 Detection of Epileptic Seizures from EEG Signals by Combining Dimensionality Reduction Algorithms with Machine Learning Models

Muhammad Zubair, Student Member, IEEE, Maria Vladimirovna Belykh, Umesh Kumar Naik M, Mohammad Fareeda Madeen Gouher, Shani Vishwakarma, Shaik Rafi Ahamed, Member, IEEE, Ramanjaneyulu Kongara, Member, IEEE

Abstract— Epilepsy is a neurological condition that affects the central nervous system. While its effects are different for each person, they mostly include abnormal behaviour, periods of loss of awareness and seizures. There are various traditional methods used to analyse EEG signals for epilepsy detection, which concludes to be time-consuming. Recently, several automated seizure detection frameworks using machine learning algorithms have been proposed to replace conventional methods. In this paper, more emphasis has been given to develop a support vector machine (SVM) model using SPPCA and SUBXPCA dimensionality reduction algorithms to increase the classification accuracy. Firstly, Discrete Wavelet Transform (DWT) is applied to EEG signals for extracting the time-frequency domain features of epilepsy such as the energy of each sub-pattern, spike rhythmicity, relative spike amplitude, Dominant Frequency (DF) and Spectral Entropy (SE). The features obtained after performing DWT on an EEG signal are extensive in number, to select the prominent features and to retain their properties, correlation feature sub-pattern-based PCA (SPPCA), and cross sub-pattern correlation-based PCA (SUBXPCA) are used as a dimensionality reduction techniques. To validate the proposed work, performance evaluation parameter such as the accuracy of the time-frequency domain features from different combinations of the dataset has been compared with the latest state-of-the-art works. Simulation results show that the proposed algorithm combined with machine learning classifiers. The best accuracy of 97% for SPPCA is achieved by CatBoost and for SUBXPCA the best accuracy of 98% is achieved by random forest classifier which clearly outperformed the other related works, both in terms of accuracy and computational complexity.

Index Terms— Epileptic Seizures, Discrete Wavelet Transform (DWT), Sub-pattern PCA (SPPCA), Cross Sub-pattern Correlation PCA (SUBXPCA), Support Vector Machine (SVM), Random forest, K-Nearest Neighbours (KNN), LightGBM, CatBoost, Multilayer Perceptron (MLP)

I. INTRODUCTION

Epilepsy is a central nervous system disorder in which the activity of the brain becomes abnormal, leading to periods of unusual behaviour, sensations, seizures and sometimes loss of awareness. Epilepsy affects 5 to 10 people for every 1000 in the world [1]. Early detection of epileptic seizure helps in the early treatment and soothing to the patients. Electroencephalogram (EEG) plays an important role in the detection of epilepsy. There are four basic EEG frequency rhythms, Beta (14-30 Hz) recorded when the person is awake with mental activity going on, Alpha (8-13 Hz) recorded when the person is awake and resting, Theta (4-7 Hz) recorded when the person is sleeping, and Delta (1-3 Hz), when the person is in a deep sleep. The seizures recorded in an EEG signal are generally in the frequency range of 0.5 Hz to 29 Hz. People with epilepsy tend to show abnormal behaviour, seizures and loss of sensations which render them weak and powerless where they can’t defend themselves.

The conventional methods for detection of epilepsy using EEG signals require close observation for extended periods to detect seizures in real-time. Moreover, this detection depends on the experience and interpretation of the administering technician, making it time-consuming and error-prone. While, the external interferences can also reduce the efficiency of diagnosis, there are several works in which the authors eliminated the sources from EEG signals effectively [32] [33]. Therefore, a faster and more appropriate approach for the clinical diagnosis of epilepsy is the need of the hour.

Probably the most common method that is utilized for identifying the epileptic seizure is pattern recognition, where the concealed patterns related to epilepsy are extracted from the EEG signals. In recent years, researchers have endeavoured
various feature extraction strategies to extricate the concealed patterns from EEG signals, for example, “DWT, CWT, FT, DFT, IDFT, STFT, and FFT” [2] [3] [4] [5] [6]. Also, various techniques, for example, “Particle Swarm Optimization (PSO), Simulated Annealing (SA), Ant Colony Optimization (ACO)” [7], and numerous different approaches to choose the best features were researched. In term of the classification models, different classifiers were utilized by the authors such as “SVM, choice tree, artificial neural networks, k-closest neighbours (k-NN), naive Bayes (NB), Gaussian blend model, adaptive neuro-fuzzy derivation frameworks, and learning vector quantization” to distinguish epileptic seizures from the EEG signals. All the works referenced above centre around expanding the accuracy of recognizing epileptic seizures with various blends of feature extraction, feature selection, and classification [8]- [9] [10] [11].

However, after investigating the above state-of-the-art works, no work has been reported on reducing the computational time of the classifiers using dimensionality reduction algorithms. Our proposed work analyses the time-frequency features extracted through discrete wavelet transform (DWT) from EEG signals. Then, by using the proposed SPPCA and SUBXPCA algorithms derived from Principal Component Analysis (PCA), the dimensionality of the features is reduced. Finally, these features are fed to various machine learning models to be further classified into epileptic or non-epileptic EEG signals. To perform DWT, we have used the signal processing and communications toolbox in MATLAB R2020b. The 1-D wavelet decomposition was performed using the wavedec function with haar as the analysing wavelet. To develop the proposed SPPCA and SUBXPCA algorithm we have used various matrix functions available in MATLAB.

