM parameters [7].

In 2016 Sharmila & Geethanjali mainly used discrete wavelet transforms (DWT) to decompose EEG data into separate sub-bands and then derive statistical features. The classifier is trained using the DWT-derived statistical features. The signals are then classified using two classifiers to decide if they are epileptic or not. The KNN and Naive Bayes classifiers are the two classifiers that are used in this study. This study compares the success of 14 different two-class epilepsy detection combinations. The results of the experiments showed that, in order to diagnose epileptic seizures, The Naive Bayes classifier achieves the highest accuracy with less computing time for most dataset combinations [8].

Swami et al. used dual-tree complex wavelet transform (DTCWT) for decomposition of signals and calculated statistical measurements then all statistical measurements were trained using by general regression neural network classifier, finally the model shows 95.24% accuracy [9].

3. Material and Methods

The goal of our work is to find the best classification algorithm for epileptic seizures by applying the PCA feature reduction technique in the dataset.
The phases will be discussed in the following sections. Figure 1. illustrates the overall flow chart of the suggested model.

3.1. Dataset Description

The epileptic seizure dataset utilized in this paper is from Bonn University, which can be found on the UCI Machine Repository website [10]. This dataset include five classes of 1 to 5, each of which is 100 signals with a length of 23.6 seconds, the Classes 5 and 4 were taken from five healthy human beings with open eyes and closed eyes. The other three classes (3, 2, and 1) were recorded from five epileptic patients. The two classes 3 and 2 were recorded, when there is no epileptic seizure, (the class 3 were recorded from pre-seizure hippocampal contralateral hemisphere and the class 2, was obtained from epileptogenic region of patient.) And the class1 was recorded during seizures.

All EEG signals are captured by a 128-channel system with a sampling rate of 173.61 Hz using a 12-bit analog-to-digital converter. There are 11,500 samples in the dataset, each having 178 attributes, and they are regularly distributed. All of the cases in classes 2, 3, 4, and 5 have never had an epileptic seizure. Only class 1 people experienced epileptic seizures [10]. As a result, for epileptic seizure and non-epileptic seizure instances, our analysis will be a binary structure, with classes 2,3,4,5. Table 1. displays the number of cases for each of the classes utilized, and we can see that all of them have an equal number of samples. Figure 2. shows the epileptic seizure dataset in a sample view.

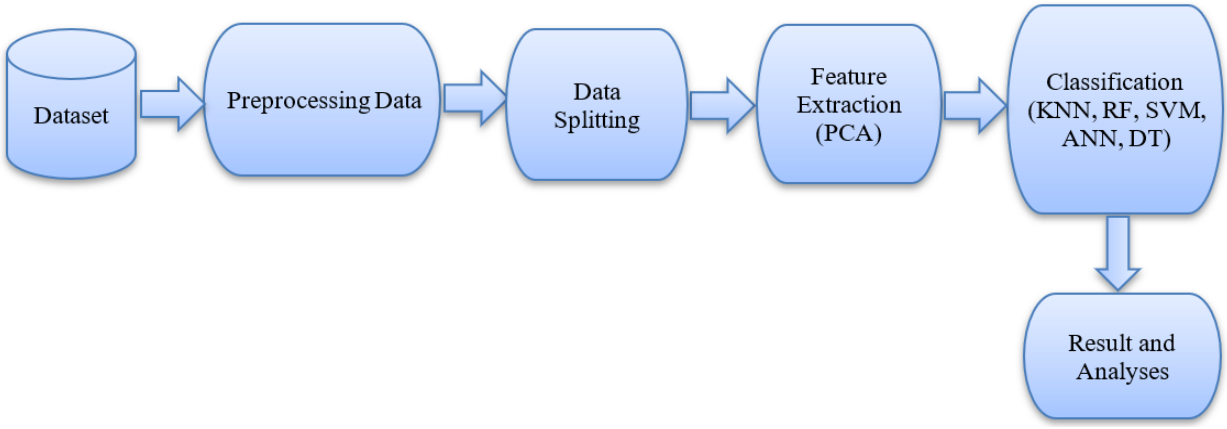

Figure 1. Proposed Model

Table 1. Dataset Description and number of cases in each class.

Classes	The Class Description	The Patient State	The Number of cases	Binary case
1	Seizure activity is recorded from epileptic patients	General epilepsy (with seizures)	2300	2300
2	The tumor was observed in epileptic patients	Partial epilepsy (without seizures)	2300	
3	The E.E.G. signal was recorded from a healthy brain region of epileptic patients	Partial epilepsy (without seizures)	2300	9200
4	eyes closed	Healthy	2300	
5	eyes opened	Healthy	2300	

	Unnamed: 0	X1	X2	X3	X4	X5	X6	X7	X8	X9	...	X170	X171	X172	X173	X174	X175	X176	X177	X178	y
11495	X22.V1.114	-22	-22	-23	-26	-36	-42	-45	-42	-45	...	15	16	12	5	-1	-18	-37	-47	-48	2
11496	X19.V1.354	-47	-11	28	77	141	211	246	240	193	...	-65	-33	-7	14	27	48	77	117	170	1
11497	X8.V1.28	14	6	-13	-16	10	26	27	-9	4	...	-65	-48	-61	-62	-67	-30	-2	-1	-8	5
11498	X10.V1.932	-40	-25	-9	-12	-2	12	7	19	22	...	121	135	148	143	116	86	68	59	55	3
11499	X16.V1.210	29	41	57	72	74	62	54	43	31	...	-59	-25	-4	2	5	4	-2	2	20	4

Figure 2. The epileptic seizure dataset in a sample view.

