An Improved Optimization Algorithm for Epileptic Seizure Detection in EEG Signals Using Random Forest Classifier

A. Phraeson Gini
Research Scholar, Department of EEE, Noorul Islam Centre for Higher Education, Noorul Islam University, Kanyakumari, India.
E-mail: Pingmehere2020@gmail.com

Dr.M.P. Flower Queen
Supervisor, Associate Professor, Department of EEE, Noorul Islam Centre for Higher Education, Noorul Islam University, Kanyakumari, India.
E-mail: mpflowerqueen@gmail.com

Received March 11, 2021; Accepted June 28, 2021
ISSN: 1735-188X
DOI: 10.14704/WEB/V18SI04/WEB18132

Abstract

Epilepsy is a psychiatric condition that has serious consequences for the human brain. The Electroencephalogram **(**EEG) may reveal a pattern that tells physicians whether an epileptic seizure is likely to occur again. EEG testing may also help the physician exclude other conditions that mimic epilepsy as a reason for the seizure. Now-a-days the researchers are showing much interest in these seizure detection because of its significance in epileptic detection. This paper is addressing an efficient soft computing framework for seizure detection from the EEG signal. The proposed pipeline of work is having the state-of-art as the possibility of achieving the maximum accuracy. The spectral features extracted from the Intrinsic mode functions (IMF) of EEG samples and it is directing the proposed flow towards the efficient detection of seizure and also the random forest algorithm based a convulsion classification is reliable for because of its learning behavior from the huge number of known dataset. The feature selection algorithm in this proposed work is stimulating the overall work towards the maximum true positive rate. This work is implemented on MATLAB platform and dataset were downloaded from the universal database such as Bonn university database. The results obtained from the proposed approach is showing the truthfulness of the approach introduced here.

Keywords

EEG, IMF, Seizure, Feature Selection, Forest at Random.

Introduction

A seizure is an uncontrollable electrical discharge disturbance in the brain. It can cause changes in your stages of consciousness, as well as in actions, gestures, or feelings. Epilepsy is diagnosed when you have two or more seizures or a history of repeated seizures. Seizures come in a variety of forms and severity levels may be described as Epileptic or Non-Epileptic seizure. Epileptic: These seizures have no apparent trigger and they occur two or more times. Epileptic seizures are called a seizure disorder or epilepsy. What causes epileptic seizures is often unknown. But they may be caused by various brain disorders, such as structural abnormalities, strokes, or tumors. In such cases, they are called symptomatic epilepsy. Non-Epileptic: These seizures are triggered by a reversible disorder or a temporary condition that irritates the brain, such as an infection, a head injury, or a reaction to a drug. Most seizures last from 30 seconds to two minutes. A seizure that lasts longer than five minutes is a medical emergency. Seizures can happen after a stroke, a closed head injury, and an infection such as meningitis or another illness. Many times, though, the cause of a seizure is unknown. Most seizure disorders can be controlled with medication, but management of seizures can still have a major influence on your day-to-day life. The good news is that you will collaborate with your doctor professional to balance seizure control and medication side effects. Sang. W et.al (2020).

The Electroencephalogram **(**EEG) may reveal a pattern that tells physicians whether a seizure is likely to occur again. EEG testing may also help your physician exclude other conditions that mimic epilepsy as a reason for your seizure. A trend in Healthcare is moving away from clinician-centered care and toward patient-centered care, in which the patient takes an active role in her own care management. Guo. V et.al (2020).

The following is a breakdown of the paper's structure: the section II is discussing the various techniques implemented in seizure detection. Section III introduces and explains the proposed methodology in seizure detection. Section IV demonstrate the results and performance analysis. Conclusion and future work are stated in Section V.

Related Works

The signal strength of each data point of the given EEG is first determined in Hanosh et.al (2019), allowing the well - known autoregressive moving average (ARMA) model to describe the dynamic behavior of the EEG time series. Even though the approach provides good results but It should be remembered that computational efficiency is an important a problem with the framework for diagnosing epilepsy.

The author Yuan et.al (2019) presented Based on complexity steps, a novel automatic detection method to differentiate between intracranial EEG time courses with seizures and those that are seizure-free. The features used to describe the EEG signals are generalized hurst exponent estimates of multi-scaling properties over a broad spectrum. To capture brain abnormalities, a unified multi-view deep learning system was developed implemented in Zheng et.al (2019).