II. RELATED WORKS

In Hsui Sen Chiang et al. [12] has developed a model for epilepsy diagnosis, which uses DWT, minimum entropy principle approach, and Associative Petri Net (APN) techniques. Authors concluded that precision rate, G-mean, F-measure and accuracy of the APN diagnosis model generated better results than other machine learning models such as the ”decision tree, neural networks, SVM and Bayes Net (BN).” The combined method of MEPA and APN can accurately determine whether the patients have epilepsy. A precision rate of 99% and negative predictive value of 90.33% has been obtained for the APN Model.

In [13] Wail Mardini et al. proposed a method for detection of epileptic seizures that uses 54 DWT mother wavelets to extract statistical features like “Shannon Entropy, Skewness, Mean Absolute Value (MAV), Average Power (AVP), Standard Deviation (SD), Variance, Mean, Max, Min, Normalized SD, Kurtosis, and Energy” from the given EEG signal. On top of it, a Genetic Algorithm (GA) has been used to reduce the feature space and select more relevant features. ML classifiers such as SVM, KNN,ANN, and NB have been employed for classification of the given EEG signal into epileptic or non-epileptic. After evaluating the metrics, they have concluded that the ANN classifier obtained high accuracy among the other dataset combinations, and it outperformed the other classifiers.

Minxing Geng et al. [14] introduced an effective seizure recognition technique for epilepsy based on Stockwell transform (S-transform) and bidirectional long short-term memory (Bi-LSTM) neural networks for intracranial EEG recordings. Segment based and event-based evaluation has been carried out to evaluate the signals. As a result, authors obtained a sensitivity of 98.09%, a specificity of 98.69%, a sensitivity of 96.3% and a false detection rate of 0.24/h, respectively.

Zixu Chen et al. [15] developed a framework, which includes two parts. In the first part, intensity of the signal is calculated for each data point, enabling the autoregressive moving average (ARMA) model to characterize the dynamic behaviour of the EEG time series. In the second part, classification has been done using pattern recognition technique and achieved an accuracy of 93% and 94%, respectively.

Zhenghui Gu et al. [16] proposed a method based on the epileptic characteristic waves extracted from training EEG signals. They have constructed a set of wavelets to perform CWT on the recorded EEG signal. Data fusion is applied to the CWT coefficient matrices corresponding to the multiple constructed wavelets to determine the boundaries of seizures.

A.Sharmila et al. [17] developed a system based on a DWT analysis of EEG signals using linear and non-linear classifiers to detect epileptic seizures. They used Naïve Bayes and K-NN classifiers to classify the statistical features obtained from DWT with an accuracy of 100%.

In [50], Coifman et al. introduced the wavelet packet transform (WPT) in which the link between wavelets and multi-resolution approximation was explained. The signals in a biomedical application generally consist of high-frequency components that are closely distributed in the frequency domain, which are also long-lasting, and the low-frequency components are closely distributed in the frequency domain. For the analysis of such types of signals, wavelets are appropriately suited as they show better frequency resolution as well as finite time resolution. To obtained better resolution in both frequency and time domain, these signals can be efficiently analyzed by using wavelets.

The difference between wavelet packet transform and wavelet transform (WT) is that in wavelet packet transform, the decomposition of both approximate and detailed coefficients takes place while in the case of the wavelet transform, the only decomposition of detailed coefficients takes place. The process of decomposition is done J times, where J is given by \( J \leq \log_2 N \). N is the total number of samples in the original signal. This will give \( J \times N \) coefficients. Hence, the tree at a resolution level of \( j \) will have \( N \) coefficients, which are divided into \( 2^j \) coefficient crystals or blocks, where \( j = 1, 2, \ldots, J \). A tree-like structure of binary WP is generated by this iterative process in which the node of a tree represents sub-spaces with distinct frequency localization components.

As we have discussed the different techniques to detect epilepsy in EEG signals, most of them have used entropy principles, S-transforms, Bi-LSTM networks, ARMA models, and basic CWT DWT analysis to perform the feature extraction tasks. While all these methods increase computational com-
plexity, the extracted features are not appropriately selected due to the dimensionality of the features. As the dimension of the features increases, the classification complexity also gets rigid. To overcome these drawbacks, more work is needed to be done on dimensionality reduction of epileptic features to obtain better results.

In the past, few researchers have implemented dimensionality reduction techniques to reduce computational complexity and increase the classification accuracy of the machine learning models. In [29], authors have used two dimensionality reduction techniques named Hilbert Transform and Hessian Local Linear Embedding (HLLE) to reduce the dimensions of epileptic features, which eventually increased the classification accuracy of the Adaboost classifier. In [30], authors deployed six dimensionality reduction algorithms to map the extracted features into other domains in the feature selection procedure. They have used "principal component analysis (PCA), independent component analysis (ICA), kernel principal component analysis (K-PCA), isometric feature mapping (ISOMAP), locally linear embedding (LLD) and Laplacian eigenmaps (LE)" to reduce the feature dimensions of epileptic electroencephalograph (EEG) signals. In [31], authors investigated two feature selection techniques based on mutual information and a wrapper method to reduce the complexity and computation costs of least-squares support vector machine (LS-SVM) and classify the nocturnal motor seizures in epileptic children from the normal nocturnal movements using accelerometer signals.