3.2. Data Preprocessing

One of the most critical and required stages in machine learning is data preparation. This technique is essential for reliable, accurate, and successful prediction outcomes when using machine learning algorithms in a data set [11]. Data preparation is a methodology that includes turning raw and unprocessed data into a suitable format for the classification process.

Data from the real world is frequently insufficient, untrustworthy, and/or lacking in specific behaviors or patterns, as well as including various mistakes. Pre-processing data is a tried-and-true way to solve such issues. Raw data is pre-processed to make it ready for subsequent processing.

In our dataset, there are no missing values (NAN). According to Table 1 we can observe that in the binary class, there is an unbalanced class distribution issue; to avoid this, we employ under sampling approaches.
Under sampling is a term that refers to a series of strategies for balancing the class distribution in a classification dataset with a skewed class distribution.

To standardize our dataset, we employed the Min-Max normalization technique. After normalization of the dataset in the proposed methodology, we used 75 percent of the data for validation and training and 25 percent of data for the testing.

3.3. Feature Extraction

The purpose of the feature extraction phase is to minimize the amount of attributes by creating new ones from the dataset's current ones. Most of the information and features in the original dataset should be summarized by this new reduced feature set. As a result of combining the original set, a concise version of the primary features can be generated [12].

As we discussed earlier, there are 178 features in our data set, if we use all the features for training, the training time will be very long, so in our study, to extract and reduce the features we, used Principal Components Analysis (PCA).

3.3.1. Principal Components Analysis (PCA)

PCA is an information extraction approach that involves projecting data from a high-dimensional space onto a lower-dimensional subspace. It seeks to maintain the data's key bits with the highest variance and delete the non-essential sections with the least variance [13].

Our objective with PCA is to discover a collection of input characteristics that can best explain the distribution of the original data by decreasing its original dimensions. PCA may achieve this by maximizing variances and decreasing reconstruction error by monitoring fragmented distances. Our main data is projected onto a set of orthogonal axes, with each axis rated in order of relevance in PCA.
PCA is an unsupervised learning method that is not interested in data labels and only cares about diversity [14].

When PCA is used, the majority of the variation in the data is concentrated in the first few components. As a consequence, only those components with significant differences are retained, while rests are ignored.

The n-dimensional mean vector mu is computed first, followed by the n x n covariance matrix R, which is then sorted in decreasing order of eigenvalues. After sorting, the largest eigenvalues are picked. Noise is considered in

the other dimensions. If we create a n x n matrix A with n eigenvectors in columns, the data can be retrieved using the below equation after pre-processing [14].

$$X'=At(x-\mu) \quad (1)$$

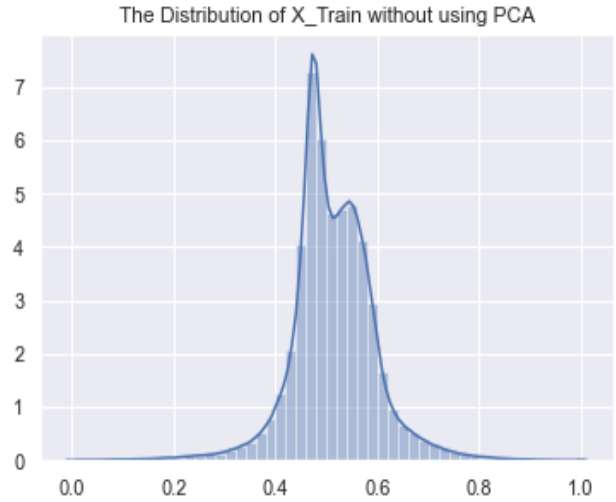

Figure 3. The distribution of X_Train without using PCA.

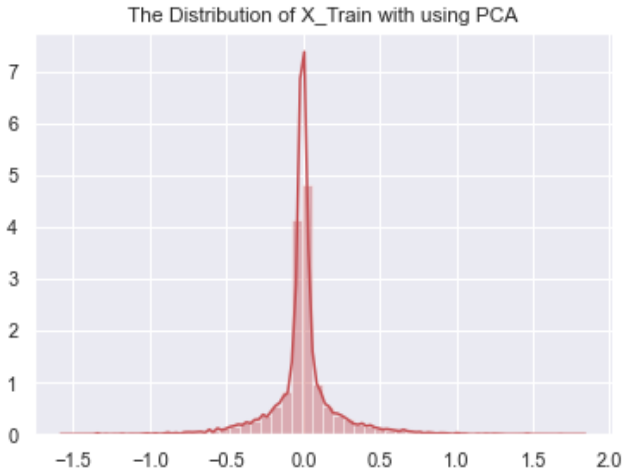

Figure 4. The distribution of X_Train using PCA.

