Journal of Ambient Intelligence and Humanized Computing (2023) 14:15497–15512
https://doi.org/10.1007/s12652-019-01613-7

ORIGINAL RESEARCH

IoT and cloud computing based automatic epileptic seizure detection using HOS features based random forest classification

Kuldeep Singh[1] · Jyoteesh Malhotra[2]

Received: 2 August 2018 / Accepted: 3 December 2019 / Published online: 11 December 2019

Abstract

Epilepsy, a fatal neurological disorder, has been emerged as a worldwide problem and is one of the major risks to human lives. There exists an urgent need for an efficient technique for early detection of epileptic seizures at its initial stage in order to save the lives of thousands of epileptic patients annually. Now a days, internet of things in combination with machine learning techniques and cloud computing services has emerged as a powerful technology to resolve many problems in healthcare sector. This paper also presents an automatic epileptic seizure detection system and its layered architecture for early detection of epileptic seizures using existing communication technologies in collaboration with machine learning and cloud computing. This model transmits sensed EEG signals from patient's scalp to cloud through 4G cellular network or Wi-Fi network. At cloud, EEG signals are processed using Fast Walsh Hadamard transform and higher order spectra (HOS) based feature extraction for extracting higher order statistics and entropy-based features. The correlation-based feature selection algorithm has been employed for reducing the dimensionality of EEG datasets so as to tackle the problem of large volume of data and to reduce delays in service offered to the end user. Random Forest algorithm has been employed for classification of EEG signals into three different seizure stages, viz., normal, preictal and ictal. For performance analysis, other well-known machine learning algorithms like Bayes Net, Naïve Bayes, Multilayer Perceptron, Radial Basis function neural network and C4.5 Decision Tree are also considered. The simulation and testing results show that Random Forest classifier provides maximum values of classification accuracy of 99.40%, sensitivity of 99.40% and specificity of 99.66%, minimum mean square error of 0.0871 along with optimum training time of 20 ms, which makes this model more real time compatible, thereby making HOS features based Random Forest algorithm's cloud model an efficient technique for early and automatic detection of epileptic seizures in real time.

Keywords Epilepsy · Cloud computing · Random forest · Machine learning · Internet of things · Healthcare · Higher order spectra

1 Introduction

Epilepsy is a fatal neurological disorder, which tends to affect the human brain through the occurrence of spontaneous and unpredictable seizures (NINDS 2018). According to the reports of the World Health Organization (WHO), approximately 50 million people around the world are suffering from epilepsy, which make it one of the most common global neurological diseases (WHO 2018); and as per the statistics provided by Indian Epilepsy Centre, there exist approximately 10 million cases of epileptic seizures in India and 0.5–1 million new cases are being added annually (IEC 2018). Because of epilepsy disease, the normal pattern of neuronal activity in the human brain gets disturbed, which leads to the rise of vital clinical signs such as strange sensations, abnormal behaviour, emotions, muscle spasms and loss of consciousness etc. (NINDS 2018). The loss of consciousness could be life threatening for an epileptic patient, while driving or walking (Sareen et al. 2016). These uncontrolled seizures have severe impact on the quality of lives of epileptic patients and their care-takers irrespective of its low duration and frequency. But such patients have

✉ Kuldeep Singh
kuldeepsinghbrar87@gmail.com

1 Department of Electronics Technology, Guru Nanak Dev University, Amritsar, India

2 Department of ECE, Guru Nanak Dev University Regional Campus, Jalandhar, India

right to live normally during non-seizure intervals (Freestone et al. 2017). Taking into account these facts, there exist an extreme requirement of an automatic epileptic seizure detection system, which could be capable of alerting the patients, their family members and nearby hospitals before the actual occurrence of epileptic seizures. Thus, this system could help epileptic patients in case of emergency for saving their lives and for improving their quality of life.

Electroencephalogram (EEG) is a widely used technique to measure electrical disturbances in the human brain for diagnosis of epileptic seizures (Shoeb and Guttag 2010). During an epileptic seizure, the normal shape of EEG signals gets modified. Thus, on the basis of variation in EEG signal characteristics, the state of epileptic patients can be categorized into three stages, viz., normal, preictal and ictal (Gajic et al. 2014). According to a clinical study (Litt et al. 2001), large amount of electrical disturbances start occurring in the brain of epileptic patient before a seizure's actual onset, which is termed as preictal stage. For detection of seizures at this stage, such electrical disturbances in the patient's brain need to be recorded during transition from normal to ictal stage. Consequently, this process of early detection of epileptic seizures at preictal stage could save the lives of patients by enabling them to take precautionary measures in order to prevent injurious and life-threatening accidents.

Now a days, Internet of things (IoT) is playing a dynamic role in healthcare sector by providing significant solutions for many medical and healthcare applications like remote health care, fitness programs, child and elderly care, detection and prognosis of many chronological diseases such as epilepsy, schizophrenia and Alzheimer (AbdulGhaffar et al. 2019; Azimi et al. 2017; Darwish et al. 2017; Malasinghe et al. 2019; Mora et al. 2017; Moreira et al. 2019; Sathyanarayana et al. 2018). IoT technologies provide continuous and real time monitoring of patient's health using wearable devices (Islam et al. 2015). These technologies are also applied for acquisition and transmission of EEG signals of epileptic patients. Along with such technologies, machine learning algorithms provide promising solutions for effective detection of seizure stages from received EEG signals. A survey of epileptic seizure detection and prediction techniques (Alotaiby et al. 2015) presents various seizure detection methods based on wavelet or time domain features using machine learning (ML) classifiers such as support vector machine (SVM), artificial neural network (ANN) or Bayesian Network. This survey concludes that epilepsy detection systems with ML classifiers provide better performance than other techniques, which are not utilising the concept of machine learning. Another work (Orosco et al. 2013) also discusses epileptic seizure detection using ANN and SVM classifiers, which provides efficient classification results with sensitivity and specificity values from 88 to 100%. As per this work, an effective epileptic seizure detection system should show at least 80% specificity and sensitivity values for providing nominal alerts of epileptic seizures.

A cloud computing based automatic epileptic seizure detection technique has been presented by Sareen et al. (2016), which takes into account Fast Walsh Hadamard Transform (FWHT) and higher order statistical analysis (HOSA) based feature extraction and k-means algorithm based classification with an accuracy of 94.6%. A similar approach presented by Sareen et al. (2016) makes use of Gaussian Process classifier with FWHT and HOSA based feature extraction, which provides classification accuracy of 85.10% for early prediction of seizures. In a ML based approach discussed by Jaiswal and Banka (2018a), the authors have presented sub pattern based principal component analysis (SpPCA) and cross-sub pattern cross-correlation based PCA(SubXPCA) feature extraction techniques with SVM classification and have obtained classification of three stages of epileptic seizures viz., normal, preictal and ictal with accuracies of 92.76% and 96.66% for SpPCA and SubXPCA respectively. In another similar technique presented by Jaiswal and Banka (2017a), global modular PCA (GmodPCA) based feature extraction and SVM classifier with RBF kernels have been employed for epileptic seizure detection and have achieved 100% accuracy for binary classification between normal and ictal seizure stages. Xiaofeng Liu et al. (2017) have proposed a SVM model utilizing incremental entropy features for epileptic seizure detection technique with accuracy of 97.32%. Another SVM based solution for automatic diagnosis of epilepsy has been discussed in their paper by Wang et al. (2017), which employs multi-domain feature extraction and provides classification accuracy of 99.25%.

In an approach of detecting epileptic seizures using different local features, Jaiswal and Banka (2018b) have presented a local transformed features-based technique with ANN classification. The authors have discussed nine different combinations of seizure stages. Among these combinations, ANN classifier has achieved classification accuracies of 98% and 98.33% for local centroid pattern (LCP) and one dimensional local ternary pattern (1D-LTP) based feature extraction techniques respectively for a combination of normal, preictal and ictal seizure stages. They have also achieved 100% classification accuracy for binary classification of normal and epileptic stages using both feature extraction techniques. In the similar manner, another ANN approach with different local pattern transformation-based feature extraction techniques has been used (Jaiswal and Banka 2017b) and the authors have achieved mean classification accuracies of 98.22 ± 0.45% and 97.06 ± 0.62% for local neighbor descriptive pattern (LNDP) and 1D- local gradient pattern (1D-LGP) feature extraction techniques in case of seizure stage combination of normal, preictal and ictal. They have also obtained 100% accuracy on testing and evaluation datasets using both feature extraction techniques with total computation time (training + testing time) of 0.985 s. In order to explore the effectiveness of ML classifiers, an idea of epileptic seizure detection using HOSA-based feature extraction has been presented by Singh and Malhotra (2018) using C4.5 decision tree classifier. Upon testing of that model, a classification accuracy of 87.668%, sensitivity of 87.7% and specificity of 93.32% has been attained for a test set of 25000 EEG samples.