All the discussed state-of-art techniques are using all the extracted epileptic features from the EEG signals, which again increases the computation and complication costs of the classifiers. To overcome these complications, we have proposed two dimensionality reductions algorithms, namely SPPCA and SUBXPCA, to reduce the computation complexity of the machine learning classifiers. To get the higher accuracy while reducing the complexity to classify the epileptic seizures, we have extracted five time-frequency domain features using DWT and compared all of them using the proposed algorithm to depict the best suitable feature for detecting epilepsy in EEG signals.

III. PROPOSED METHODOLOGY

In this work, we have proposed an algorithm for epilepsy detection using different time-frequency domain features extracted from an EEG signal and in order to increase the accuracy of the machine learning models to detect epileptic seizures, we have developed two dimensionality reduction algorithm named SPPCA and SUBXPCA.

A. Data Description

The dataset used in our work is taken from the publicly available database of department of epileptology from university of Bonn, Germany [23]. The Dataset is comprised of five groups, namely A to E, each group contains 100 single-channel EEG signals, and each signal is of 23.6-sec in length with 4097 examples.

B. Decomposition of EEG signal using DWT

In this work, DWT is used to decompose the EEG signal into different frequency components in the time-frequency space [19]. DWT is not only an efficient way to analyse the signal, but it also reduces the computational time. In DWT, the given signal is decomposed into finer details by using two functions as elaborated in (1) [17].

\[ S(t) = \sum_{i \in \mathbb{Z}} 2^j c_j(k) \varphi \left( 2^j t - k \right) + \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} 2^j d_j(k) \psi \left( 2^j t - k \right) \tag{1} \]

Here, \( \varphi(t) \) is a basic scaling function, and \( \psi(t) \) mother wavelet. The first summation is an approximation of \( S(t) \) which depends on the scaling index \( j_0 \), and the second summation is adding more details and is varied with respect to the value of \( j \).

The WT can be acquired by using the blend of low and high pass filters to dissect the low and high frequency segments of the signal.

\[ c_j(k) = \int_{-\infty}^{\infty} S(t) \varphi \left( 2^j t - k \right) dt \tag{2} \]

\[ d_j(k) = \int_{-\infty}^{\infty} S(t) \psi \left( 2^j t - k \right) dt \tag{3} \]

Where, \( c_j \) and \( d_j \) are approximate and detail coefficients of the decomposed signal.

Firstly, the EEG signal is decomposed into a high-frequency component \( g(n) \) and the low-frequency component \( h(n) \) respectively. The approximation of the low-frequency component and the high-frequency component is obtained when the signal is down-sampled by two. This process is to be repeated until we acquire the desired decomposition level. In our proposed work, decomposition is performed till the fifth level. Secondly, we have selected half of the frequencies between 0Hz and data Nyquist frequency 173.6Hz as a
threshold frequency for decomposition. Hence, the EEG signals of frequency range (0–86.8)Hz are decomposed into approximate components (43.4–86.4)Hz and detail components (0–43.4)Hz. In the second-level of decomposition, the approximate components are again decomposed into two groups having a frequency between (0 – 21.7)Hz as approximate components and frequencies between (21.7 – 43.4)Hz as detailed components. This process is continued till the fifth-level of decomposition. After the decomposition process, the high-frequency component are excluded because most of them replicate the noise. Finally, after the fifth level of decomposition, the components we obtained are (21.7 – 43.4)Hz, (10.85 – 21.7)Hz, (5.425 – 10.85)Hz, (2.71 – 5.425)Hz and (0 – 2.71)Hz. All the five levels of frequency components extracted using DWT are shown in Fig. 1. After decomposing the EEG signal into different sub-bands, we have found that majority of the energy is distributed in delta and theta bands during seizures. To validate this claim, we have calculated various time-frequency domain features such as dominant frequency, spectral entropy, spike rhythmicity, and wavelet transform. Finally, our study found that energy of the EEG signal was giving higher accuracies with minimum computational time to detect epileptic seizures.

C. Time domain features extraction

In our work, we have considered three time domain features, energy of each sub-pattern, Spike Rhythmicity (SR) and Relative Spike Amplitude (RSA) [20]. After evaluating all three feature it is found that energy is giving best results when compared to the other features.

Energy of each sub-pattern can be calculated by summing the square value of each sub-pattern samples as given in (4). The graphical representation of energy is shown in Fig. 2(c).

\[
\text{Energy} (E_i) = \sum_{j=1}^{N} D_{ij}^2
\]  

(4)

where, D can be number of samples in each sub-pattern. The other two features were obtained by following the steps mentioned below:

- A desired length median filter is used to smoothen the EEG signal.
- Now, the difference signal is obtained by subtracting the filtered EEG signal from the original signal.
- The demarcation of the EEG signal from noise is done by enhancing the absolute value of the difference signal.
- Threshold value is set to 50% of the maximum spike amplitude. We can identify the number of spikes using this threshold value.

Finally, the SR value of a particular EEG segment can be evaluated as the number of spikes above the threshold value in that segment as shown in Fig. 2(d). Moreover, the RSA for a specific section is equivalent to the maximum amplitude value of all the peaks that crosses the threshold value.