Figure 3. illustrates the distribution of X_Train without the use of PCA, whereas Figure 4. represents the distribution of X_Train with the use of PCA, in here we can see applying PCA reduces the size of the data.

3.4. Algorithms used for Classification

The algorithms and practical machine learning methods for categorization and estimation utilized throughout this article are:

3.4.1. K-Nearest Neighbors (KNN)

K-NN is among the supervised learning techniques frequently used in classification studies. It is a nonparametric and basic technique that classifies objects in the input space based on the nearest samples [15]. While the number of neighbors is shown with k, it is named k-nearest neighbor algorithm because its proximity to the data to be classified is important. The KNN Classification method aims to address both classification and regression problems. The KNN algorithm is one of the algorithms that take a long time to learn [16].

K-NN is a learning algorithm based on calculating the sample distance, each time the algorithm encounters a new data sample, the distance is calculated on all samples in the new data. After this calculation, class labels are classified by finding k closest neighbors from previously known data samples and comparing them with the examples in the education data in the new example and looking at the similarities between them [17].

Figure 5. shows the Confusion matrix for KNN algorithm when k=1 is selected, and the PCA feature reduction technique is used to forecast epilepsy. In figure 6. we can see the best value for k is one because it shows the highest accuracy in training and testing.

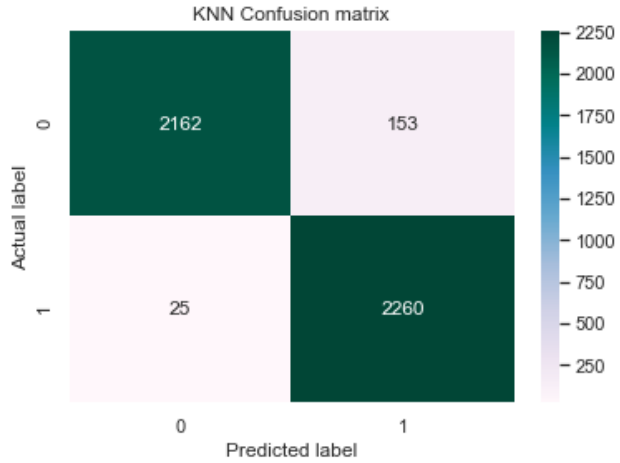

Figure 5. Confusion matrix for KNN algorithm with using of PCA feature reduction technique.

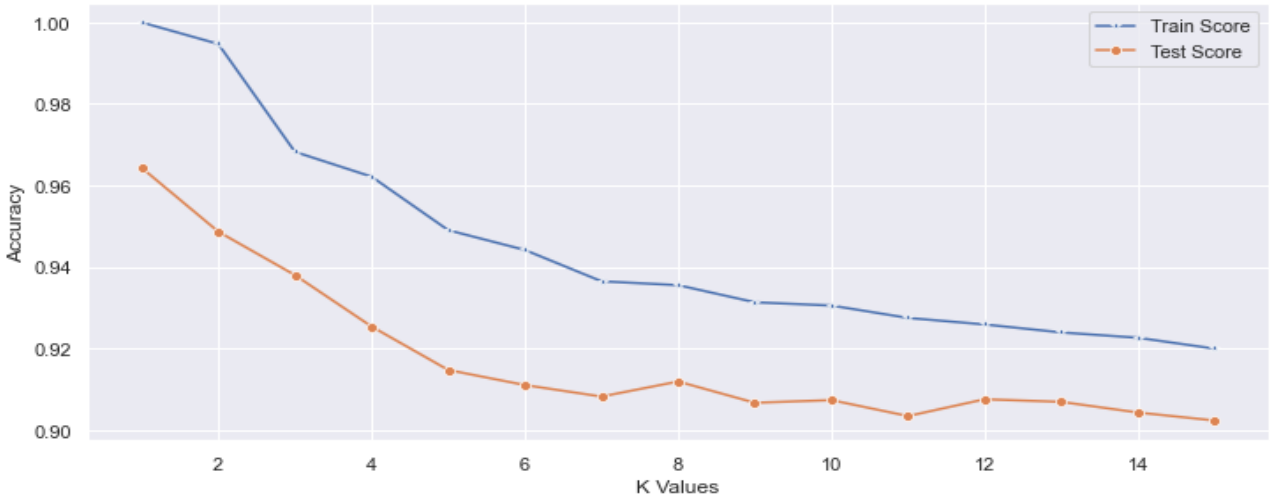

Figure 6. The Train Score and Test Score of KNN algorithm with different k values.