For detecting epileptic seizures in multichannel EEG data samples, Cui et al. (2018) have presented bag-of-waves based approach for learning EEG synchronization patterns using extreme learning machine (ELM) and have provided accurate prediction of seizures with sensitivity of 88.24%. A similar approach given in Hu et al. (2019) presents convolutional neural network (CNN) based feature extraction using mean amplitude spectrum of 19 subbands taken from 18 EEG channels and seizure classification using SVM. This technique provides classification results with accuracy of 86.25%.

In the similar way, various other classification techniques are also proposed by several researchers for automatic epilepsy seizure detection, which include kernel extreme machine learning (ELM) based classification (Liu et al. 2017), Tunable Q Wavelet Transform (TQWT) and Kraskov entropy based feature extraction technique with least square support vector machine (LS-SVM) (Patidar and Panigrahi 2017), Discrete Short Time Fourier Transform based feature extraction and Multilayer Perceptron (MLP) (Samiee et al. 2015), ELM using optimized sample entropy based feature extraction (Song et al. 2012) and fuzzy sugeno classifier using entropy based features (Acharya et al. 2012).

In this era of IoT technologies and cloud computing services, a large number of patients can make use of such cloud-based epileptic seizure detection models, which could result in generation of large volume of EEG data. Therefore, for real time deployment of these models, the large volume of raw EEG data samples urgently needs to be handled with least possible delay for serving end users. This delay could be minimized by reducing processing and classification times of data samples with optimum selection of classification algorithms and desired features characterizing each class of EEG data samples. Therefore, these time constraints need to be addressed on priority in order to make an accurate and real time compatible model of automatic epileptic seizure detection.

The present work focuses on an automatic epileptic seizure detection system and its layered architecture for early detection of epileptic seizures. This system processes sensed EEG data samples by its decomposition into frequency domain using Fast Walsh Hadamard Transform (FWHT), which emphasizes the epileptic spikes of EEG signals in spectral domain for fast and accurate prediction of seizures. In the next step, it employs higher order spectra (HOS) based feature extraction, which is highly suitable for analysing non-linear characteristics of EEG signals because of its ability to preserve original information of a signal from deviations caused by non-linearities. In order to consider amplitude and phase behaviour of EEG signals, bispectrum and bicoherence parameters of HOS analysis are taken into account, which are further processed to extract their amplitude, moments and entropy-based features. After extracting various features, correlation-based feature selection procedure has been carried out to select appropriate features in order to reduce dimensionality of given dataset and to achieve accurate classification by discarding redundant features. Finally, the classification among three different seizure stages, viz. normal, preictal and ictal, has been performed using Random Forest, which is an accurate and fast machine learning algorithm. Because of its capability to take optimally lesser time for training and prediction, robustness against non-linear behaviour of data and ability to efficiently run on large databases, Random Forest classifier is preferred in this research work for ensuring its effectiveness to handle large volume of non-linear EEG data in real time scenarios. For effective analysis of results, the performance of given classifier has been tested and compared with other well-known machine learning (ML) algorithms, which are extensively used in biomedical applications. These classifiers include Bayes Net, Naïve Bayes, Multilayer Perceptron, Radial Basis Function Neural Networks, and C4.5 Decision tree algorithms. The performance of these algorithms has been evaluated using performance metrics, which are widely acceptable for analysis in biomedical fields, such as accuracy, sensitivity, specificity, Youden's index, positive likelihood, negative likelihood, discriminant power, training time, testing time and various error measures. After carrying-out several simulations, the proposed model of Random Forest classifier exhibits excellent results for classification of seizure stages with high accuracy and optimally low training and testing times for given EEG data samples. Finally, the analysis of simulation results of this framework reveals that Random Forest algorithm with HOS feature extraction is an effective and suitable technique for accurate detection of epileptic seizures in real time.

The present paper is framed into different sections. Section 1 briefly introduces the problem of epileptic seizures and various techniques proposed for its detection in recent literature. The overview of proposed automatic epileptic seizure detection system and its layered architecture has been presented in Sect. 2, whereas, Sect. 3 explains methodology adopted and implementation of the proposed model. The simulation results provided by given model have been discussed in Sect. 4 to ensure the effectiveness of this model for the detection of epileptic seizures. Finally, conclusions have been discussed in Sect. 5.

2 Proposed automatic epileptic seizure detection system

The basic building blocks of the proposed model of automatic epileptic seizure detection system are shown in Fig. 1. This proposed model consists of bluetooth enabled EEG headset, mobile phone and its communication section, which is termed as device layer and the cloud storage and computing services section, which is termed as cloud layer.

In device layer, the first task of acquisition of EEG data is performed using an acquisition module, which is a wearable EEG headset. The function of EEG headset is to collect EEG signals through one or multiple electrodes, which detect voltage of current waves flowing through brain neurons of different segments of the human brain. The raw EEG signals acquired from EEG headset are transmitted to patient's mobile phone through bluetooth. These signals are retrieved by an android based application in digital form, and are further transmitted to cloud-based server on cloud layer through Wi-Fi or 4G cellular network.

2.1 Layered architecture of the proposed system

The proposed layered approach for automatic epileptic seizure detection consists of six layers, namely: Data Acquisition Layer (DAL), Data Transmission Layer (DTL), Information Mining Layer (IML), Sample Classification Layer (SCL), Decision Making Layer (DML) and Cloud Storage Layer (CSL). The layered architecture of this proposed system is illustrated in Fig. 2. Each of these layers provides efficient services by performing its requisite functions, which are discussed in the following subsections.

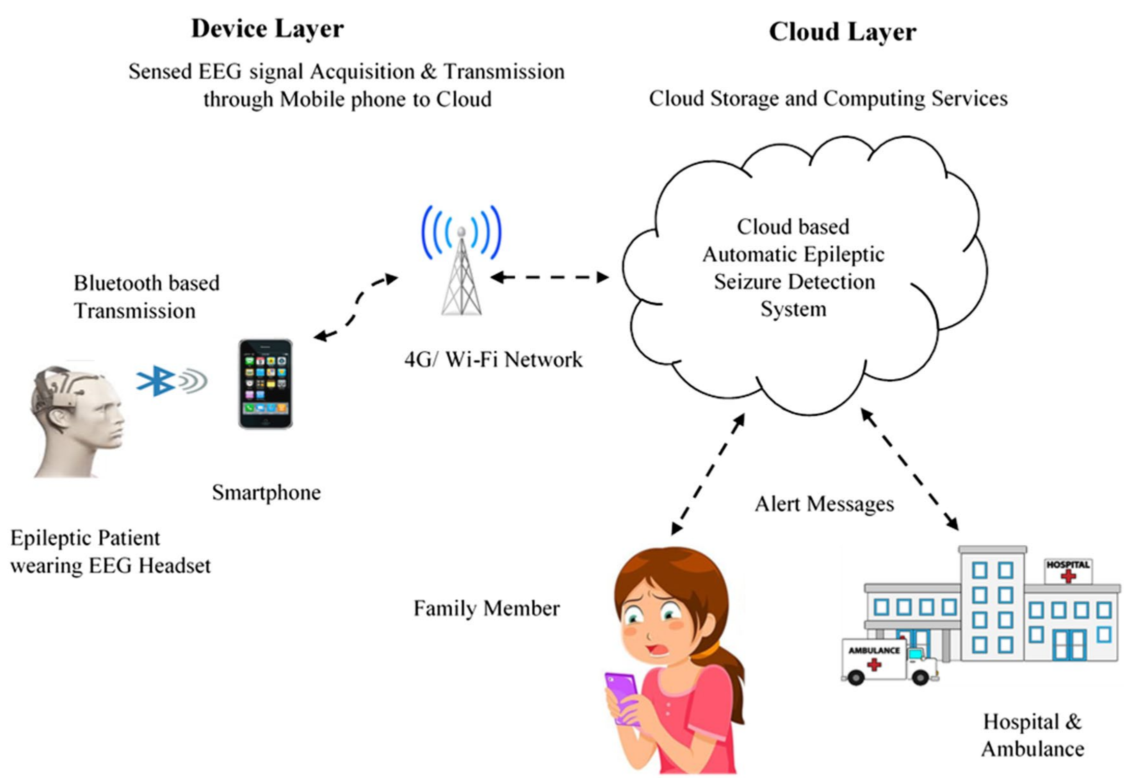

Fig. 1 Overview of proposed automatic epileptic seizure detection system

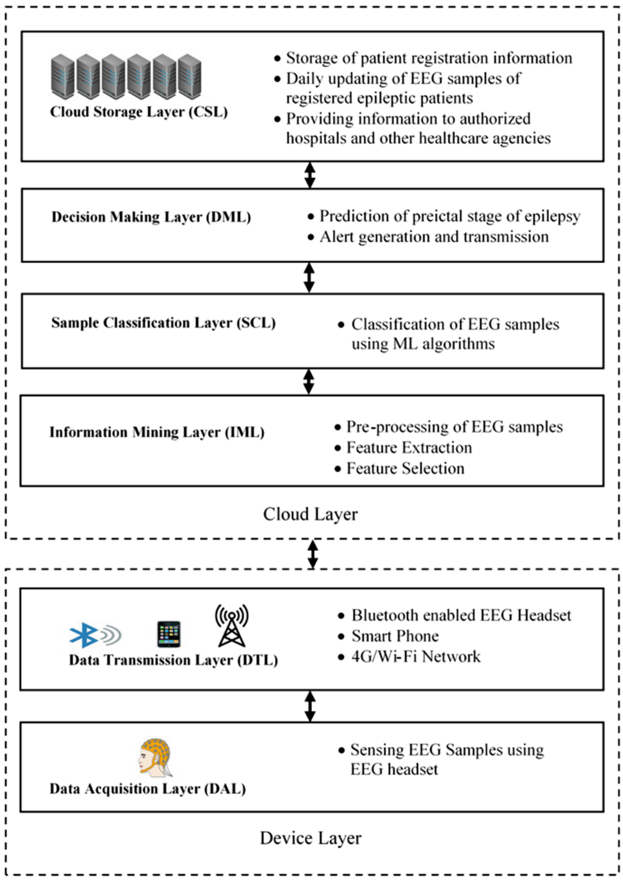

Fig. 2 Layered architecture of proposed automatic epileptic seizure detection system