The usefulness of Passive Infra-Red (PIR) sensors to detect human body motion during sleep triggered by epileptic seizures was investigated in Sony. J (2019). The authors Paul. Y (2018) developed a remote wireless control system focused on a single wrist-worn device accelerometer device that can diagnose seizures with a limited duration and is responsive to various forms of convulsive seizures. The use of Intracortical Micro-Electrode Early identification of human epileptic seizures using Array (MEA) signals was implemented in Wang et. al (2019) and the signal classification was done with the assistance of a vector support machine classification.

The hybrid model that comprise the classifiers used are K-Nearest Neighbor, a Radial Basis Support Vector Machine, and a Linear discriminant Analysis in seizure classification in Shumuel, A (2018). The Canonical polyadic Decomposition (CPD) and Block Term Decomposition (BTD) of the EEG data expressed as a third order tensor yield the classifier features. Wavelet or Hilbert-Huang transforms are used to convert the EEG into a tensor. The authors Zhu et.al (2019) used the evolution of model parameters to detect the starting point of the seizure in a D&F-model-driven approach to solving the problem of detecting early epileptic seizures. It is normally diagnosed clinically by experienced physicians using continuous electroencephalography (EEG) Particularly for seasoned doctors, this is time-consuming. Meanwhile, amplitude integrated electroencephalography (AIE) is being studied (EEG) has shown potential to find out whether you're having an epileptic seizure In Van Huffe S (2018) algorithm for hybrid seizure detection algorithm for hybrid seizure detection was proposed by Detecting seizures by combining a EEG-based seizure detection algorithm with an EEG-based seizure detection algorithm. The paper O. Berian (2018) was implemented with a low-energy Ada Boost cascade system for long-term seizure stability and a seizure candidate detection stage efficiently classified the given EEG signal.

In this paper, an efficient soft computing framework for seizure detection from the EEG signal was addressed. The proposed pipeline of work is having the state-of-art as the possibility of achieving the maximum accuracy. The a low-energy AdaBoost cascade system for long-term seizure stability of EEG samples and it is directing the proposed

flow towards the efficient detection of seizure and also the random forest algorithm based seizure classification is reliable for because of its learning behavior from the huge number of known dataset. The feature selection algorithm in this proposed work is stimulating the overall work towards the maximum true positive rate.

Methodology

The main frame work of the seizure detection workflow is shown in the following figure1. The proposed approach is having four important stages.

(i) Preparation of materials
(ii) Extraction of features
(iii) Range of features
(iv) Seizure detection

The detail of these subsections are briefed in below.

(i) Preprocessing

The EEG signals usually have some undesired artifacts as a part due to the interference phenomenon. This artifacts are unwanted noises that should be eliminated before the signals are being involved with seizure detection process. The Least-squares filter (LMS) is a This type of adaptive filter is used to replicate a desired filter by determining the filter coefficients that are related to generating the least mean squared error square of the signal of error It's a stochastic gradient descent approach in which the filter is only adjusted in response to the current error. The unwanted noise / artifacts can be eliminated at this stage and the pure EEG signal will allowed for the next process.