3.4.2. Random Forest (RF)

The RF classifier, which has been developed in recent years, provides an advantage over Acceleration [18] and Bagging [19] methods, which are known as two very good methods in collective learning, in terms of both fast and high accuracy. Compared to learning methods, the RF classifier is much faster during the training phase, especially than the Acceleration method. It is a very useful classifier with its efficiency and accuracy [20].

RF is an easy-to-use ML algorithm that, even without changing its meta parameters, often delivers great results. This algorithm is one of the most commonly utilized machine learning algorithms for both “Classification” and “Regression” due to its simplicity and usability. This algorithm would randomly create a forest. The built “forest” is actually a “Decision Trees” band. This strategy can undoubtedly deal with huge datasets.

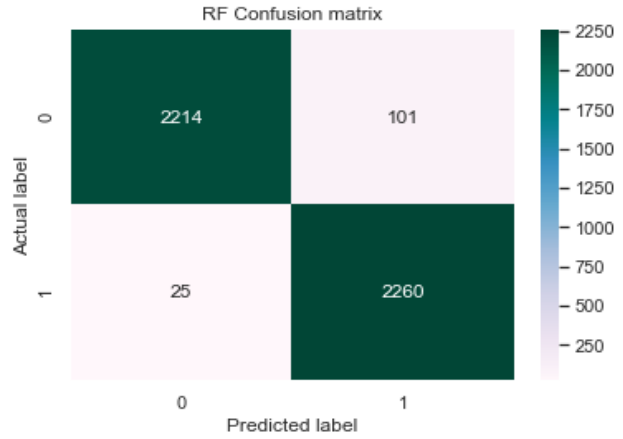

Figure 7. Confusion matrix for Random forests algorithm using PCA feature reduction technique.

Random Forest was created by Leo Bremen. It selects samples randomly from the dataset then builds a decision tree for each sample. A prediction result is measured from each decision tree. Then vote the prediction result after that the most votes consider the final prediction model [21].

Figure 7. illustrates the Confusion matrix for the RF algorithm when 100 estimators are selected, and the PCA feature reduction technique is used to forecast epilepsy.

3.4.3. Support Vector Machine (SVM)

SVM stands for supervised machine learning. Vapnik and Chervonenkis were the first to present the SVM in 1963. The SVM tries to locate an ideal hyperplane ready to isolate the examples of any class. If groups can be divided linearly, hyperplanes with maximum margins might be used to identify them. Otherwise, if the data are not linearly separable, they can be transferred to a bigger space to separate them linearly (i.e. feature space). The kernel function is the name for this conversion. This classifier specifies the hyperplane that isolates the spots to put the most noteworthy number of points of a similar class on a similar side while expands the interval of each class to such a hyperplane. The support vectors comprise of the closest points of the hyperplane. The interval from a class to a hyperplane is the littlest interval among them and the spots in that class [16].

The hyperplane can be utilized for grouping or regression moreover. SVM separates examples in particular groups and can likewise characterize the substances which are not upheld by data. Detachment is finished by through hyperplane plays out the partition to the nearest training spot of any group.

Figure 8. Shows the Confusion matrix for SVM algorithm using of PCA feature reduction technique to predict epilepsy.

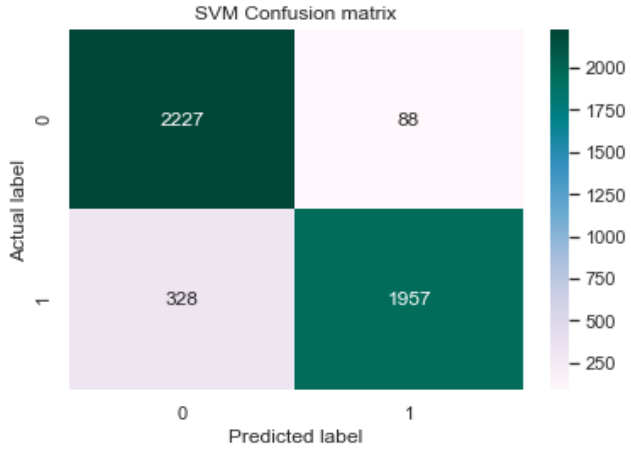

Figure 8. Confusion matrix for SVM algorithm with PCA feature reduction technique.

3.4.4. Artificial Neural Network (ANN)

ANNs are a network structure composed of a series of interconnected elements called neurons, each of which has an input and output and performs a relatively simple operation. Neural networks generally learn their function through a learning process. In fact, by processing data, they discover the law underlying them and transmit it to the network.
In fact, these networks are a software programs that can act like human beings in such a way that:

a. More experienced, through over the time and by more interaction with the environment.
b. In addition to performing calculations, able to draw logical conclusions.
c. Provide a suitable solution in new conditions.

Artificial neural networks are computational structures modeled on the human brain. ANN is made up of many interconnected unit operations that cooperate to process data. They often deliver beneficial outputs as a result of it. In general, the ANN is comprised of network layers and network tasks, which the network layers namely the input layer, hidden layer and output layer. For the data mining model, the input neurons determine all the input attribute values [22].