2.1.1 Data acquisition layer (DAL)

Data acquisition layer (DAL) performs the function of retrieving EEG data samples from epileptic patient's scalp using wearable EEG headset. This EEG headset senses patient data every second to make this model appropriate for real time usage, so that this model can play the role of life saviour for epileptic patients effectively. There are large variety of EEG headsets available in the market such as Emotiv EPOC headset (Emotiv 2018), NeuroSky Mindwave EEG headset (NeuroSky 2018) and other devices (Neurotech 2016). For this research work, Emotiv EPOC headset has been proposed because it contains 14 electrodes, which are located along the scalp for sensing EEG signals from different parts of the brain. This layer also converts sensed data into textual format for its easy transmission and utilization on cloud. It is a front-end layer and acts a sub-layer of the device layer in this model.

2.1.2 Data transmission layer (DTL)

This layer performs the function of transmission of sensed EEG data from EEG headset to cloud server and also acts as a sub-layer of the device layer for this model. The EEG data samples sensed by EEG headset are transmitted to a smart phone using low energy bluetooth technology, which are collected by an android-based application within the smart phone in digital format. Smart phones further transmit these data samples to cloud server through internet connectivity provided by 4G network or Wi-Fi services.

2.1.3 Information mining layer (IML)

IML is a sub-layer of cloud layer. It performs the tasks of pre-processing of raw EEG samples, its feature extraction and feature selection for dimensionality reduction. The preprocessing of EEG samples is performed by converting them from time domain to frequency domain through fast Walsh Hadamard transform (FWHT). Because of non-linear and dynamically changing nature of EEG data samples (Subha et al. 2010; Upadhyay et al. 2015), higher order spectra (HOS) (Nikias and Mendel 1993) based non-linear feature extraction technique seems to be an efficient way of feature extraction. Therefore, these EEG samples are fed to higher order spectra (HOS) analysis for obtaining its bispectrum and bicoherence values. Various features characterizing each epilepsy stage are extracted from these HOS based values. These features are related to magnitude, entropy values and moments of bispectrum and bicoherence. In order to handle the problem of big data of EEG signals due to large number of patients utilizing this service, the dimensionality reduction is performed using Correlation based feature selection (CFS) algorithm (Singh and Agrawal 2011), which discards redundant features and selects only suitable features.

2.1.4 Sample classification layer (SCL)

This layer performs the function of classification of EEG data samples retrieved from epileptic patient's scalp into different stages of epileptic seizures such as normal, preictal and ictal stage. The EEG signals characterized by various selected features obtained from IML, are classified using Random Forest classification algorithm (Cutler et al. 2012). This layer has great significance in correctly predicting preictal stage with maximum accuracy and least classification time so as to offer real time services effectively with least delay.

2.1.5 Decision making layer (DML)

This layer is responsible for determining whether a patient is in safe state or not. If SCL gives preictal stage as its outcome, then DML make decision about alerting the patient, his/her family members and nearby hospital or ambulance services through an alert message in case of occurrence of seizures and preventive measures may be taken accordingly to save the life of patient. The patient's current geographical location is also being tracked through GPS based location tracking system and shared along with alert message.

2.1.6 Cloud storage layer (CSL)

This layer plays a significant role in reception and aggregation of sensed EEG data samples from epileptic patients on daily basis. The raw EEG signals acquired from epileptic patient are stored on this layer along with patient's personal information in order to uniquely identify received EEG signals for a particular patient. The patient's information includes patient's social security number such as Unique ID (UID), name, age, gender, home address, family member's name, family member's mobile number etc. This layer provides certain information to DML such as nearby hospital location and contact numbers as well as contact numbers of registered family members, for handling emergency situations. Summarized EEG data samples are also provided to many hospitals, healthcare agencies and research and development organizations for developing new medicine or vaccine and for carrying out further research in the field of early detection of epileptic seizures.

3 Methodology and implementation

This section discusses the procedure adopted for analysis and classification of EEG signals. For this research work, MATLAB 2014b has been used for feature extraction on personal computer having configuration of Intel i3 processor with 3 GB RAM with Windows 10 operating system. Also, Weka 3.8.1 software tool (Holmes et al. 1994) has been used for feature selection, training and testing of performance of different classifiers on local server having configuration of Intel (R) Xeon (R) CPU E5-2620v3 2.40 GHz processor with 16 GB RAM and Windows Server 2012R2 Standard operating system. The methodology adopted for carrying out this research has been clearly demonstrated by pseudo algorithm 1.

Algorithm 1 Analysis of EEG data samples for epileptic seizure detection

Input: Database of Epileptic Patients containing EEG samples.
Output: Classification of different epileptic seizure stages, namely: normal, preictal and ictal.
%...........**Uploading and Storage of EEG data**..........%
Read patient's UID generated by the system;
Read personal data and EEG data from the patient through his/her mobile phone;
if UID already exist
then
Update the records with new EEG data;
else
Create two new records to save personal and EEG data of the patient for his/her UID;
Add patient's personal details in personal data record securely;
Add patient's newly generated EEG data for his/her UID securely using generated secret keys;
end for
Share secret keys to patient, hospitals and other healthcare agencies;
Send (UID, data) to Cloud server for storage;
endif

%**Pre-processing, Feature extraction & selection**%
Let n be the number of EEG samples and d be the number of data values in each EEG sample.
Let $wh[.]$ and $np[.]$ be the one dimensional matrices, which are used to store FWHT and normalized FWHT coefficients for each EEG sample of given registered UID.
for all $i \leq n$ ***do***
Read EEG data sample from cloud storage;
Compute FWHT coefficients for given EEG sample and save them in $wh[.]$ matrix;
for all $i \leq d$ ***do***
Find mean and standard deviation (std) of each FWHT coefficient;
Normalize each FWHT coefficient wh[.] using equation

$$np[i] = (wh[i] - mean(wh[i]))/std(wh[i])$$

end for
Calculate bispectrum and bicoherence values using HOSA toolbox of *Matlab*;
Calculate Bispectrum mean magnitude, normalized Shannon entropy, log energy entropy, norm entropy, sum of logarithmic amplitudes of bispectrum, sum of logarithmic amplitudes of diagonal elements of bispectrum, first order and second order spectral moments of amplitudes of diagonal elements of bispectrum features using bispectrum and bicoherence values.
end for
Load feature dataset in *Weka* tool;
Apply Correlation based Feature selection algorithm of *Weka* to select appropriate features;
end for

%................**Classification of EEG samples**.................%
Execute Random Forest Classification algorithm of *Weka*;
endif
Label three output classes of EEG samples as normal, preictal and ictal;
end
exit

The methodology used for automatic epileptic seizure detection is also described by the flowchart shown in Fig. 3. This flowchart clearly demonstrates different steps performed for detection of epileptic seizures. These steps include EEG signal clinical dataset, preprocessing of raw EEG signals, feature extraction using higher order spectra (HOS) based technique, feature selection using CFS algorithm, training of developed ML classifier using training dataset in offline phase, testing of developed ML classifier using up-sampled testing dataset in online phase and performance analysis of different ML classifiers for classification of different stages of epileptic patient. All these steps are individually elaborated in the following sub-sections.

3.1 EEG signals clinical dataset

The publicly available Andrzejak et al. (2001) dataset from University of Bonn, Germany has been taken into consideration for this research work. This dataset consists of five subsets of EEG data segments, denoted as A, B, C, D and E. Each subset is comprised of 100 single channel EEG segments of 23.6 s duration. These single channel EEG segments are measured using 14 surface electrodes placed on the skull of patients using standard 10–20 placement method (Homan et al. 1987). These EEG signal records are captured by a 128-channel amplifier and are converted into digital form at a sampling rate of 173.61 Hz and 12-bit analog to digital resolution. Single channel EEG signals are categorized into three stages of epileptic patients such as normal, preictal, and ictal. From these five subsets, first two subsets A and B are taken from five healthy persons and are termed as Normal. Other three subsets C, D, and E are recorded from five epileptic patients. Among these three subsets, the subsets C and D contain EEG signals recorded from epileptic hemisphere and opposite hemisphere of the human brain respectively during seizure-free intervals and are termed as Preictal. The remaining subset E contains the EEG signals associated with seizure activity only and are termed as Ictal. Therefore, the given dataset is grouped into A–B (Normal), C–D (Preictal) and E (Ictal) groups and this dataset of 500 samples consists of 200 samples of normal stage, 200 samples of preictal stage and 100 samples of ictal stage. Figure 4 clearly represents EEG signal plots for normal stage in case of healthy person and preictal and ictal stages in case of epileptic persons.