D. Frequency domain features extraction

We have considered two frequency domain features, Dominant Frequency (DF) and Spectral Entropy (SE) [20]. Firstly, to get the frequency domain features, time domain signal need to be transformed into the frequency domain by using DFT as given in (5).

\[
S(k) = \sum_{n=0}^{N-1} s[n] e^{-j \frac{2\pi}{N} kn}
\]  

(5)

1) Dominant Frequency (DF): Dominant frequency is defined as the maximum peak frequency of each sub-domain. This feature is very useful in epilepsy detection as it provides better results. We have calculated the dominant frequency of each epoch of size 32 for an EEG signal as shown in Fig. 2(a).

2) Spectral Entropy (SE): Spectral entropy is calculated between two frequencies \( f_1 \) and \( f_2 \) for each sub-pattern. It can be formulated as given in (6).

\[
SE = \sum_{f_i=f_1}^{f_2} P_n(f_i) \times \log \left( \frac{1}{P_n(f_i)} \right)
\]

(6)

Where, \( f_i \) is the frequency component ranges from \( f_1 \) to \( f_2 \), and \( P_n(f_i) \) is the normalized power spectrum component at \( f_i \).

And then the normalized spectrum entropy (NSE) is obtained by dividing the SE by \( \log(N) \)

\[
NSE = \frac{SE}{\log(N)}
\]

(7)

SE of an EEG signal using 32-point FFT is shown in Fig. 2(b).
E. Feature Dimensionality Reduction using SPPCA and SUBXPCA

Principal Component Analysis (PCA) is generally used to reduce the dimension of large datasets into smaller datasets while still retaining most of the information of the larger dataset like features. To apply PCA on the given data of dimension $n \times m$, (where $n$ is the number of data and $m$ is the number of dimensions of each element), the following procedure is used:

- Firstly, the covariance matrix of size $(m \times m)$ is calculated from the extracted features.
- Then, calculate the eigenvalues ($\lambda^i$) and corresponding eigenvectors $E^i$ of the covariance matrix.
- Choose ‘$k$’ ($k < m$) largest eigenvalues and their corresponding eigenvectors. Let $E^i$ represent the set of these ‘$k$’ chosen eigenvectors.
- To obtain the PCs for feature vectors, projection of feature vectors $S^i$ onto $E^i$ is performed to obtain the features with reduced dimension as per the following relation given in (8).

$$Y_i = S_i \times E_i \tag{8}$$

In this work, we have used two variants of PCA to reduce the dimensionality of features [21]. The two algorithms are elaborated as:

1) SPPCA: In SPPCA, the signal is first divided into different equal sub-parts. We have divided an EEG signal into eight as they were giving the desired results. In the second step, PCA is applied to each sub-part of the signal, and $k$ number of Principal Components (PCs) are selected. Here, the $k$ should be less than the dimension of the signal. After the dimensionality reduction, all the sub-parts are horizontally concatenated to get appropriate global data. Finally, the reduced dimension signal is applied to the classifier.

2) SUBXPCA: This algorithm is obtained by incorporating some additional steps in the SPPCA algorithm. To obtain SUBXPCA, we need to select $t$ no of PCs from the SPPCA data and then project this data on to $t$ no of eigenvectors. As the processing steps in both the algorithms are approximately the same, there is a small difference in accuracy between SPPCA and SUBXPCA. The detailed process of the proposed algorithm is explained in Fig. 3. And the output EEG signal after dimensionality reduction using SPPCA and SUBXPCA is shown in Fig. 4.
TABLE I: Comparison between computational time of the proposed method with the other techniques

| Authors                        | Computational time (s) |
|--------------------------------|------------------------|
| Qu & Goatman [41]              | 9.60                   |
| Osrio et al. [42]              | 2.10                   |
| Grewal & Goatman [43]          | 17.1                   |
| Gardener et al. [44]           | 7.58                   |
| Aarahi et al. [45]             | 11.0                   |
| Zhang et al. [46]              | 10.8                   |
| Gatien Hocepied et al. [47]    | 14.8                   |
| N Arunkumar et al. [48]        | 1.14                   |
| P.T. Krishnan et al. [49]      | 0.48                   |
| Z. Chen et al. [15]            | 0.31                   |
| J-L. Song et al. [34]          | 7.71                   |
| Proposed SPPCA                 | 1.02                   |
| Proposed SUBXPCA               | 0.71                   |

F. Time complexity reduction using PCA, SPPCA and SUBXPCA

Let us consider $P_1, P_2, P_3, ..., P_N$ be the input patterns of class $P$ with a dimension $d$. By applying PCA, the time complexity can be determined as:

$$T(PCA) = O(Nd^2 + d^3)$$ (9)

In the proposed SPPCA and SUBXPCA algorithms, the input patterns are divided into $L$ subpatterns. So the time complexity of SPPCA is given as:

$$T(SPPCA) = O(L(Nr^2 + r^3))$$ (10)

For SUBXPCA, as one additional step is begun added the time complexity can be given as:

$$T(SUBXPCA) = O(L(Nr^2 + r^3)) + O(N(L.k)^2 + (L.k)^3)$$ (11)

From the above equations, it can be concluded that $T(SPPCA) \leq T(PCA)$ and $T(SUBXPCA) \leq T(PCA)$.