A significant number of new headways have been made in the field of Artificial Intelligence, utilizing Artificial Neural Networks including Voice Recognition, Image Recognition and Robotics.

In Figure 9. The Confusion matrix for ANN algorithm using the PCA feature reduction technique to predict epilepsy is shown.

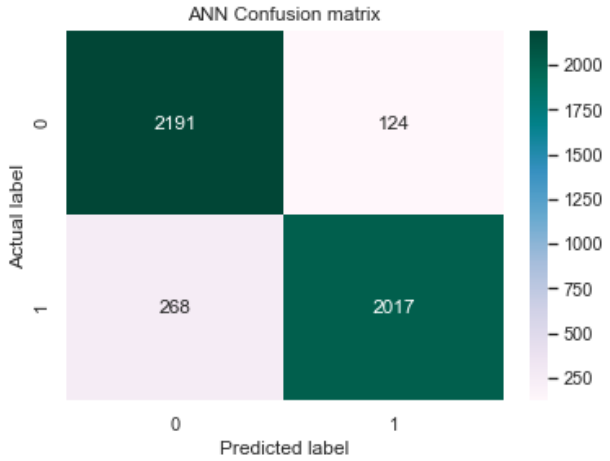

Figure 9. Confusion matrix for ANN algorithm using PCA feature reduction technique.

3.4.5. Decision Tree (DT)

A DT is a decision-making aid that employs tree modeling. DT is a fundamental classification and regression technique. In operations and research, the decision tree is commonly used. DT model that has a tree structure can be used to define the mechanism of classifying instances based on characteristics [23]. When the result attribute is categorical, a decision tree is used.

The root node, branches, and leaves make up the decision tree graph. Classification takes place on the leaves, while the outcomes of each occurrence are stored on the branches. The pathways from the root node to the leaf nodes are taken into account while creating categorization rules [24].

Both nominal and numerical features are provided by the decision tree algorithm. It has the potential to tolerate noise and unstable values. The decision tree uses a top-down approach to categorize the whole qualified dataset by partitioning the nodes from the topmost to the class node. Every node represents the instance's test attribute, with each node representing one of the several likely values for that feature attribute.

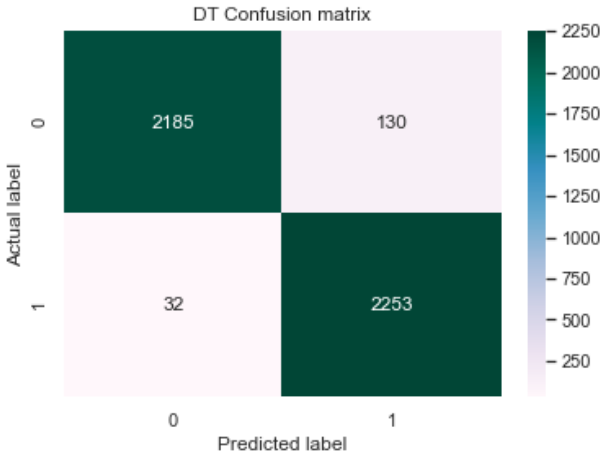

Figure 10. Confusion matrix for DT algorithm using PCA feature reduction technique.

From the top node to the attack class node level by level, a decision tree can easily turn the specified set of instances into meaningful patterns.

In Figure 10. the Confusion matrix for DT algorithm using the PCA feature reduction technique to predict epilepsy is Shown.

4. Results and Discussion

The outcomes of utilizing several classifiers to classify the epilepsy data set and the performance of the classifications using PCA and without using PCA feature reduction are analyzed in this section.
Table 2 and Figure 11. shows the Comparison of the different Classification Techniques using PCA feature reduction and setting 75% of the dataset for training and 25% for testing. According to the Table 2. We see that the RF Classification algorithm with 97 % accuracy and with low Computational times (training time and test time) shows the best result, after the RF the KNN and DT algorithms with 96 % accuracy and too low Computational times are shown the best result. It is worth to mentioning that the Computational times is also related to power of the computer workstation, it means that in a powerful workstation the Computational times will be too low.

Table 2. Comparison of the different Classification Techniques using PCA feature reduction and setting 75% of the dataset for training and 25% for testing.

Classification Techniques	Accuracy	Precision	Recall	F1 score	Train time (S)	Test time (S)
KNN	96%	96%	96%	96%	0.015	0.29
RF	**97%**	**97%**	**97%**	**97%**	0.92	0.05
SVM	90%	91%	90%	90%	2.23	0.41
ANN	91%	91%	91%	91%	6.70	0.005
DT	96%	96%	96%	96%	**0.057**	**0.0009**