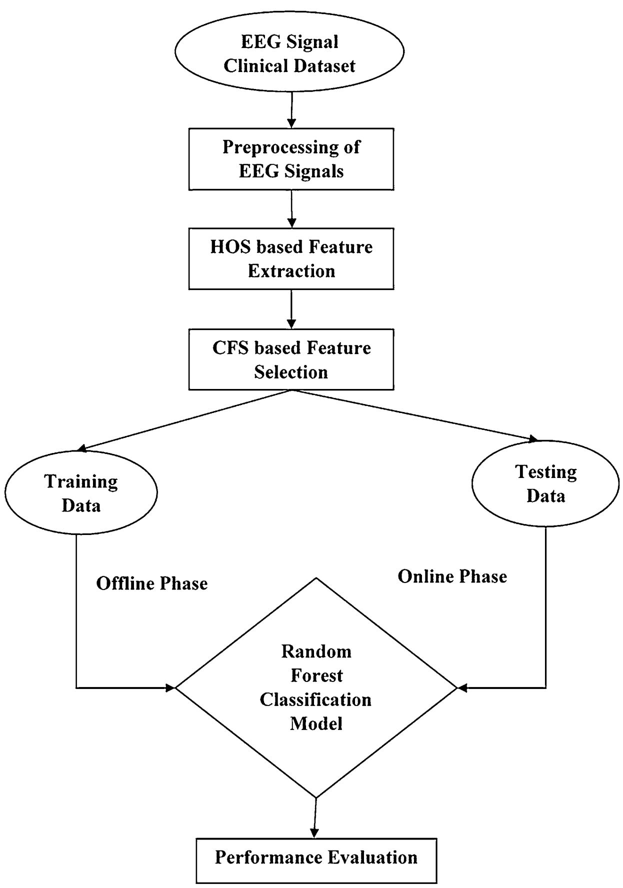

Fig. 3 Flowchart of methodology for automatic epileptic seizure detection system

3.2 Preprocessing of EEG signals

In preprocessing step, received raw EEG signals are decomposed into frequency domain samples by converting its time domain samples using Fast Walsh Hadamard Transform (FWHT). FWHT emphasizes the epileptic spikes of EEG signals in frequency or spectral domain in order to make

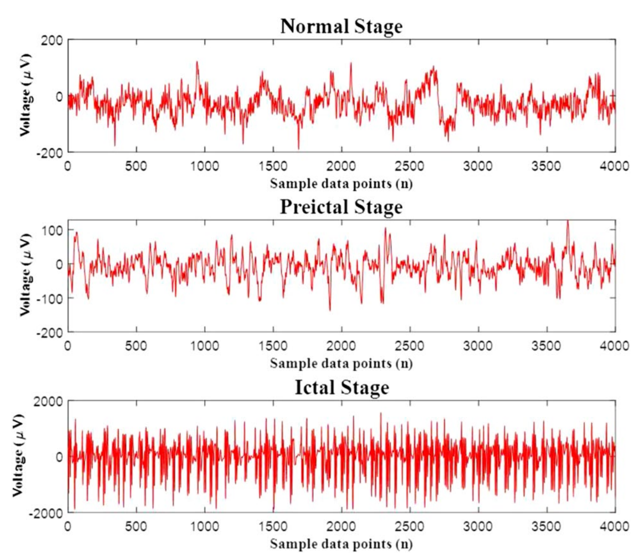

Fig. 4 EEG signal plots for normal stage, preictal stage and ictal stage

fast and more accurate classification or clustering of such signals.

The Fast Walsh Hadamard Transform (FWHT) of a signal $x(t)$ having length N can be defined by equation 1:

$$y_n = \frac{1}{N} \sum_{i=0}^{N-1} x_i \, WAL(n, i) \tag{1}$$

where $i = 0, 1, 2, 3, \ldots, N-1$ and $WAL(n,\ i)$ are Walsh functions.

FWHT produces unique sequency values, which are assigned to all Walsh functions individually. These values are used for estimation of frequencies values of the original signal. FWHT is able to perform highly accurate detection of signals containing high discontinuities with less processing time and having very few coefficients.

In present work, the obtained coefficients of FWHT are normalized using their mean and standard deviation values, so that any kind of error occurring due to inadequate number of features being extracted can be removed. Then, normalized FWHT values are applied to higher order spectra (HOS) based analysis procedure for extracting its bispectrum and bicoherence values.

3.3 HOS based feature extraction

In this section, the preprocessed and decomposed EEG signals using FWHT are fed to higher order spectra (HOS) based procedure for analysing the characteristics of epileptic seizures in terms of its bispectrum and bicoherence values.

Higher Order Spectra (HOS) based procedure or Higher Order Spectral Analysis (HOSA) (Nikias and Mendel 1993) is the spectral representation of higher order statistics of a signal. This procedure is a dominant tool for studying and analysing the non-linear characteristics of given EEG signals because of its ability to preserve the original information of the signal due to deviations from degrees of non-linearities and Gaussianity in the time series (Chua et al. 2010). The higher order statistics are represented in terms of moments, cumulants and its spectra. The third order and fourth order statistics and its Fourier transform are known as bispectrum and trispectrum respectively (Seijas et al. 2013).

The bispectrum $B(f_1,f_2)$ of a given signal is the Fourier transform of the third order correlation of the signal and it is a function of two frequencies. It is given by Eq. 2:

$$B(f_1,f_2) = E[X(f_1)X(f_2)X^*(f_1+f_2) \tag{2}$$

where $B(f_1,f_2)$ is the bispectrum with bifrequency (f_1,f_2), $X(f)$ is the discrete time Fourier transform of the given signal, $*$ represents complex conjugate and $E[.]$ gives the value of statistical expectation operation over an ensemble of possible realizations of the given signal.

Bicoherence $B_{CO}(f_1,f_2)$ Chua et al. (2010) is another important parameter in HOS, which is obtained from squared magnitude of normalized bispectrum. It is given by following equation 3:

$$B_{co}(f_1,f_2) = \frac{B(f_1,f_2)}{\sqrt{E[P(f_1)P(f_2)]E[p(f_1+f_2)]}} \tag{3}$$

where $P(f)$ is the power spectrum of the given signal. Bispectrum display symmetry has been estimated in the principal domain region (Ω) (Chua et al. 2009). Both bispectrum and bicoherence are obtained by applying direct FFT method in HOSA toolbox of MATLAB.

The bispectrum analysis is an essential technique for obtaining inter-frequency phase coherence, which characterizes the EEG signals for different stages of epilepsy. Bispectrum value consists of amplitude as well as degree of phase coupling of EEG signals, while Bicoherence values consist of degree of phase coupling only. For effectively characterizing the different stages of EEG signals, both amplitude and phase values are equally significant.

The present work makes use of bispectrum and bicoherence parameters obtained from normalized FWHT values of EEG signals with N coefficients. Eight features are extracted from these parameters using an algorithm implemented in MATLAB. Among these eight features, the five features are extracted from bispectrum, such as bispectrum mean magnitude (F_1) (Yuvaraj et al. 2018), sum of logarithmic amplitudes of bispectrum (H_1) (Yuvaraj et al. 2018), sum of logarithmic amplitudes of diagonal elements of bispectrum (H_2) (Yuvaraj et al. 2018), first order spectral moments of amplitudes of diagonal elements of bispectrum (H_3) (Yuvaraj et al. 2018) and second order spectrum moment of amplitudes of diagonal elements of bispectrum (H_4) (Yuvaraj et al. 2018). The remaining three entropy-based features are extracted from bicoherence values. These entropy features are Normalized Shannon Entropy (F_2) (Sareen et al. 2016), Log Energy Entropy (F_3) (Sareen et al. 2016) and Norm Entropy (F_4) (Sareen et al. 2016). The mathematical expressions for these features are given by following equations.