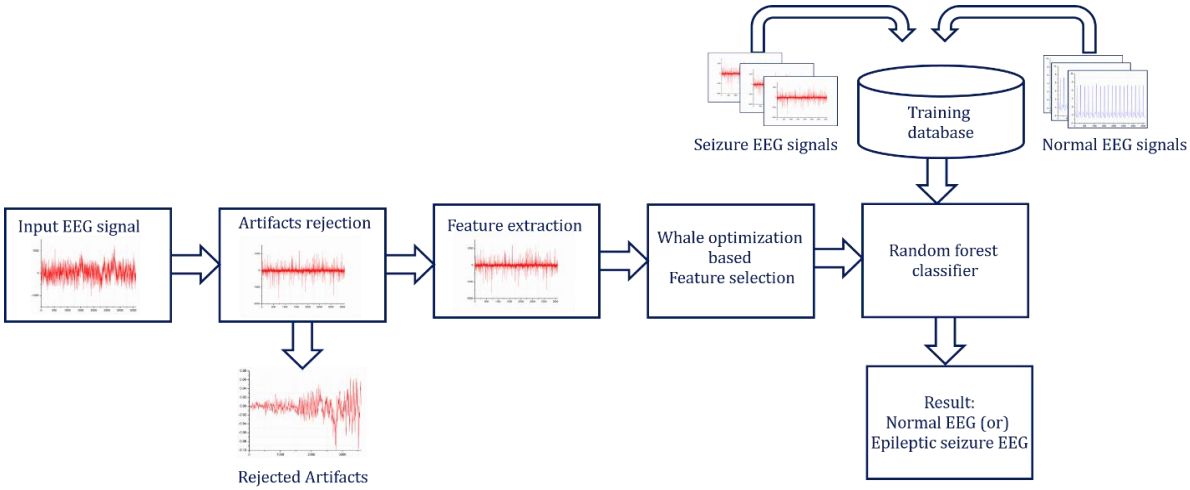

Figure 1 Pipeline of proposed work

(ii) Feature Extraction

One important strength of the proposed approach is that it is capable of conducting a spectral analysis of the signals in feature extraction phase. The features / attributes of the EEG are extracted from the decomposed result of Empirical Mode Decomposition. EMD is a method of decomposing a signal into a variety of oscillatory components known as intrinsic mode functions that is data based. (IMFs). The spectral features obtained from IMFs can thus give a wealth of knowledge about the physiology of the EEG signals. Visual analysis of this EEG spectrum shows that the statistics can be used as relevant features.

(a) Spectral Centroid

The centroid frequencies of the extracted from EEG When supervised clustering is used on the signals, they form distinct classes. These respective groups are indicative of the seizure and non-seizure EEG signals. The centroid frequency is therefore a distinctive feature that can be used for the characterization of EEG signals.

$$C_s = \frac{\sum_\omega \omega P(\omega)}{\sum_\omega P(\omega)} \qquad (1)$$

Where the P() is the frequency bin's amplitude. in the spectrum.

(b) Variation Coefficient

The variation of normal and seizure EEG signals gets varied so it can be used for their characterization. This variation coefficient can be calculated as follows:

$$\sigma_s^2 = \frac{\sum_\omega (\omega - C_s)^2 P(\omega)}{\sum_\omega P(\omega)} \quad (2)$$

Where C_s is the spectral centroid.

(c) Spectral Skew

Skewness is the third order moment and it measures the symmetry/asymmetry of a distribution. The skewness of the power spectrum can discriminate normal signals from seizure signals. It can be calculated as

$$\beta_s = \frac{\sum_\omega \left(\frac{\omega - C_s}{\sigma_s}\right)^3 P(\omega)}{\sum_\omega P(\omega)} \quad (3)$$

These three spectral features are extracted from the extracted EEG signals where the feature metric is framed for each signal.

(iii) Feature Selection

In a wrapper-based approach, the whale optimization algorithm (WOA) is used for feature selection. The principal characteristic of the wrapper-based methodology is to apply the classification approach as a method for selecting features based on an optimised component; a feature set that has been chosen. This is a paper about applies the Random forest of classification to maintain the consistency of the chosen feature set (15) The WOA is used adaptively in this paper to find the best function subset that maximises the return on investment. classification performance. The whales in WOA constantly change their positions to every point in space, starting with the best search agent. Then, as seen in equations (4) and (5), they work to update their positions against the best search agent. In the dataset, each solution is represented as a continuous vector with the same dimension. The values of the solution vectors are continuous and bounded. to the range [0, 1]. The values of the solution fitness assessment are represented as binary. In most cases, the fitness function used is the same as the one used in the previous step. classification quality, as well as the number of features chosen. The relevant equation for this WOA is expressed below.

$$\vec{D} = \left|\vec{C}.\overrightarrow{X^*}(t) - \vec{X}(t)\right| \qquad (4)$$