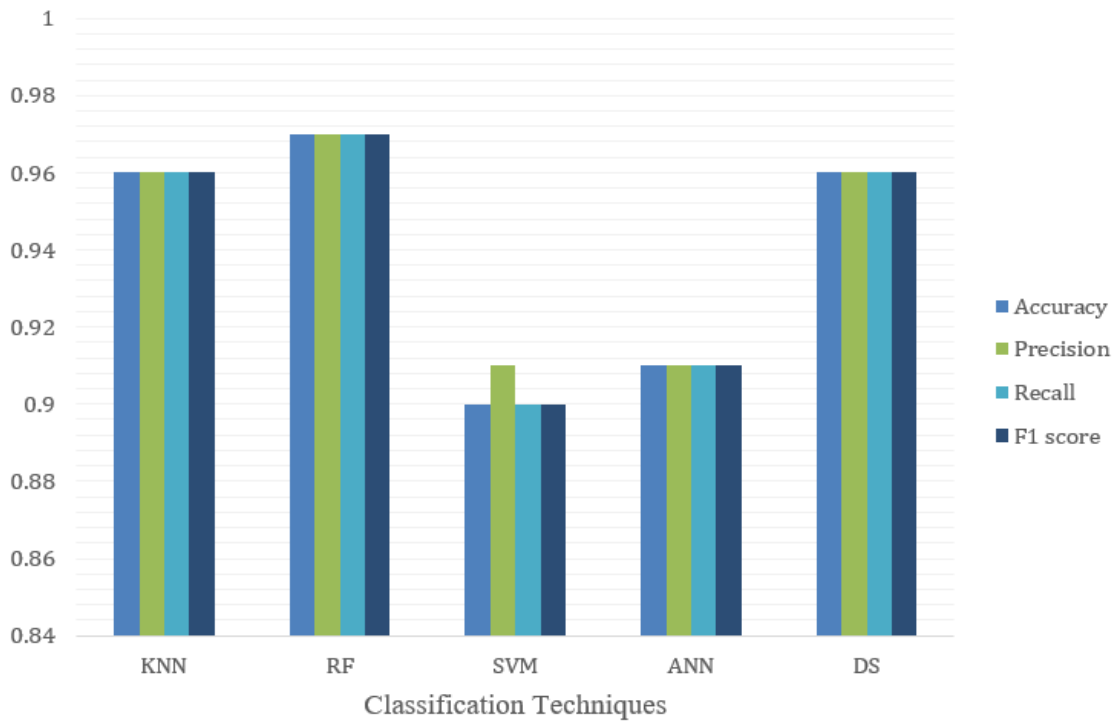

Figure 11. Comparison of the different Classification Technique using PCA feature reduction and setting 75% of dataset for training and 25% for testing

Table 3. Comparison of the different Classification Techniques without using the PCA feature reduction and setting 75% of dataset for training and 25% for testing.

Classification Techniques	Accuracy	Precision	Recall	F1 score	Train time (S)	Test time (S)
KNN	**99%**	**99%**	**99%**	**99%**	1.00	15.68
RF	**99%**	**99%**	**99%**	**99%**	6.83	0.072
SVM	97%	97%	97%	97%	10.15	2.517
ANN	92%	92%	92%	92%	22.11	0.014
DT	97%	97%	97%	97%	3.13	0.009

Table 4. Comparison of the different classification techniques with using PCA feature reduction and setting 70% of dataset for training and 30% for testing.

Classification Techniques	Accuracy	Precision	Recall	F1 score	Train time (S)	Test time (S)
KNN	96%	96%	96%	96%	0.014	0.331
RF	**97%**	**97%**	**97%**	**97%**	0.986	0.068
SVM	90%	91%	90%	90%	1.735	0.424
ANN	91%	91%	91%	91%	5.9	0.005
DT	96%	96%	96%	96%	0.0519	0.0009

According to Table 3 We see that without using the PCA feature reduction technique the classification algorithms show high accuracy, but with high Computational times, in this case the KNN and random Forest (RF) Classification algorithms with 99 % accuracy shows best result. After the KNN and RF the DT and SVM Classifiers with 97 % accuracy has good result.

Table 4 illustrates the comparison of the different classification techniques using the PCA feature reduction technique and setting 70% of dataset for training and 30% for testing. In this case, we see that the RF classifier shows the best result with 97% accuracy.

Table 5. Comparison of the different classification techniques without using the PCA feature reduction and setting 70% of dataset for training and 30% for testing.

Classification Techniques	Accuracy	Precision	Recall	F1 score	Train time (S)	Test time (S)
KNN	**99%**	**99%**	**99%**	**99%**	0.961	16.20
RF	**99%**	**99%**	**99%**	**99%**	5.62	0.075
SVM	97%	97%	97%	97%	8.56	2.43
ANN	92%	92%	92%	92%	22.297	0.0139
DT	97%	97%	97%	97%	2.81	0.008

Table 6. Accuracy Comparison of Classification Techniques using PCA and without using PCA

Classification Techniques	Accuracy using the PCA	Accuracy without using the PCA
KNN	**96%**	**99%**
RF	**97%**	**99%**
SVM	90%	97%
ANN	91%	92%
DT	96%	97%

Table 7. Computational time evaluation of Classification Techniques using PCA and without using the PCA.