- Bispectrum Mean Magnitude (F_1)

$$F_1 = \frac{1}{N} \sum_{\Omega} |B(f_1,f_2)| \tag{4}$$

- Normalized Shannon Entropy (F_2)

$$F_2 = -\sum_i q_i \, log(q_i) \tag{5}$$

where

$$q_i = \frac{B_{co}(f_1,f_2)}{\sum_{\Omega} |B_{co}(f_1,f_2)|}$$

- Log Energy Entropy (F_3)

$$F_3 = -\sum_i log(s_i) \tag{6}$$

where

$$s_i = \frac{|B_{co}(f_1,f_2)|^2}{\sum_\Omega |B_{co}(f_1,f_2)|^2}$$

- Norm Entropy (F_4)

$$F_4 = \sum_i log(p_i) \tag{7}$$

where

$$p_i = \frac{|B_{co}(f_1,f_2)|^\rho}{\sum_\Omega |B_{co}(f_1,f_2)|^\rho}$$

- Sum of Logarithmic Amplitudes of Bispectrum (H_1)

$$H_1 = \sum_\Omega log(|B(f_1,f_2)|) \tag{8}$$

- Sum of Logarithmic Amplitudes of Diagonal Elements of Bispectrum (H_2)

$$H_2 = \sum_\Omega log(|B(f_D,f_D)|) \tag{9}$$

- First Order Spectral Moment of amplitudes of Diagonal Elements of Bispectrum (H_3)

$$H_3 = \sum_{m=1}^{N} m\, log(|B(f_D,f_D)|) \tag{10}$$

- Second Order Spectrum Moment of Amplitudes of Diagonal Elements of Bispectrum (H_4)

$$H_4 = \sum_{m=1}^{N} (m - S3)^2\, log(|B(f_D,f_D)|) \tag{11}$$

3.4 CFS algorithm based feature selection

After feature extraction, feature selection is another important task for dimensionality reduction to handle big data problem and for proper selection of desirable features in order to avoid redundant features, so that accurate classification can be performed for detection of different stages of epileptic patients. Therefore, correlation based feature selection (CFS) algorithm (Singh and Agrawal 2011) of Weka tool has been employed for this task. This algorithm estimates the worth of different features by considering the predictive ability of each individual feature and degree of redundancy among them. The features having high degree of correlation with the output class label and having low inter-correlation among others are preferred and other features are discarded. After implementation of this algorithm on the dataset of extracted features, the number of features is reduced from eight to six features, which are F_1, F_2, F_3, F_4, H_1 and H_2.

3.5 Random forest algorithm based classification

In this work of detecting epileptic seizures, Random Forest classification algorithm (Cutler et al. 2012; Rokach 2010) has been employed for classification of the various stages of epileptic patients such as normal, preictal and ictal. This algorithm is an ensemble machine learning technique, widely used for classification tasks. It is a combination of many individual and unpruned decision trees as a "forest" and aggregates the results of different trees for classification. The use of multiple decision trees in this algorithm make it more accurate classifier with better learning performance.

Random Forest technique is able to handle large number of input attributes and is very fast technique with lesser training and prediction times, thereby, making it suitable for handling the problem of big data (Manogaran and Lopez 2017). Moreover, this algorithm is also stable, robust against overfitting and non-linear behaviour of EEG data, able to effectively handle categorical features and a preferred approach for multilabel classification (Jia et al. 2013; Subasi et al. 2019; Zhang et al. 2017). On the other hand, the other popular traditional ML algorithms like SVM and k-nearest neighbor (kNN) encounter many problems such as their inability of handling large volume of non-linear multilabel EEG data due to slow speed and taking lot of computation time for training and prediction, sensitivity to irrelevant features and high memory usage irrespective of their high accuracy (Bentlemsan et al. 2014; Senagi et al. 2017; Zhang et al. 2017). Due to all these factors, Random Forest algorithm is given preference over other ML classifiers in many biomedical applications such as epileptic seizures, sleep disorders, migraine and Alzheimer etc. (Alickovic et al. 2019; Fraiwan et al. 2012; Klok et al. 2018; McDonald et al. 2014; Subasi et al. 2019; Zhang et al. 2017).

For this work, Random Forest algorithm has been trained for 500 samples epileptic patients in Weka 3.8.1 and the number of iterations for effective model building are kept to 10 only in order to reduce training time.

3.6 Testing of the proposed model

The proposed model of epileptic seizure detection system has been tested using Weka software on local server, whose configuration has been discussed in Sect. 3. For effective testing of this model in online phase, the dataset of 500 samples obtained from five persons is not sufficient. In order to effectively realize the results of epileptic seizure detection for big data problem, the dataset obtained after feature extraction and selection process has been up-sampled to big data of 50,000 data samples using resampling technique in Weka tool. The testing experiment has been conducted with first iteration of 5000 samples, then increasing sample size with repeated addition of extra 5000 samples. This experiment continuous until total number of samples reaches to 50,000.

4 Results and discussion

This section discusses results of the proposed model for classification of various epileptic seizure stages for a testing dataset of 50,000 epileptic samples in terms of various performance metrics. In order to efficiently demonstrate the performance of the proposed model and for proper clarification of results, performance of Random Forest classifier given in the proposed model has been compared with the performance of other well-known machine learning classifiers, which are widely used in biomedical signal processing. These classifiers include Bayes Net (Cheng and Greiner 2001), Naïve Bayes (Pop 2006), Multilayer Perceptron (MLP) (Singh and Agrawal 2011), Radial Basis Function Neural Network (RBF) (Singh and Agrawal 2011), and C4.5 decision tree (DT) algorithm (Korting 2006).

The commonly used performance metrics for performance analysis of machine learning classifiers being used in biomedical field are accuracy (Baratloo et al. 2015), sensitivity (Baratloo et al. 2015), specificity (Baratloo et al. 2015) and training time and testing time. But because of random nature of EEG data, other performance metrics, which are typically used for performance analysis and efficient discrimination of classifiers in medical diagnosis, such as Youden's index (Sokolova et al. 2006), likelihoods (Sokolova et al. 2006) and discriminant power (Sokolova et al. 2006), are also taken into consideration. These performance metrics are derived from confusion metrics (Ting 2017) of various classes of epileptic seizure stages, which consist of values of true positive (TP), true negative (TN), false positive (FP), and false negative (FN). All these performance metrics are discussed in the following subsections.

- Classification Accuracy (*ACC*): It is the ability of a binary classifier to differentiate between positive and negative classes such as different stages epileptic patient and healthy person. The present work takes into account percentage accuracy for analysing the performance of given ML classifiers.

$$ACC = \frac{TP + TN}{TP + FP + FN + TN} \tag{12}$$

- Sensitivity (*SENS*): It is the ability of a binary classifier to identify positive classes correctly such as any epileptic stage in case of epileptic patient. In this paper, percentage sensitivity has been used.

$$SENS = \frac{TP}{TP + FN} \tag{13}$$

- Specificity (*SPEC*): It is the ability of a binary classifier to identify negative classes correctly such as normal stage in case of healthy person. This term is also measured in percentage.

$$SPEC = \frac{TN}{TN + FP} \tag{14}$$

- Youden's Index (γ): This parameter evaluates classifier's ability to avoid failure. This parameter equally weights classifier's performance on positive and negative classes. A higher value of Youden's index indicates better ability to avoid failure.

$$\gamma = SENS - (1 - SPEC) \tag{15}$$

- Likelihoods (ρ_+, ρ_-): If a performance analysis accommodates both sensitivity and specificity, but treats them separately. Then, positive and negative likelihood values are used to evaluate a classifier's performance to precise degree.

$$\rho_+ = \frac{SENS}{1 - SPEC} \tag{16}$$

$$\rho_- = \frac{1 - SENS}{SPEC} \tag{17}$$

A higher value of positive likelihood (ρ_+) and lower value of negative likelihood (ρ_-) are preferable for better performance of a classifier on positive and negative classes respectively.

- Discriminant Power (*DP*): This parameter evaluates the ability of a classifier to distinguish between positive and negative classes.

$$DP = \frac{\sqrt{3}}{\pi}(logX + logY) \tag{18}$$

where

$$X = \frac{SENS}{1 - SENS}$$
$$Y = \frac{SPEC}{1 - SPEC}$$

A classifier is poor discriminant if DP < 1, limited if DP < 2, fair if DP < 3 and good for other cases.

- Training Time (T_{train}): It is the time taken by the system to build classification model. It is measured in milliseconds (ms). In this research work, average training time has been taken into consideration for number of patients from 5000 to 50000.

- Testing Time (T_{test}): It is the time taken by a given machine learning classification model to predict or classify given number of data samples. It is also measured in milliseconds (ms)

Table 1 shows values of overall performance of various ML classifiers in terms of performance metrics such as accuracy, average sensitivity, average specificity Youden's Index, positive and negative likelihoods, discriminant power, training time and testing time in order to detect different epileptic seizure classes for a dataset of 50,000 EEG samples of epileptic patients.

This table evidently exhibit that among all given ML classifiers, Random Forest classifier provides maximum classification accuracy of 99.40% along with maximum sensitivity of 99.40%, maximum specificity of 99.66%, Youden's index's maximum value of 0.9906, maximum positive likelihood of 292.35 and maximum discriminant power values of 5.9467, while having minimum negative likelihood value of 0.006 for 50000 data samples of epileptic patients, which makes it an excellent classifier for classification of different epileptic seizure stages among other given classifiers. The training time for Random Forest Classifier is 20 ms, which is slightly greater than that of Naïve Bayes, Bayes Net and C4.5 DT classifiers. The testing time for Random Forest classifier is 480 ms which is slightly great than that of C4.5 DT and MLP classifiers. But these values of training and testing times for Random Forest classifier are optimally small, thereby making it more real time compatible. Thus, the total computation time including training and testing times for Random Forest classifier is 0.5 s (20 ms + 480 ms) only, which is smaller than that of all other classifiers, except for C4.5 decision tree, which is having total computational time of 375 ms (14 ms + 361 ms). These results also demonstrate that MLP classifier having maximum values of training time is not an appropriate choice for online and real time classification of epileptic seizures. Apparently, Naïve Bayes, Bayes Net and RBF Neural Network classifiers are also not suitable for this application because of their very poor classification accuracy of 57.486%, 64.446% and 64.814% respectively. C4.5 DT classifier performs slightly better than Bayes Net, Naïve Bayes, RBF Neural Network and MLP classifiers for epileptic seizure detection with classification accuracy of 88.12%.