$$\vec{X}(t+1) = \overrightarrow{X^*}(t) - \vec{A}.\vec{D} \qquad (5)$$

$$\vec{X}(t+1) = \overrightarrow{X_{rand}} - \vec{A}.\vec{D} \qquad (6)$$

Where the $\vec{A}\&\vec{C}$ are the vectors of coefficients and $\vec{X}\&\overrightarrow{X^*}$ are the vector of position and best location vector respectively.

```
1) Initialize the whales population Xi (i = 1, 2, 3, ..., n).
2) Compute the fitness of each whale.
3) Set X* as the best whale.
4) while (t < maximum number of iterations) do
        for (each search whale) do
                Update a, A, C, l and p.
                        if (p < 0.5) then
                                if (|A| < 1) then
                                        The whale position is updating by the Eq. (4).
                                        else
                                if (|A| ≥ 1) then
                                        Select the random whale Xrand.
                                        The whale position is updating by the Eq. (6).
                                end
                        end
                else
                        if (p ≥ 0.5) then
                                Modify the whale position by the Eq. (5).
                        end
                end
        end
        Check if any search agent goes beyond the search space and amend it.
        Compute the fitness of each search agent.
        Update X* if there is a better solution.
        t = t + 1
end
```

(iv) Seizure Detection

This project uses a random forest classifier to detect seizures in EEG signals. Random forests, also known as random decision forests, are a form of ensemble learning system that can be used for classification and regression and other tasks in which a large number of decision trees are built during training and the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees is output. Random decision forests correct the tendency of decision trees to overfit their training collection.

Around one-third of the cases are left out of the study when the training set for the current tree is drawn using sampling with replacement. As trees are introduced to the forest, this OOB (out-of-bag) data is used to get a running unbiased estimate of the classification error. It's also used to calculate variable importance estimates.

After each tree is constructed, all of the data is run down the tree, and proximities for each pair of cases are computed. For two cases share a terminal node, their distance is increased by one. At the conclusion of the book the distances are normalised by dividing by the number of trees in the race. Proximities are used to replace missing data, find outliers, and create illuminating low-dimensional data views. The seizure detection wants very clear and transparent procedure where this random forest is most suitable for it.

Results and Discussion

(i) Dataset

The information analysed in this research is globally obtainable dataset as Bonn university database (Guo et al., 2020). The data was sampled at a rate of 173.61 Hz. Please see the manuscript for a more detailed summary of the details. Please keep in mind that the time series has a spectral component bandwidth of the acquisition device, which ranges from 0.5 to 85 Hz. Since the application of a 40 Hz low-pass filter, as defined in the manuscript, is considered the first stage of study, it is not performed on the downloadable time series.

(ii) Simulation Results

The figure 2 shows the plot of a normal Electroencephalogram (EEG) recorded in conjunction with help the electroencephalogram (EEG) measuring setup. The frequency and the amplitude are the notable things here and both are uniform and smooth.

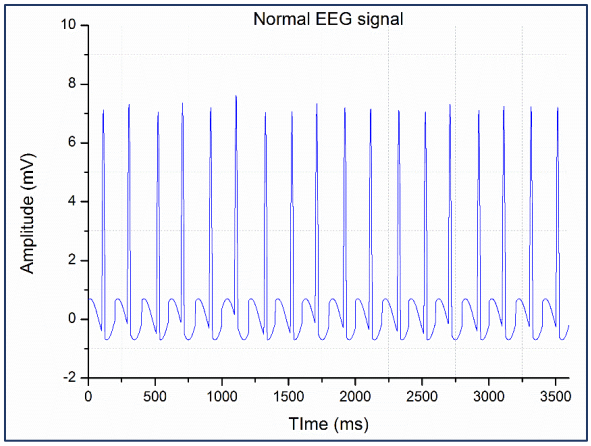

Figure 2 Normal EEG signal

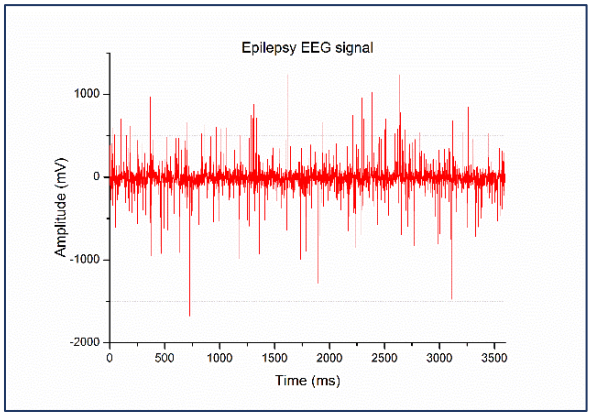

Figure 3 Seizure EEG signal

The above the electroencephalogram (EEG) signal is having it is unwanted artifacts and the noises are eliminated in preprocessing stage shown in figure 4. The response of the preprocessing stage is shown in the following figure where the artifacts was separated from the given input signal.