Classification Techniques	Using PCA		Without Using PCA	
	Train time (S)	Test time (S)	Train time (S)	Test time (S)
KNN	0.015	0.29	1.00	15.68
RF	0.92	0.05	6.83	0.072
SVM	2.23	0.41	10.15	2.517
ANN	6.70	0.005	22.11	0.014
DT	**0.057**	**0.0009**	3.13	0.009

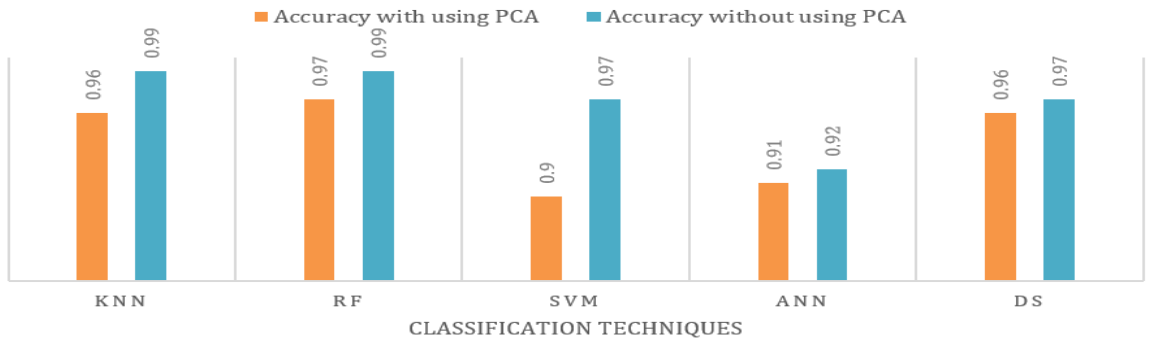

Figure 12. Accuracy Comparison of Classification Techniques using the PCA and without using the PCA.

Table 8. The result comparisons of different research studies with our study.

Research Studies	Number of Classes	Methods	The best method	Accuracy	
Almustafa[5]	2	RF, DT, K-NN, Naïve Bayes, Logistic Regression, Random Tree, J48 and Stochastic Gradient Descent (S.G.D.)	RF	97%	
Nandy et al., [4]	2	SVM classifier for classification, Linear Discriminant Analysis (LDA) and Quadratic Linear Discriminant Analysis (QLDA) for comparison.	SVM	97%	
Usman et al., [6]	2	PCA for feature extraction and SVM classifier to classification.	SVM	Average sensitivity 93.1%	
Swami et al.,[9]	2	Used dual-tree complex wavelet transform (DTCWT) for decomposition of signals and calculate statistical measurements, and general regression neural network classifier for classification	Neural network	95%	
				Without PCA	Using PCA
This Paper	2	K-NN, RF, SVM, ANN & DT	K-NN	99%	96%
			RF	99%	97%
			DT	97%	96%

Table 5 also illustrates the Comparison of the different classification techniques without using the PCA feature reduction technique and setting 70% of dataset for training and 30% for testing. In this case, we see that the KNN and RF classifier also show the best result with 99% accuracy, in addition, shows a little low training and testing time compared to setting the 75% of dataset to training.

According to Table 6, we can see that the RF classifier shows the best result using the PCA feature reduction technique with 97 %. At the same time, without using the PCA feature reduction technique, the KNN and RF with 99 % accuracy, it shows the best result.

Table 7 shows the Computational times evaluation of Classification Techniques using the PCA and without using the PCA feature reduction technique. According to this table we can say that by using PCA feature reduction technique, we reduce the computational times. At the same time, according to Figure 12. the accuracy of the classifiers is reduced.

Table 8 compares the results of other research papers in the field of Epileptic Seizure classification with our paper.

5. Conclusion

Epileptic seizures are currently one of the leading reasons for morbidity and mortality in the world. With the rise of epileptic seizures around the world and their effect on people's lives, it's more important than ever to get an accurate and timely diagnosis.

The fundamental goal of this paper was to discover the best classification algorithm for epileptic seizures by applying the Principal Components Analysis (PCA) feature reduction technique in the dataset.
In this paper, we applied K-Nearest Neighbors (KNN), Random Forest (RF), Support Vector Machine (SVM), Artificial Neural Network (ANN), and Decision Tree (DT) algorithms by using the PCA feature reduction technique in the dataset to predict epilepsy, and the performance of classifiers are analyzed with using PCA and without using the PCA technique.

It has been demonstrated that the Random Forest classifier (RF) with an accuracy of 97% and with low Computational times (training time and test time) shows the best result by using the PCA feature reduction in the dataset. In addition, the K-Nearest Neighbors (KNN) and random forest classifiers (RF) with 99 % accuracy without using PCA feature reduction in dataset, shown the best result.

It is worth to mentioning that using PCA feature reduction technique reduces the computational times. At the same time, the accuracy of the classifiers is also reduced. In both cases, using PCA and without using PCA after the random forest classifier (RF) the KNN Classifier with 96% accuracy by using PCA and 99 % accuracy without using PCA has shown good result.