The efficient performance of Random Forest classifier along with other given ML techniques for epileptic seizure detection in terms of classification accuracy with respect to number of epileptic patient samples is also observable from graph shown in Fig. 5. This figure demonstrates the consistency of given ML classifier over increased number of EEG samples. Figure 6 provides the increasing values of testing time with respect to number of samples of epileptic patients. This figure clearly reveals that testing time for Random Forest classifier is increasing a very low rate with respect to number of patients as compared to that of Naïve Bayes and RBF Neural Network classifiers and is having slightly higher value of testing time than that of Bayes Net, MLP and C4.5

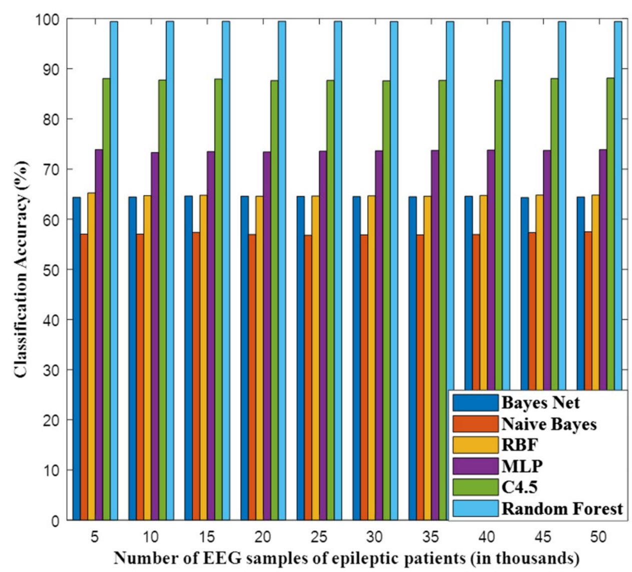

Fig. 5 Performance comparison of given ML classifiers in terms of classification accuracy w.r.t. number of EEG samples of epileptic patients

Table 1 Comparison of given ML classifiers for epileptic seizure detection of 50000 EEG samples of epileptic patients

S. no.	ML classifiers	*ACC* (%)	*SENS* (%)	*SPEC* (%)	γ	*Likelihoods*		*DP*	T_{train} (ms)	T_{test} (ms)
						ρ_+	ρ_-			
1.	Bayes Net	64.446	64.4	76.187	0.4059	2.7044	0.4673	0.9676	10	498
2.	Naïve Bayes	57.486	57.5	73.496	0.31	2.1695	0.5783	0.7287	1	813.3
3.	RBF NN	64.814	64.8	77.467	0.4227	2.8758	0.4544	1.0169	20	710
4.	MLP	73.848	73.8	84.03	0.5783	4.6212	0.3118	1.4858	690	363.3
5.	C4.5 DT	88.12	88.1	93.706	0.8181	13.997	0.1270	2.5916	14	361
6.	Random Forest	99.40	99.40	99.66	0.9906	292.35	0.006	5.9467	20	480

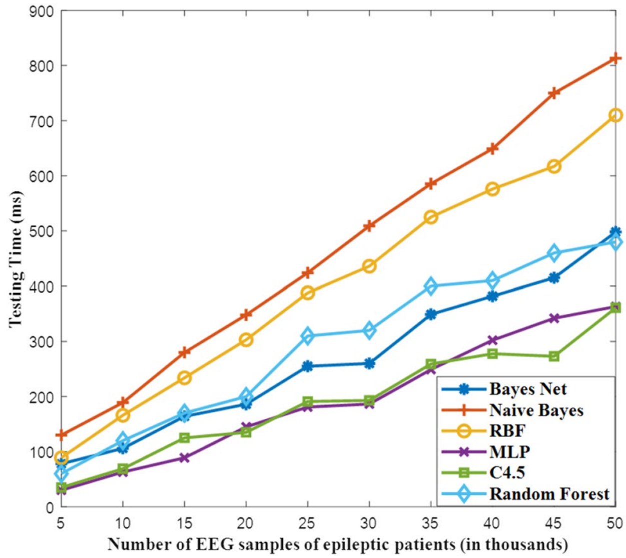

Fig. 6 Performance comparison of given ML classifiers in terms of testing time w.r.t. number of EEG samples of epileptic patients

DT classifiers. Overall value of testing time of Random Forest classifier remains within acceptable limits, making it suitable for real time classification of epileptic seizures.

The performance of Random Forest classification algorithm for total 50,000 samples of epileptic patients in terms of sensitivity and specificity values of individual classes is given in Table 2. It is evident from Table 2 that Random Forest classifier is giving optimum performance in terms of sensitivity and specificity values for detection of different epileptic seizure stages such as normal, preictal and ictal. Apart from given performance analysis discussed in this section, the other performance measures related to different errors being occurred during classification process are also taken into consideration. These performance measures related to errors are mean square error (MSE), root mean squared error (RMSE), relative absolute error (RAE) and root relative absolute error (RRAE). Table 3 shows the performance evaluation of various ML classifiers in terms different error values. It is again apparent from Table 3 that Random Forest provides best performance among all other given ML classifiers with least values of all the errors such as mean square error value of 0.0871, root mean square error value of 0.1489, relative absolute error value of 30.4071% and root relative absolute error value of 32.2471% only.

Table 2 Performance analysis of Random Forest classifier in terms of sensitivity and specificity for detection of different epileptic seizure stages

S. no.	Epileptic seizure stages	*SENS* (%)	*SPEC* (%)
1	Normal	100	99.33
2	Preical	100	99.66
3	Ictal	97	100

Table 3 Comparison of error performance for given ML classifiers for prediction of epileptic seizures with given dataset of 50,000 EEG samples of epileptic patients

S. no.	ML classifiers	MSE	RMSE	RAE (%)	RRAE (%)
1	Bayes net	0.2419	0.4146	56.6786	89.7596
2	Naïve Bayes	0.2853	0.463	66.851	100.2448
3	RBF	0.3114	0.3945	72.9778	85.1418
4	MLP	0.229	0.3534	53.6596	76.51
5	C4.5 decision tree	0.1184	0.2425	27.7317	52.496
6	Random forest	0.0871	0.1489	20.4071	32.2471

Similarly, the performance of this proposed HOS based feature extraction and Random Forest based classification technique for epileptic seizure early detection system is also compared with other existing techniques of epileptic seizure detection from the open literature in Table 4. As per this table, a technique presented by Jaiswal and Banka (2017b) has achieved 100% accuracy on test and evaluation datasets at the cost of total combinational time (training time + testing time) of 0.985 s. On the other hand, the proposed model has achieved 99.40% accuracy with very small value of total computation time of 0.5 s only. Although, the proposed model is also capable of getting 100% accuracy with increase in number of iterations to 100, but it will be achieved at the cost of higher computational time, which is not good enough for real time applications. Some other researchers (Jaiswal and Banka 2017a, 2018a, b) have also obtained 100% accuracy for binary classification of normal vs. preictal, normal vs. ictal and preictal vs. ictal stages. But present research work has explored 3-class or 3-label classification among three seizure stages such as normal vs. preictal vs. ictal, which is essentially more convenient way to address the issue of epileptic seizure detection. The other techniques discussed in this table show lower performance as compared to the proposed system in terms of classification accuracy, sensitivity and specificity. Thus, it has been again proved that the proposed HOS based Random Forest classification technique is an optimum way of detecting epileptic seizures accurately in real time scenario.

Thus, from above discussion, it is evident that Random Tree classification algorithm provides better performance among all other given classifiers for classification of different epileptic seizure stages such as normal, preictal and ictal stage. Therefore, this classifier makes the proposed model an effective technique for real time and early detection of epileptic seizures using cloud computing.