G. Classification using various machine learning models

For the classification purpose firstly, the dataset is divided into 10 folds by using K-fold method. Then for each model and performance metric, the hyper-parameter tuning was done by GridSearchCV. As per the received parameters, we calculated performance metrics for the fitted models and aggregated them into the DataFrame. Finally, we accumulated performance metrics of all folds from the same dataset in order to get their average.

In this paper we are using several models such as:

- K-Nearest Neighbor based on dividing dataset on classes according to distances between points of classes.
- Random forest consisting of an ensemble of trees, in which each tree gives a prediction of the class, and the class with the most votes becomes the prediction of our model.
- Multi-layer Perceptron (MLP) is a supervised learning algorithm, consisting of three or more levels of nodes.
- Light GBM and CatBoost both based on gradient boosted decision trees.

For the better visualisation of the accuracies of different features extracted using SPPCA and SUBXPCA are graphically represented in Fig. 5 and Fig. 6.

IV. Results

The results highlighted in this paper represents the accuracies of various features related to epilepsy detection using SPPCA and SUBXPCA algorithms. The samples in the datasets were randomly divided into training and testing sets; we have designated 70% data for training and 30% for testing.

As we have extracted the different features related to epilepsy and reduced the dimension of the data, as well as to extract the correlation features using SPPCA and SUBXPCA algorithms, we divide these feature vectors into $L$ different portions. The value of $L$ can be any value less than the dimension of feature vectors. In our experiment, it is found that $L = 8$ gives satisfactory results. Then we are selecting $p$-numbers of PCs from each sub-portion. Here, the value of $p$ decides the accuracy of the proposed technique. In this work,
we are using 10-fold cross-validation for calculating various classification parameters in different cases.

We have evaluated the performance by considering seven combinations of the datasets: (A-E), (B-E), (C-E), (D-E), (AB-E), (CD-E), (ABCD-E). Based on the epileptic features, accuracy, precision, recall, and F1 score are calculated for SPPCA and SUBXPCA algorithms by using various machine learning algorithms such as SVM, KNN, Random Forest, Multilayer Perceptron (MLP), LightGBM, and CatBoost; acquired results are shown in Table I. After computing the average of all the parameters, we have noted the accuracy, recall, and precision of SPPCA algorithm as 94%, 90%, and 98% respectively. Similarly, for SUBXPCA the computed average values are 96%, 93%, and 98% respectively. Since one model should be superior for SPPCA and one for SUBXPCA, we rank models in each dataset and choose best model according to places in their ranks. The best model for SPPCA is CatBoost with default parameters, the best model for SUBXPCA is Random Forest classifier.
V. DISCUSSION

As of late, several automated seizure detection frameworks using machine learning (ML) algorithms have been proposed to supplant conventional methods. In this study, we have compared the performance of our proposed method with existing approaches for the classification of a normal, epileptic seizure, and seizure-free classes from the EEG signal, as illustrated in Table III. In the recent study, J. A. de la O Serna et al. [24] has developed an "EEG-rhythm specific Taylor-Fourier filter-bank implemented with O-splines for the detection and classification of epilepsy from the EEG signal." In this work KNN and least square SVM were employed to obtained an accuracy of 94.88% respectively. Similarly, U. R. Acharya et al. [25] have used "Entropy features for epilepsy detection and D-CNN for the classification of epilepsy in EEG signals and obtained an accuracy of 88.70%.

Moreover, P. Li et al. [26] have extracted various entropy features from the EEG signal as a part of FE, and SVM is used for the classification between seizure-free and seizure classes while they obtained an accuracy of 93%. Similarly, L. C. Amarantidis and D. Ab ‘asolo [27] have also used various entropy measures to extract the epileptic features from the EEG signals and they have utilised KNN classifier for the detection of seizures and achieved an accuracy of 73.50%. In another study, W. Mardini et al. [13] have extracted the features from the EEG signal of fourteen different combinations of datasets using 54 DWT mother wavelets. They obtained an accuracy of 97.82% with ANN, 97.58% with KNN, 97.15% with SVM and 97.32% with NB classifiers respectively. Similarly, Z. Chen et al. [15] have extracted features using EMD and performed classification using pairwise one-class SVMs on the publicly-available Bern-Barcelona and CHB-MIT EEG database and achieved an accuracy of 94%. Moreover, D. Wu et al. [28] have extracted Multi-domain features and classified them using Random forest (RF) classifier. Here the author has experimented on CHB-MIT dataset, the accuracy obtained by the hybrid method was 99.36%.

In the proposed work, more emphasis has been given to develop a machine learning model using the time-frequency domain features extracted from the EEG signal. And we have proposed dimensionality reduction algorithms namely SPPCA and SUBXPCA to reduce the feature dimensions of the seizures. Furthermore, for classification purpose, we have used an six different ML models to validate the effectiveness of the proposed framework which obtained satisfactory results compared to [13] [15] [24]- [28] [34] [35] [36] [37] [38] [39] [40] and achieved an accuracy of 97%. The detailed comparison of all the classification parameters are compared in Table III.