References

[1] Fisher, R., Acevedo, C., Arzimanoglou, A., Bogacz, A., Cross, J., Elger, C., et al. (2014). ILAE official report: a practical clinical definition of epilepsy, *Epilepsia*, 55,4, 475-482.

[2] Ramgopal, S., Thome-Souza, S., Jackson, M., & Kadish, N. E., Sánchez Fernández, I., Klehm, J., Bosl, W., Reinsberger, C., Schachter, S., & Loddenkemper, T. (2014). Seizure detection, seizure prediction, and closed-loop warning systems in epilepsy. *Epilepsy & behavior: E&B*, 37,291–307.

[3] Lehnertz, K., Mormann, F., Kreuz, T., Andrzejak, R. G., Rieke, C., David, P., & Elger, C. E. (2003). Seizure prediction by nonlinear EEG analysis. *IEEE engineering in medicine and biology magazine: the quarterly magazine of the Engineering in Medicine & Biology Society*, 22,1,57–63.

[4] Nandy, A., Alahe, M. A., Nasim Uddin, S. M., Alam, S., Nahid, A. A., & Awal, M. A. (2019). Feature Extraction and Classification of EEG Signals for Seizure Detection. *2019 International Conference on Robotics,Electrical and Signal Processing Techniques (ICREST)*.

[5] Almustafa, K. M. (2020). Classification of epileptic seizure dataset using different machine learning algorithms. Informatics in Medicine Unlocked, 21, 100444.

[6] Usman, S. M., Latif, S., & Beg, A. (2019). Principal components analysis for seizures prediction using wavelet transform. International *Journal of Advanced and Applied Sciences*, 6, 3, 50–55.

[7] Hamad, A., Houssein, E. H., Hassanien, A. E., & Fahmy, A. A. (2017). A Hybrid EEG Signals Classification Approach Based on Grey Wolf Optimizer Enhanced SVMs for Epileptic Detection. *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics*, 108–117.

[8] Sharmila, A., & Geethanjali, P. (2016). DWT Based Detection of Epileptic Seizure from EEG Signals Using Naive Bayes and k-NN Classifiers. *IEEE Access*, 4, 7716–7727.

[9] Swami, P., Gandhi, T. K., Panigrahi, B. K., Tripathi, M., & Anand, S. (2016). A novel robust diagnostic model to detect seizures in electroencephalography. *Expert Systems with Applications*, 56, 116–130.

[10] Andrzejak, R. G., Lehnertz, K., Mormann, F., Rieke, C., David, P., & Elger, C. E. (2001). Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E*, 64, 6.

[11] Soni, M., & Varma, S. (2020), Diabetes Prediction using Machine Learning Techniques, *International Journal of Engineering Research & Technology (IJERT)*, 9, 9.

[12] Ippolito, P. P. (2019). Feature Extraction Techniques - Towards Data Science. Retrieved, from https://towardsdatascience.com/feature-extraction-techniques-d619b56e31be, (Date of access: December 27, 2020).

[13] Qiu, J., Wang, H., Lu, J., Zhang, B., & Du, K. L. (2012). Neural Network Implementations for PCA and ItsExtensions. *ISRN Artificial Intelligence*, 2012, 1–19.

[14] Ghosh-Dastidar, S., Adeli, H., & Dadmehr, N. (2008). Principal Component Analysis-Enhanced Cosine Radial Basis Function Neural Network for Robust Epilepsy and Seizure Detection. *IEEE Transactions on Biomedical Engineering*, 55(2), 512–518.

[15] Yağanoğlu, M., & Köse, C. (2018). Real-Time Detection of Important Sounds with a Wearable Vibration Based Device for Hearing-Impaired People. *Electronics*, 7, 4, 50.

[16] Rodrigues, J. D. C., Filho, P. P. R., Peixoto, E., N, A. K., & de Albuquerque, V. H. C. (2019). Classification of EEG signals to detect alcoholism using machine learning techniques. *Pattern Recognition Letters*, 125, 140–149.

[17] Mitchell, T. M. (1997). Does Machine Learning Really Work?. *AI Magazine*, 18, 3, 11..

[18] Freund, Y., Schapire, R.E., 1996. Experiments with a new boosting algorithm. In: Machine Learning. *Proceedings of the Thirteenth International Conference*. 148–156.

[19] Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24, 2, 123–140.

[20] Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random Forests for land cover classification. *Pattern Recognition Letters,* 27, 4, 294–300.

[21] Fawagreh, K., Gaber, M. M., & Elyan, E. (2014). Random forests: from early developments to recent advancements. *Systems Science & Control Engineering*, 2, 1, 602–609.

[22] Walczak, S., & Cerpa, N. (2003). Artificial Neural Networks. Encyclopedia of Physical Science and Technology, 631–645.

[23] Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1, 1, 81–106.

[24] Geetha, A., & Nasira, G. M. (2014). Data mining for meteorological applications: Decision trees for modeling rainfall prediction. 2*014 IEEE International Conference on Computational Intelligence and Computing Research*.