Table 4 Performance comparison of the proposed model with other models given in recent literature

Authors, year	Feature extraction	ML classifiers	Results
Jaiswal and Banka (2018a)	Sub pattern PCA (SpPCA) and Cross-sub pattern cross-correlation PCA (SubXPCA)	SVM	Accuracy = 92.76% (SpPCA), Accuracy = 96.66% (SubXPCA)
Singh and Malhotra (2018)	HOSA	C4.5 decision tree	Accuracy = 87.668%, Sensitivity = 87.7%, Specificity = 93.32%
Hu et al. (2019)	CNN based features containing mean amplitude spectrum of 19 subbands taken from 18 EEG channels	SVM	Accuracy = 86.25%
Cui et al. (2018)	Bag-of-waves based feature extraction for multichannel EEG	Extreme learning machine (ELM)	Sensitivity = 88.24%
Jaiswal and Banka (2018b)	Local centroid pattern (LCP) and 1D- local ternary pattern (1D-LTP)	ANN	Accuracy = 98% (LCP), Accuracy = 98.33% (1D-LTP)
Xiaofeng Liu et al. (2017)	Incremental entropy	SVM	Accuracy = 97.32%, Sensitivity = 95.34%, Specificity = 99.3%
Liu et al. (2017)	Crest factor, kurtosis, impulse factor, signal factor, wavelet packet energy and entropy features	Kernel ELM	Accuracy = 96.5%
Wang et al. (2017)	Multi domain feature extraction in time domain, frequency domain, time-frequency domain and information theory based non-linear analysis features, feature selection using PCA and ANOVA	SVM	Accuracy = 99.25%
Patidar and Panigrahi (2017)	Tunable Q wavelet transform (TQWT) and Kraskov entropy	LS-SVM	Accuracy = 97.75%, Sensitivity = 97%, Specificity = 99%
Jaiswal and Banka (2017b)	Local neighbor descriptive pattern (LNDP) and 1D- local gradient pattern (1D-LGP)	ANN	Accuracy = 97.06±0.62% (1D-LGP), Accuracy = 98.22±0.45% (LNDP), Accuracy = 100% (on test set), total comp. time = 0.985 s
Sareen et al. (2016)	Entropy based feature Extraction using FWHT and HOSA	k-means classification algorithm	Accuracy = 94.6%, Sensitivity = 93.8%, Specificity = 92.3%
Sareen et al. (2016)	FWHT and HOSA based entropy features	Gaussian Process	Accuracy = 85.10%, Sensitivity = 83.6%, Specificity = 16.3%
Samiee et al. (2015)	Discrete short time fourier transform	MLP	Accuracy = 98.1%, Sensitivity = 99.2%, Specificity = 93.8%
Song et al. (2012)	Optimized sample entropy	ELM	Accuracy = 99%
Acharya et al. (2012)	Approximate entropy (ApEn), sample entropy (SampEn), phase entropy 1 (S1), and phase entropy 2 (S2) features	Fuzzy Sugeno classifier	Accuracy = 98.1%
Proposed model	Fast walsh hadamard transform (FWHT) and Higher order spectra (HOS) based features and feature selection using Correlation based feature selection algorithm	Random forest	Accuracy = 99.4%, Sensitivity = 99.4%, Specificity = 99.66% , Training time = 20 ms, Testing time = 480 ms, Total comp. time = 0.5 s

5 Conclusion

In the age of internet of things (IoT), real time early detection system for epileptic seizures may act as a life savior for thousands of epileptic patients. This paper has proposed an automatic epileptic seizure early detection system in order to detect epileptic seizures at early stage. This model analyses sensed EEG data at cloud server using FWHT based preprocessing, HOS based feature extraction and Random Forest based classification.

The performance of this model has been tested and compared with other well-known machine learning algorithms like Bayes Net, Naïve Bayes, Multilayer Perceptron, RBF Neural Network and C4.5 decision tree classifiers. Random Forest classifier gives maximum accuracy of 99.40%, maximum sensitivity value of 99.40% and maximum specificity of 99.66% with an optimum training and testing times of

20 ms and 480 ms only for real time classification of epileptic seizures. This algorithm also provides 0.9906 value of Youden's index, 292.36 value of positive likelihood and 5.9467 value of discriminant power which are maximum among all other given ML algorithms, while having minimum value of 0.006 for negative likelihood and minimum values of error measures among other ML algorithms. This overall analysis and comparison with existing epileptic seizure detection techniques conclude that Random Tree classifier is an effective and fast machine learning algorithm for early detection of epileptic seizures in real time.

This model can play an effective role in medical healthcare by facilitating epilepsy patients through early detection epileptic seizures, so that patients suffering from this chronological disorder, their family members and doctors in nearby hospitals can be alerted in the preictal stage of the patient only and lives of patients can be saved. This model could enable the doctors in remote monitoring of epileptic patients living at distant places in remote and underprivileged areas.

Acknowledgements The authors of this paper are highly grateful to Research and Development laboratory of Guru Nanak Dev University, Amritsar, Punjab, India for providing research facilities to carry out this research work. Also, the authors would like to thank the reviewers in advance for their invaluable comments and suggestions.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

AbdulGhaffar A, Mostafa SM, Alsaleh A, Sheltami T, Shakshuki EM (2019) Internet of things based multiple disease monitoring and health improvement system. J Ambient Intell Hum Comput. https://doi.org/10.1007/s12652-019-01204-6

Acharya UR, Molinari F, Sree SV, Chattopadhyay S, Ng KH, Suri JS (2012) Automated diagnosis of epileptic eeg using entropies. Biomed Signal Process Control 7(4):401–408. https://doi.org/10.1016/j.bspc.2011.07.007

Alickovic E, Subasi A, Initiative ADN, et al. (2019) Automatic detection of Alzheimer disease based on histogram and random forest. In: International conference on medical and biological engineering, Springer, pp 91–96

Alotaiby T, El-Samie FEA, Alshebeili SA (2015) Ahmad I (2015) A review of channel selection algorithms for eeg signal processing. EURASIP J Adv Signal Process 1:66. https://doi.org/10.1186/s13634-015-0251-9

Andrzejak RG, Lehnertz K, Mormann F, Rieke C, David P, Elger CE (2001) Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. Phys Rev E 64(6):061907

Azimi I, Rahmani AM, Liljeberg P, Tenhunen H (2017) Internet of things for remote elderly monitoring: a study from user-centered perspective. J Ambient Intell Hum Comput 8(2):273–289. https://doi.org/10.1007/s12652-016-0387-y

Baratloo A, Hosseini M, Negida A, El Ashal G (2015) Part 1: simple definition and calculation of accuracy, sensitivity and specificity. Emergency 3(2):48–49

Bentlemsan M, Zemouri E, Bouchaffra D, Yahya-Zoubir B, Ferroudji K (2014) Random forest and filter bank common spatial patterns for eeg-based motor imagery classification. In: 2014 5th International conference on intelligent systems, modelling and simulation, pp 235–238. https://doi.org/10.1109/ISMS.2014.46

Cheng J, Greiner R (2001) Learning Bayesian belief network classifiers: Algorithms and system. In: Conference of the Canadian society for computational studies of intelligence. Springer, pp 141–151

Chua K, Chandran V, Acharya UR, Lim C (2009) Automatic identification of epileptic electroencephalography signals using higher-order spectra. Proc Inst Mech Eng [H] 223(4):485–495

Chua KC, Chandran V, Acharya UR, Lim CM (2010) Application of higher order statistics/spectra in biomedical signals—a review. Med Eng Phys 32(7):679–689

Cui S, Duan L, Qiao Y, Xiao Y (2018) Learning eeg synchronization patterns for epileptic seizure prediction using bag-of-wave features. J Ambient Intell Hum Comput. https://doi.org/10.1007/s12652-018-1000-3

Cutler A, Cutler DR, Stevens JR (2012) Random forests. In: Ensemble machine learning. Springer, pp 157–175

Darwish A, Hassanien AE, Elhoseny M, Sangaiah AK, Muhammad K (2017) The impact of the hybrid platform of internet of things and cloud computing on healthcare systems: opportunities, challenges, and open problems. J Ambient Intell Hum Comput. https://doi.org/10.1007/s12652-017-0659-1

Emotiv (2018) Emotiv EPOC wearable EEG Headset. https://www.emotiv.com/epoc/. Accessed 01 Apr 2018

Fraiwan L, Lweesy K, Khasawneh N, Wenz H, Dickhaus H (2012) Automated sleep stage identification system based on time-frequency analysis of a single eeg channel and random forest classifier. Comput Methods Progr Biomed 108(1):10–19

Freestone DR, Karoly PJ, Cook MJ (2017) A forward-looking review of seizure prediction. Curr Opin Neurol 30(2):167–173

Gajic D, Djurovic Z, Di Gennaro S, Gustafsson F (2014) Classification of eeg signals for detection of epileptic seizures based on wavelets and statistical pattern recognition. Biomed Eng Appl Basis Commun 26(02):1450021

Holmes G, Donkin A, Witten IH (1994) Weka: a machine learning workbench. In: Proceedings of ANZIIS '94—Australian New Zealand intelligent information systems conference, pp 357–361. https://doi.org/10.1109/ANZIIS.1994.396988

Homan RW, Herman J, Purdy P (1987) Cerebral location of international 10–20 system electrode placement. Electroencephalogr Clin Neurophysiol 66(4):376–382. https://doi.org/10.1016/0013-4694(87)90206-9

Hu W, Cao J, Lai X, Liu J (2019) Mean amplitude spectrum based epileptic state classification for seizure prediction using convolutional neural networks. J Ambient Intell Hum Comput. https://doi.org/10.1007/s12652-019-01220-6

IEC (2018) Indian Epilepsy Centre, New Delhi. http://www.indianepilepsycentre.com/what-is-epilepsy.html. Accessed 01 Feb 2018