The main advantage of our work is to identify epileptic seizures from the EEG signals with less computation time and high accuracy. In the present scenario, complex observation has to be performed by the neurologist. This complexity is gradually reduced as we have used dimensionality reduction algorithms (SPPCA and SUBXPCA), which reduces the computational time to detect seizures, reduces human errors while studying EEG signals, increases the accuracy of machine learning models and reduces the burden on the physician while detecting epilepsy from the EEG signals. And, the disadvantages are not all features can be detected apart from epilepsy using this method. In our future work, we would be developing a new algorithm to perform on other channels of EEG signals to detect other problems (i.e., seizures, head injuries, dizziness, headaches, brain tumors and sleeping problems) as well.

VI. CONCLUSION

In this work, we propose a novel approach to deal with the detection of epileptic seizures present in EEG signals using SPPCA and SUBXPCA dimensionality reduction algorithms to improve the accuracy of SVM classifier. The experimental outcomes indicated that the time-frequency domain features extracted using DWT when combined with some machine learning algorithms had given recognizable outcomes. The new proposed framework can reduce the dimensions of the features with less computational complexity and can detect seizures with high accuracy. In our work, we extracted time-frequency domain features using DWT, and afterwards, the dimensions of these features are reduced using SPPCA and SUBXPCA algorithms. The features extracted at this stage are later used for epileptic seizure classification using various machine learning models. The effectiveness of the proposed technique is verified using 7 different combinations of datasets. The outputs acquired by our proposed work achieved good results in terms of performance metrics. Hence, it can be concluded that features extracted through DWT and reduced using SPPCA and SUBXPCA gives satisfactory results when compared with the latest state-of-art works. In terms of classification, the proposed SPPCA algorithm combined with CatBoost classifier achieved the best accuracy of 97.0% and SUBXPCA combined with Random forest achieved an accuracy of 98.5% when compared with different classifiers.

ACKNOWLEDGEMENT

The authors thank Prof. Vincent Jacquemet (Université de Montréal, Canada) for his valuable comments and suggestions on our work.
REFERENCES

[1] Mehndiratta, Man Mohan, and Swati Anil Wadhai. “International Epilepsy Day - A day notified for global public education & awareness.” The Indian journal of medical research vol. 141,2 (2015): 143-4.
doi:10.4103/0971-5916.155531