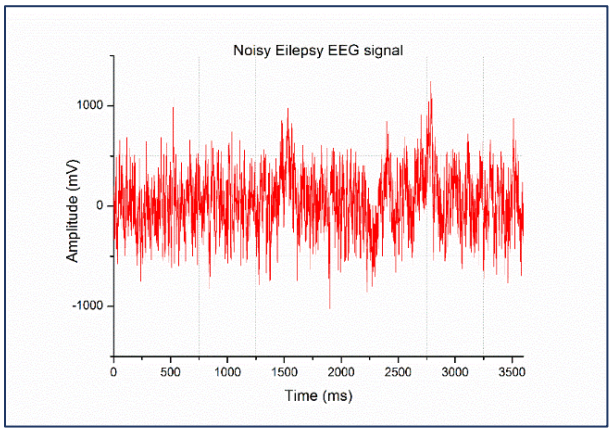

Figure 4 Eliminated artifacts signal

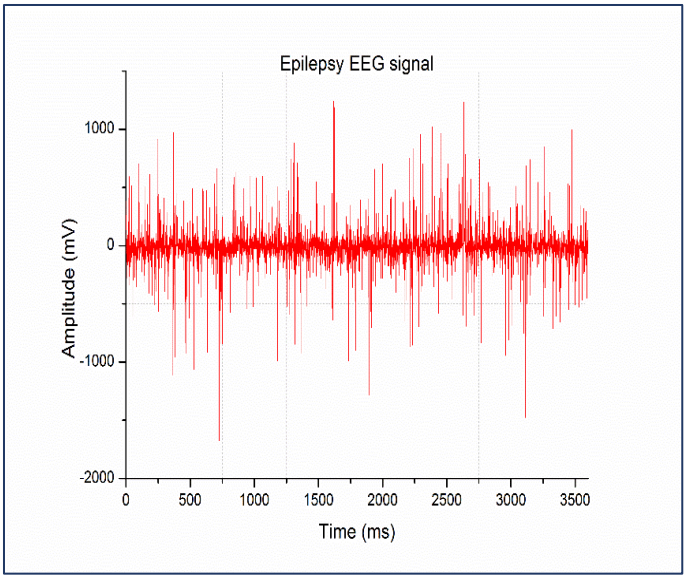

Figure 5 Filtered seizure EEG signal

The features extraction phase is done by extracting spectral features from the 5 different IMF signals of a test EEG. The figure 6 shows the 5 IMF of test EEG signal.

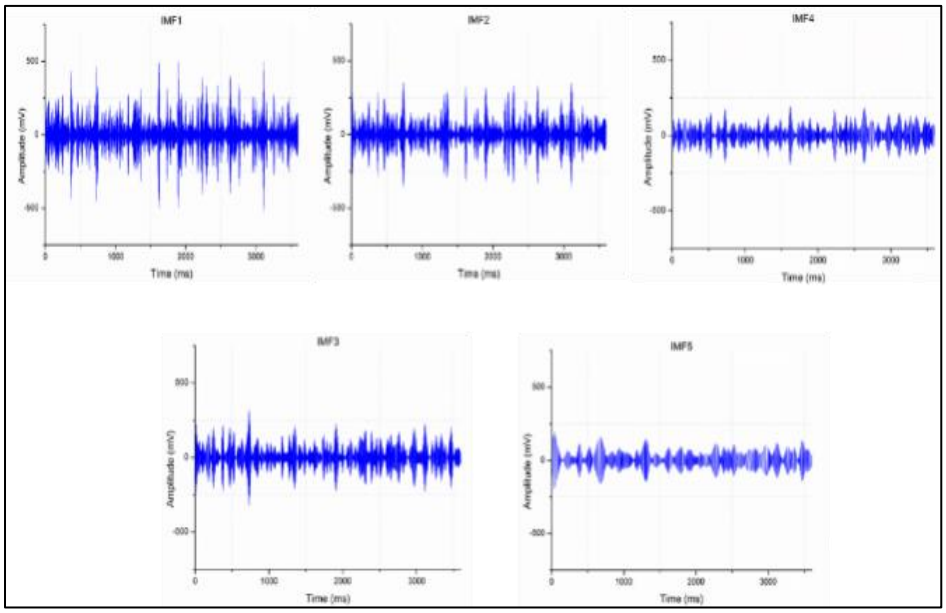

Figure 6 IMFs of test EEG signal

The seizure signal is having multi-frequency signal so the spectral features can reveal the basic discriminating features for the classification purpose. The values of extracted features of IMF1 to IMF3 is shown in the below table.