Islam SMR, Kwak D, Kabir MH, Hossain M, Kwak K (2015) The internet of things for health care: a comprehensive survey. IEEE Access 3:678–708. https://doi.org/10.1109/ACCESS.2015.2437951

Jaiswal AK, Banka H (2017a) Epileptic seizure detection in eeg signal with gmodpca and support vector machine. Bio-Med Mater Eng 28(2):141–157

Jaiswal AK, Banka H (2017b) Local pattern transformation based feature extraction techniques for classification of epileptic eeg signals. Biomed Signal Process Control 34:81–92. https://doi.org/10.1016/j.bspc.2017.01.005

Jaiswal AK, Banka H (2018a) Epileptic seizure detection in eeg signal using machine learning techniques. Aust Phys Eng Sci Med 41(1):81–94. https://doi.org/10.1007/s13246-017-0610-y

Jaiswal AK, Banka H (2018b) Local transformed features for epileptic seizure detection in eeg signal. J Med Biol Eng 38(2):222–235. https://doi.org/10.1007/s40846-017-0286-5

Jia S, Hu X, Sun L (2013) The comparison between random forest and support vector machine algorithm for predicting β-hairpin motifs in proteins. Engineering 5(10):391

Klok AB, Edin J, Cesari M, Olesen AN, Jennum P, Sorensen HBD (2018) A new fully automated random-forest algorithm for sleep staging. In: 2018 40th annual international conference of the ieee engineering in medicine and biology society (EMBC), pp 4920–4923. https://doi.org/10.1109/EMBC.2018.8513413

Korting TS (2006) C4. 5 algorithm and multivariate decision trees. Image Processing Division, National Institute for Space Research–INPE Sao Jose dos Campos–SP, Brazil

Litt B, Esteller R, Echauz J, D'Alessandro M, Shor R, Henry T, Pennell P, Epstein C, Bakay R, Dichter M, Vachtsevanos G (2001) Epileptic seizures may begin hours in advance of clinical onset: a report of five patients. Neuron 30(1):51–64. https://doi.org/10.1016/S0896-6273(01)00262-8

Liu Q, Zhao X, Hou Z, Liu H (2017) Epileptic seizure detection based on the kernel extreme learning machine. Technol Health Care 25(S1):399–409

Malasinghe LP, Ramzan N, Dahal K (2019) Remote patient monitoring: a comprehensive study. J Ambient Intell Hum Comput 10(1):57–76. https://doi.org/10.1007/s12652-017-0598-x

Manogaran G, Lopez D (2017) A survey of big data architectures and machine learning algorithms in healthcare. Int J Biomed Eng Technol 25(2–4):182–211

McDonald AD, Lee JD, Schwarz C, Brown TL (2014) Steering in a random forest: Ensemble learning for detecting drowsiness-related lane departures. Hum Factors 56(5):986–998

Mora H, Gil D, Terol RM, Azorín J, Szymanski J (2017) An iot-based computational framework for healthcare monitoring in mobile environments. Sensors 17(10):2302

Moreira MWL, Rodrigues JJPC, Carvalho FHC, Chilamkurti N, Al-Muhtadi J, Denisov V (2019) Biomedical data analytics in mobile-health environments for high-risk pregnancy outcome prediction. J Ambient Intell Hum Comput 10(10):4121–4134. https://doi.org/10.1007/s12652-019-01230-4

NeuroSky (2018) NeuroSky MindWave EEG Headset. https://www.neurosky.com/biosensors/eeg-sensor/biosensors/. Accessed 01 Apr 2018

Neurotech (2016) Neurotech Hardware Roundup 2016. http://www.autodidacts.io/neurotech-hardware-roundup-eeg-bci-tdcs-neurofeedback/. Accessed 05 Apr 2018

Nikias CL, Mendel JM (1993) Signal processing with higher-order spectra. IEEE Signal Process Mag 10(3):10–37

NINDS (2018) National Institute of Neurological Disorders and Stroke. https://www.ninds.nih.gov/Current-Research/Focus-Research/Focus-Epilepsy. Accessed 25 Jan 2018

Orosco L, Correa AG, Laciar E (2013) A survey of performance and techniques for automatic epilepsy detection. J Med Biol Eng 33(6):526–537

Patidar S, Panigrahi T (2017) Detection of epileptic seizure using kraskov entropy applied on tunable-q wavelet transform of eeg signals. Biomed Signal Process Control 34:74–80

Pop I (2006) An approach of the naive bayes classifier for the document classification. Gen Math 14(4):135–138

Rokach L (2010) Ensemble-based classifiers. Artif Intell Rev 33(1):1–39. https://doi.org/10.1007/s10462-009-9124-7

Samiee K, Kovács P, Gabbouj M (2015) Epileptic seizure classification of eeg time-series using rational discrete short-time fourier transform. IEEE Trans Biomed Eng 62(2):541–552. https://doi.org/10.1109/TBME.2014.2360101

Sareen S, Sood SK, Gupta SK (2016) An automatic prediction of epileptic seizures using cloud computing and wireless sensor networks. J Med Syst 40(11):1–18. https://doi.org/10.1007/s10916-016-0579-1

Sareen S, Sood SK, Gupta SK (2016) A cloud-based seizure alert system for epileptic patients that uses higher-order statistics. Comput Sci Eng 18(5):56–67. https://doi.org/10.1109/MCSE.2016.82

Sathyanarayana S, Satzoda RK, Sathyanarayana S, Thambipillai S (2018) Vision-based patient monitoring: a comprehensive review of algorithms and technologies. J Ambient Intell Hum Comput 9(2):225–251. https://doi.org/10.1007/s12652-015-0328-1

Seijas C, Caralli A, Villazana S (2013) Neuropathology classifier based on higher order spectra. J Comput Commun 1(04):28

Senagi K, Jouandeau N, Kamoni P et al (2017) Using parallel random forest classifier in predicting land suitability for crop production. J Agric Inform 8(3):23–32

Shoeb A, Guttag J (2010) Application of machine learning to epileptic seizure detection. In: Proceedings of the 27th international conference on international conference on machine learning, Omnipress, USA, ICML'10, pp 975–982

Singh K, Agrawal S (2011) Performance evaluation of five machine learning algorithms and three feature selection algorithms for ip traffic classification. IJCA Spec Issue Evol Netw Comput Commun 1:25–32

Singh K, Agrawal S (2011) Comparative analysis of five machine learning algorithms for ip traffic classification. In: 2011 International conference on emerging trends in networks and computer communications (ETNCC), pp 33–38. https://doi.org/10.1109/ETNCC.2011.5958481

Singh K, Malhotra J (2018) Iot enabled epileptic seizure early detection system using higher order spectral analysis and c 4.5 decision tree classifier. In: 2018 5th International Conference on "Computing for Sustainable Global Development(IndiaCom-2018), Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), New Delhi (INDIA), pp 1105–1110

Sokolova M, Japkowicz N, Szpakowicz S (2006) Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation. In: Australasian joint conference on artificial intelligence, Springer, pp 1015–1021

Song Y, Crowcroft J, Zhang J (2012) Automatic epileptic seizure detection in eegs based on optimized sample entropy and extreme learning machine. J Neurosci Methods 210(2):132–146. https://doi.org/10.1016/j.jneumeth.2012.07.003

Subasi A, Ahmed A, Aličković E, Hassan AR (2019) Effect of photic stimulation for migraine detection using random forest and discrete wavelet transform. Biomed Signal Process Control 49:231–239

Subha DP, Joseph PK, Acharya R, Lim CM (2010) Eeg signal analysis: a survey. J Med Syst 34(2):195–212

Ting KM (2017) Confusion matrix. In: Encyclopedia of machine learning and data mining, pp 260–260

Upadhyay R, Manglick A, Reddy D, Padhy P, Kankar P (2015) Channel optimization and nonlinear feature extraction for electroencephalogram signals classification. Comput Electr Eng 45:222–234

Wang L, Xue W, Li Y, Luo M, Huang J, Cui W, Huang C (2017) Automatic epileptic seizure detection in eeg signals using multi-domain feature extraction and nonlinear analysis. Entropy 19(6):222

WHO (2018) World Health Organization. https://www.who.int/mentalhealth/neurology/epilepsy/en/. Accessed 27 Jan 2018

Xiaofeng Liu, Aimin Jiang, Ning Xu (2017) Automated epileptic seizure detection in eegs using increment entropy. In: 2017 IEEE

30th Canadian conference on electrical and computer engineering (CCECE), pp 1–4. https://doi.org/10.1109/CCECE.2017.7946705
Yuvaraj R, Acharya UR, Hagiwara Y (2018) A novel parkinson's disease diagnosis index using higher-order spectra features in eeg signals. Neural Comput Appl 30(4):1225–1235
Zhang Y, Xin Y, Li Q, Ma J, Li S, Lv X, Lv W (2017) Empirical study of seven data mining algorithms on different characteristics of datasets for biomedical classification applications. BioMed Eng OnLine 16(1):125. https://doi.org/10.1186/s12938-017-0416-x