[2] A. Hamad, E. H. Houssein, A. E. Hassanien and A. F. Mahy, "Feature extraction of epilepsy EEG using discrete wavelet transform", Proc. 12th Int. Comput. Eng. Conf. (ICENCE), pp. 190-195, Dec. 2016.
[3] U. R. Acharya, S. Vinitha Sree, G. Swapna, R. J. Martis and J. S. Suri, "Automated EEG analysis of epilepsy: A review", Knowl.-Based. Syst., vol. 45, pp. 147-165, Jun. 2013.
[4] P. Sarma, P. Tripathi, M. P. Sarma and K. K. Sarma, "Pre-processing and feature extraction techniques for EEGBCI applications-a review of recent research", ADBU J. Eng. Technol., vol. 5, no. 1, pp. 1-8, 2016.
[5] C. Umale, A. Vaidya, S. Shirude and A. Raut, "Feature extraction techniques and classification algorithms for EEG signals to detect human stress-a review", Int. J. Comput. Appl. Technol. Res., vol. 5, no. 1, pp. 8-14, Jan. 2016.
[6] K. Polat and S. Güneş, "Artificial immune recognition system with fuzzy resource allocation mechanism classifier principal component analysis and FFT method based new hybrid automated identification system for classification of EEG signals", Expert Syst. Appl., vol. 34, no. 3, pp. 2039-2048, Apr. 2008.
[7] T. K. Gandhi, "Classification of EEG signal using wavelet transform and ant colony classifier", Proc. IEEE Int. Conf. Commun. (ICC), pp. 3529-3534, Jun. 2014.
[8] S. A. Aljawarneh, P. Radhakrishna and A. Cheruvu, “VRKSHA: A novel tree structure for time-profiled temporal association mining” in Neural Computing and Applications, Cham, Switzerland: Springer, pp. 1-29, 2017, [Online] Available: https://link.springer.com/article/10.1007%2Fs00521-018-3776-7
[9] V. H. Chiang, M. Cheng and Y. Huang, “Wavelet-based EEG Processing for Epilepsy Detection Using Fuzzy Entropy and Associative Petri Net,” in IEEE Access, vol. 7, pp. 103255-103262, 2019, doi:10.1109/ACCESS.2019.2929266.
[10] W. Mardini, M. M. Bani Yassein, R. Alrawshdeh, S. Aljawarneh, Y. Khamayseh, R. Vangipuram, V. K. Puligadda and J. Vinjamuri, “ASTRA—A novel tree structure for time-profiled temporal association mining” in Neural Computing and Applications, Cham, Switzerland: Springer, pp. 1-29, 2017, [Online] Available: https://link.springer.com/article/10.1007%2Fs00521-018-3776-7
[11] V. Radhakrishna, S. A. Aljawarneh, P. Veereswara Kumar and V. Janaki, “ASTRA—A novel interest measure for unearthing latent temporal associations and trends through extending basic Gaussian membership function”, Multimedia Tools Appl., vol. 78, no. 4, pp. 4217-4265, Feb. 2019.
[12] V. Radhakrishna, S. A. Aljawarneh, V. Janaki and P. Kumar, “Looking into the possibility for designing normal distribution based dissimilarity measure to discover time profiled association patterns”, Proc. Int. Conf. Soft. Eng. and MIS (ICEMIS), pp. 1-5, May 2017.
[13] H. Chiang, M. Chen and Y. Huang, “Wavelet-Based EEG Processing for Epilepsy Detection Using Fuzzy Entropy and Associative Petri Net,” in IEEE Access, vol. 7, pp. 103255-103262, 2019, doi:10.1109/ACCESS.2019.2929266.
[14] W. Mardini, M. M. Bani Yassein, R. Alrawshdeh, S. Aljawarneh, Y. Khamayseh and O. Meqdadi, “Enhanced Detection of Epileptic Seizure Using Fuzzy Entropy Algorithm,” in IEEE Access, vol. 7, pp. 103255-103262, 2019, doi:10.1109/ACCESS.2019.2904949.
[15] Prabhakar, S. & Raguraju, Harikumar. (2018). Adaboost Classifier with Dimensionality Reduction Techniques for Epilepsy Classification from EEG. 10.1007/978-981-10-6778-4_31.
[16] Milošević, M., Van de Vel, A., Cappelletti, M. et al. Feature selection methods for acceleration-based seizure detection in children. Med Biol Eng Comput 55, 151–165 (2017). https://doi.org/10.1007/s11517-016-1506-9
[17] J. A. de la O Serna, M. R. A. Paternina, A. Zamora-Méndez, R. K. Tripathy and R. B. Pachori, “EEG-Rhythm Specific Taylor–Fourier Filter Bank Implemented With O-Splines for the Detection of Epilepsy Using EEG Signals,” in IEEE Sensors Journal, vol. 20, no. 12, pp. 6542-6551, 15 June 2020, doi: 10.1109/JSEN.2020.2976519.
[18] A. Hamad, E. H. Houssein, A. E. Hassanien and A. A. Fahmy, “Feature extraction of epilepsy EEG using discrete wavelet transform and ant colony classifier”, Proc. IEEE Int. Conf. Commun. (ICC), pp. 3529-3534, Jun. 2014.
[19] A. Hamad, E. H. Houssein, A. E. Hassanien and A. A. Fahmy, “Feature extraction of epilepsy EEG using discrete wavelet transform and ant colony classifier”, Proc. IEEE Int. Conf. Commun. (ICC), pp. 3529-3534, Jun. 2014.
[20] P. Li, C. Karmakar, J. Yearwood, S. Venkatesh, M. Palaniswami, and C. Liu, “Detection of epileptic seizure based on entropy analysis of short-term EEG,” PloS one, vol. 13, no. 3, p. e0193691, 2018.
[21] L. C. Amarantidis and D. Ab’asolto, “Interpretation of entropy algorithms in the context of biomedical signal analysis and their application to EEG analysis in epilepsy.”, Entropy, vol. 21, no. 9, p. 840, 2019.
[22] S. A. Aljawarneh, P. Radhakrishna and A. Cheruvu, “VRKSHA: A novel tree structure for time-profiled temporal association mining” in Neural Computing and Applications, Cham, Switzerland: Springer, pp. 1-29, 2017, [Online] Available: https://link.springer.com/article/10.1007%2Fs00521-018-3776-7
[23] V. H. Chiang, M. Chen and Y. Huang, “Wavelet-Based EEG Processing for Epilepsy Detection Using Fuzzy Entropy and Associative Petri Net,” in IEEE Access, vol. 7, pp. 103255-103262, 2019, doi:10.1109/ACCESS.2019.2929266.
[24] W. Mardini, M. M. Bani Yassein, R. Alrawshdeh, S. Aljawarneh, Y. Khamayseh and O. Meqdadi, “Enhanced Detection of Epileptic Seizure Using EEG Signals in Combination With Machine Learning Classifiers,” in IEEE Access, vol. 8, pp. 24046-24055, 2020, doi: 10.1109/ACCESS.2020.2970012.
[25] M. Geng, W. Zhou, G. Liu, C. Li and Y. Zhang, “Epileptic Seizure Detection Based on Stockwell Transform and Bidirectional Long Short-Term Memory,” in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 28, no. 3, pp. 573-580, March 2020, doi: 10.1109/TNSRE.2020.2969290.
[26] Z. Chen, G. Lu, Z. Xie and W. Shang, “A Unified Framework and Method for EEG-Based Early Epileptic Seizure Detection and Epilepsy Diagnosis,” in IEEE Access, vol. 8, pp. 20080-20092, 2020, doi: 10.1109/ACCESS.2020.2969055.
[27] Z. Gu, G. Yan, J. Zhang, Y. Li and Z. L. Yu, "Automatic Epilepsy Detection Based on Wavesets Constructed From Data," in IEEE Access, vol. 6, pp. 53133-53140, 2018, doi: 10.1109/ACCESS.2018.2887662.
[28] A. Sharma and P. Geethanjali, "DWT Based Detection of Epileptic Seizure From EEG Signals Using Naive Bayes and k-NN Classifiers," in IEEE Access, vol. 4, pp. 7716-7727, 2016, doi:10.1109/ACCESS.2016.2585661.
[29] R. Panda, P. S. Khobragade, P. D. Jambhule, S. N. Jengte, P. R. Pal and T. K. Gandhi, "Classification of EEG signal using wavelet transform and support vector machine for epileptic seizure detection," 2010 International Conference on Systems in Medicine and Biology, Kharagpur, 2010, pp. 405-408, doi: 10.1109/ICSMB.2010.5753541.
of intrinsic mode functions, Expert Systems with Applications, Volume 42, Issue 3, 2015, Pages 1106-1117, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2014.08.030.