Table 1 Sample matrix of Extracted IMF spectra features

	IMF1			IMF2			IMF3		
	Spectral Centroid	Variation coefficient	Skewness	Spectral Centroid	Variation coefficient	Skewness	Spectral Centroid	Variation coefficient	Skewness
Test EEG1	-4.16	20.79	-17.13	38.02	-2.30	-64.73	1.60	10.25	9.41
Test EEG2	65.09	28.07	-13.94	-16.03	-28.94	-32.77	-27.45	-82.70	-98.71
Test EEG3	5.05	-34.10	-75.12	-72.63	-125.79	-166.03	-95.31	-58.68	-71.97
Test EEG4	-81.55	-140.93	-79.33	-72.83	-88.61	-121.56	-155.97	-91.81	-86.30
Test EEG5	-107.07	-72.40	-39.46	-51.78	-22.70	3.56	32.79	61.84	22.06

The dimension of the feature matric is huge and so the dimension reduction technique is mandate for this proposed methodology. As discussed in section C, the Whale optimization based feature selection technique is implemented in this work. The specifications of WOA is tabulated below and the feature selection algorithm is simulated for selecting best 6 features from the given large feature matric. As a result, WOA performed well towards the minimization of losses in classification process the following figure is showing the convergence curve for the WOA where the optimum solution ie. Optimum selection of best 6 features is obtained within 50 iteration. This shows how fast the feature selection happens.

Table 2 Parameters of WOA

Name of the WOA parameter	Selected value
Maximum number if iteration	50
Number of search agent	30
Boundaries of whale [min max]	[1 15]
Dimension	6
Evaluation function	Classifier Loss function

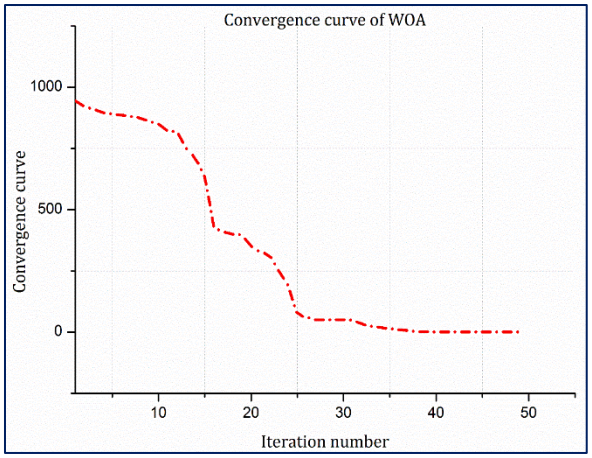

Figure 7 Convergence curve of WOA based feature selection algorithm Filtered seizure EEG signal

The optimum features are listed by the WOA algorithm where the selected most common discriminating features are listed out here.

Table 3 Optimum features selected by WOA

IMF1 – F1	IMF1 – F3	IMF2 – F3	IMF3 – F1	IMF3 – F2	IMF3 – F3
Spectral Centroid	Skewness	Skewness	Spectral Centroid	Variation coefficient	Skewness

The random forest classifier is trained with 275 training samples where 180 samples are normal EEGs and rest of them are seizure EEGs. RF classifier can give two results as hard class and soft class during the testing phase. The hard class is exact class of the signal and the soft class is the crispy value of the class. Hence the random forest classifier is providing an acute class of the given input signal.

The procedure for determining performance was achieved by comparing the classification the outcomes of the known class. The validation method yielded four values: true positive (TP), false positive (FP), true negative (TN), and false negative (FN) (FN). True positives is a term that refers to a group of people who believe in something. signals correctly recognised as epileptic seizure, A variety of things can cause a false positive signals True negatives are a variety of items that have been wrongly flagged signals as identified correctly normal False negative (FN) is a term that refers to a variety of things that aren't true signals erroneously marked as epileptic seizure.

Table 4 Performance parameters

	Epileptic Seizure	Normal
Found	Positive in every way (TP)	Falsely optimistic (FP)
There was nothing found.	Negative False (FN)	True Negative (TN)

For purposes of calculation, each parameter is calculated signal in the set of data As precision tests, sensitivity, specificity, and predictivity are used.

$$\text{Sensitivity} = (\text{TP} + \text{FN})/(\text{TP} + \text{FN})/(\text{TP} + \text{FN})/(\text{TP} + \text{FN})/(\text{TP} +$$

$$\text{Specificity} = \frac{\text{TN}}{(\text{FP} + \text{TN})}$$

$$\text{Predictivity} = \frac{\text{TP}}{(\text{TP} + \text{FP})}$$

The above parameters are measured for the whole dataset using k-fold validation technique and performance attributes are tabulated in Table 5.

Table 5 Parameter results

K fold	Sensitivity	Specificity	Predictivity
k=3	0.9796	0.5810	0.9796
k=5	0.9524	0.6667	0.9756
k=10	0.9062	0.6667	0.9667

The evaluated results are compared with the results obtained using support vector classifier. The samples are directly classified by SVM with feature selection process this will reduce the computational effort but the accuracy needs to be compromised in SVM based work. While comparing the performance of both the classification method, proposed Random forest based work is providing better results compared with the latter method. The graphical comparisons are shown in figure 8 to Figure 9.

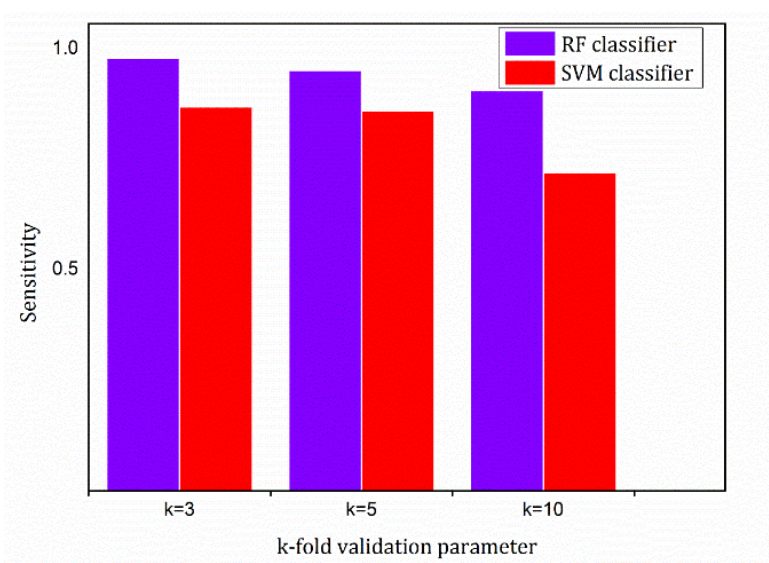

Figure 8 Sensitivity of proposed system

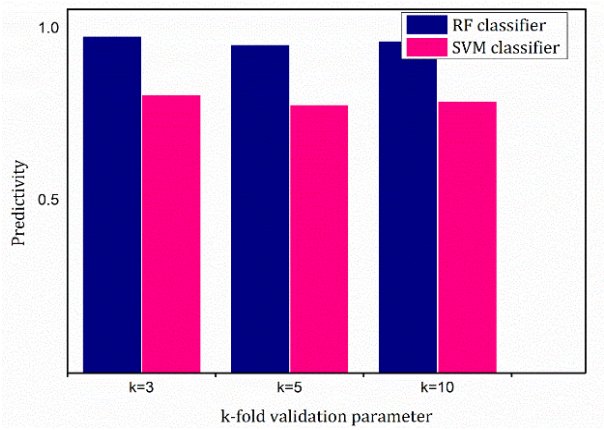

Figure 9 Predictivity of proposed system

Final thoughts and Future Work

The EEG is an electroencephalogram (EEG) signal seizure detection that is dependent is addressed in this paper. The EEG signals are downloaded from the Bonn university dataset. The recorded EEG signals are having artifacts due to interference effects which may induce computational error in diagnosis so the artifacts are eliminated by the filtering at preprocessing stage. The spectral features are extracted from the artifacts-free signals are extracted. This large number of feature metric will not help for producing good classification rate. The whale optimization based feature selection technique is implemented in this proposed work where the dominant features are selected by the optimization approach this also avoiding the memory overhead issues by reducing the dimension of the feature metric. The Random forest classifier is trained with a huge number of EEG signals during offline training phase and it is creating the model for testing the unknown test EEG signals. The classification process are validated by means of K-fold validation approach and it reveals the efficiency of the proposed work that how the system is responding to an unknown input. As result shows, the classification accuracy is achieved as 97.76% as maximum for this seizure detection and this showcases the stat-of-art of the proposed work.