[39] A. K. Tiwari, R. B. Pachori, V. Kanhangad and B. K. Panigrahi, "Automated Diagnosis of Epilepsy Using Key-Point-Based Local Binary Pattern of EEG Signals," in IEEE Journal of Biomedical and Health Informatics, vol. 21, no. 4, pp. 888-896, July 2017, doi: 10.1109/JBHI.2016.2589971.

[40] Ahnaf Rashik Hassan, Abdulhamit Subasi, Automatic identification of epileptic seizures from EEG signals using linear programming boosting, Computer Methods and Programs in Biomedicine, Volume 136, 2016, Pages 65-77, ISSN 0169-2607, https://doi.org/10.1016/j.cmpb.2016.08.013.

[41] H. Qu et al., "A seizure warning system for long-term epilepsy monitoring", Neuroil, vol. 45, no. 12, pp. 2250-2254, 1995.

[42] I. Osorio et al., "Real-time automated detection and quantitative analysis of seizures and short-term prediction of clinical onset", Epilepsia, vol. 39, no. 6, pp. 615-627, 1998.

[43] S. Grewal et al., "An automatic warning system for epileptic seizures recorded on intracerebral EEGs", Clin. Neurophysiol., vol. 116, no. 10, pp. 2460-2472, 2005.

[44] A. B. Gardner et al., "One-class novelty detection for seizure analysis from intracranial EEG", J. Mach. Learn. Res., vol. 7, pp. 1025-1044, 2006.

[45] A. Aarabi et al., "A fuzzy rule-based system for epileptic seizure detection in intracranial EEG", Clin. Neurophysiol., vol. 120, no. 9, pp. 1648-1657, 2009.

[46] Y. Zhang et al., "An automatic patient-specific seizure onset detection method in intracranial EEG based on incremental nonlinear dimensionality reduction", Comput. Biol. Med., vol. 40, no. 11-12, pp. 889-899, 2010.

[47] G. Hocepid et al., "Early detection of epileptic seizures based on parameter identification of neural mass model", Comput. Biol. Med., vol. 43, pp. 1773-1782, 2013.

[48] N. Arunkumar, K. Ram Kumar and V. Venkataraman, "Entropy features for focal EEG and non focal EEG", J. Comput. Sci., vol. 27, pp. 440-444, Jul. 2018.

[49] P. T. Krishna and P. Balasubramanian, "Automated EEG seizure detection based on S-transform", Proc. IEEE Int. Conf. Comput. Intell. Comput. Res. (ICCC), pp. 1-5, Dec. 2016.

[50] Coifman, Ronald & Meyer, Y. & Wickerhauser, Mladen. (1992). Wavelet analysis and signal processing,” in: Wavelets and Their Applications. Jones and Bartlett. 153-178.

Muhammad Zubair (Student Member, IEEE) is presently pursuing bachelor’s degree in electronics and communication engineering at Prasad V. Potluri Siddhartha Institute of Technology, Vijayawada, India; affiliated to JNTU Kakinada, Andhra Pradesh, India. He is expecting to complete his bachelor’s by July 2021.

Marina Belykh received her bachelor’s degree in mathematics from the Ural Federal University, Yekaterinburg, Russia, in 2018. She is currently pursuing a Master’s degree in Applied Mathematics from the National Research University Higher School of Economics. Her research interests are in Machine Learning, Computer Vision, Deep Learning, Scientific Visualization.

Umesh Kumar Naik M is presently pursuing bachelor’s degree in electronics and communication engineering at Prasad V. Potluri Siddhartha Institute of Technology, Vijayawada, India; affiliated to JNTU Kakinada, Andhra Pradesh, India.

Mohammad Fareeda Madeen Gouher is an undergraduate student pursuing bachelor’s in electronics and communication engineering at Prasad V. Potluri Siddhartha Institute of Technology, Vijayawada, India. She’s an AI and ML enthusiast. Her fields of interest are Medical Imaging, Computer Vision, and Machine Learning.

Shani Vishwakarma Shani Vishwakarma graduated in Electrical Engineering from Shri Shankaracharya Group of Institutions, Bhilai in 2017, and received his Masters in Electronics and Electrical Engineering with Signal Processing specialisation from IIT Guwahati, India in 2020. He is currently working as Digital Engineer in Tata Digital Mumbai.

Shaik Rafi Ahamed (Member, IEEE) received his Ph.D. degree from IIT Kharagpur, India, in 2008. He is currently a Professor with the Department of Electronics and Electrical Engineering, IIT Guwahati, India. His teaching and research interests include digital, adaptive, biomedical, and VLSI signal processing and machine learning.

Ramanjaneyulu Kongara (Member, IEEE) received his PhD from Andhra University in 2012. He is working as a Professor in the Department of ECE, PVP Siddhartha Institute of Technology, Vijayawada. His research interests are in the areas of Digital Watermarking, Image Processing and its Applications, Artificial Intelligence and Wireless communications.