In addition to the seizure detection, seizure classification will assist the physician more and this will also helpful for the patients where they can immediately go for the concern diagnosis. It is planned to implement seizure classification in future.

References

Chen, Z., Lu, G., Xie, Z., & Shang, W. (2020). A unified framework and method for EEG-based early epileptic seizure detection and epilepsy diagnosis. *IEEE Access*, *8*, 20080-20092.

Guo, W., Liu, T., Dai, F., & Xu, P. (2020). An improved whale optimization algorithm for feature selection. *Comput., Mater. Continua*, *62*(1), 337-354.

Hanosh, O., Ansari, R., Younis, K., & Cetin, A.E. (2019). Real-time epileptic seizure detection during sleep using passive infrared sensors. *IEEE Sensors Journal*, *19*(15), 6467-6476.

Park, Y.S., Cosgrove, G.R., Madsen, J.R., Eskandar, E.N., Hochberg, L.R., Cash, S.S., & Truccolo, W. (2019). Early detection of human epileptic seizures based on intracortical microelectrode array signals. *IEEE Transactions on Biomedical Engineering*, *67*(3), 817-831.

Song, J.L., Li, Q., Zhang, B., Westover, M.B., & Zhang, R. (2019). A new neural mass model driven method and its application in early epileptic seizure detection. *IEEE Transactions on Biomedical Engineering*, *67*(8), 2194-2205.

Wu, D., Wang, Z., Jiang, L., Dong, F., Wu, X., Wang, S., & Ding, Y. (2019). Automatic epileptic seizures joint detection algorithm based on improved multi-domain feature of cEEG and spike feature of aEEG. *IEEE Access*, *7*, 41551-41564.

Xu, K., Zheng, Y., Zhang, F., Jiang, Z., Qi, Y., Chen, H., & Zhu, J. (2019). An energy efficient adaboost cascade method for long-term seizure detection in portable neurostimulators. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *27*(11), 2274-2283.

Aldana, Y.R., Hunyadi, B., Reyes, E.J.M., Rodriguez, V.R., & Van Huffel, S. (2018). Nonconvulsive epileptic seizure detection in scalp EEG using multiway data analysis. *IEEE journal of biomedical and health informatics*, *23*(2), 660-671.

Kusmakar, S., Karmakar, C.K., Yan, B., O'Brien, T.J., Muthuganapathy, R., & Palaniswami, M. (2018). Automated detection of convulsive seizures using a wearable accelerometer device. *IEEE Transactions on Biomedical Engineering*, *66*(2), 421-432.

Lahmiri, S., & Shmuel, A. (2018). Accurate classification of seizure and seizure-free intervals of intracranial EEG signals from epileptic patients. *IEEE Transactions on Instrumentation and Measurement*, *68*(3), 791-796.

Paul, Y. (2018). Various epileptic seizure detection techniques using biomedical signals: a review. *Brain informatics*, *5*(2), 1-19

Yuan, Y., Xun, G., Jia, K., & Zhang, A. (2018). A multi-view deep learning framework for EEG seizure detection. *IEEE journal of biomedical and health informatics*, *23*(1), 83-94.

Faust, O., Acharya, U.R., Adeli, H., & Adeli, A. (2015). Wavelet-based EEG processing for computer-aided seizure detection and epilepsy diagnosis. *Seizure*, *26*, 56-64.

Oweis, R.J., & Abdulhay, E.W. (2011). Seizure classification EEG signals using Hilbert-Huang transform. *BioMedical Engineering OnLine*, *10*(1), 38.

Rahimi, F., Riahinia, N., Nourmohammadi, H., Sotudeh, H., & Tavakoli Zadeh-Ravari, M. (2019). How academia and society pay attention to climate changes: A bibliometric and altmetric analysis. *Webology*, *16*(2), 108-127.