

Classification of EEG Signals in Seizure Detection System Using Ellipse Area Features and Support Vector Machine



Dattaprasad A. Torse, Veena Desai and Rajashri Khanai

Abstract Epilepsy is a brain disorder, characterized by transitory and impulsive electrical signal of the brain. The electroencephalogram (EEG)-based seizure detection system used for automated diagnosis of epilepsy requires optimum classification of signals. This paper presents an optimized method for classification as normal and epileptic signals using Empirical-Mode Decomposition (EMD) technique. The Intrinsic-Mode Functions (IMFs) are few symmetric and band-limited signals obtained by applying EMD to the signals. However, optimum selection of IMF features is crucial step in deciding feature set for classification. The 95% confidence ellipse area is calculated from the Second-Order Difference Plot (SODP) of selected IMFs to form features for classification. The feature space is used by the Cosine Similarity Measure Support Vector Machine (CSM-SVM) classifier with optimum feature selection. It is observed that the features formed using ellipse area of dissimilar combination of IMFs have given superior classification performance on EEG data available from the Bonn University, Germany.

Keywords Empirical-mode decomposition · Intrinsic-mode function · Ellipse area · Cosine similarity measure support vector machine

D. A. Torse (✉) · V. Desai

Department of Electronics and Communication Engineering, KLS Gogte Institute of Technology, Belagavi, Karnataka 590008, India
e-mail: datorse@git.edu

V. Desai

e-mail: veenadesai@git.edu

R. Khanai

Department of Electronics and Communication Engineering, KLE's Dr. M.S.S.C.E.T, Belagavi, Karnataka 590008, India

© Springer Nature Singapore Pte Ltd. 2019

A. J. Kulkarni et al. (eds.), *Proceedings of the 2nd International Conference on Data Engineering and Communication Technology*, Advances in Intelligent Systems and Computing 828, https://doi.org/10.1007/978-981-13-1610-4_9

1 Introduction

Epilepsy is an acute and recurring neurological confusion generally evident by frequent seizures which has an effect on about 1% of the world's population [1]. It results in worsening of consciousness and may show random and recurrent body convulsions. The developing countries contribute about 85% of the 60 million people affected by epilepsy worldwide. EEG recording characterizes electrical activities of the brain by measuring electrical potential of neurons. The EEG is recorded a prominent tool in the detection of epileptic seizures and includes recognition of spikes in the epilepsy identification [2]. Further, the accurate diagnosis of epilepsy helps in deciding the course of the antiepileptic medication. The automated seizure detection system generally involves three steps, namely preprocessing, signal transformation feature extraction and classification [3]. In joint time-frequency techniques, one finds two-dimensional (2D) time-frequency representation of filtered EEG [4], using Hilbert–Huang Transform (HHT) and its variant Empirical-Mode Decomposition (EMD) [5].

The proposed method employs two signal processing techniques for segregation of normal and epileptic EEG signals. In the first step, the EMD method is applied on the EEG signals for the extraction of the IMFs followed by SODP which measures the rate of variability of individual IMFs [6]. In the second step, the 95% confidence ellipse areas of SODPs of two IMFs is computed and used as a feature set for categorization of the EEG signal information using Cosine Similarity Measure Support Vector Machine (CSM-SVM) classifier [7].

The remaining paper is ordered as follows: Sect. 2 explains the data set, EMD, SODP and ellipse area formulation, selection of features based on CSM, CSM-SVM-based classification, and estimation of statistical parameters. The result and discussion are presented in Sect. 3. Conclusion has been drawn in Sect. 4.

2 Methodology

2.1 Data Set

The EEG data set information that is accessible online publicly and explained in [8] is used to validate the results of this work. The data set with normal and seizure EEG signals of five sub-signals is represented as A, B, C, D and E, each having 100 numbers of single-channel EEG signals. The duration of each sample is 23.6 s with sampling rate of 173.61 Hz. The subsets A and B are recorded with international 10–20 standard electrode placement scheme with surface EEG recordings. The set A is for healthy subjects with eyes open and B for eyes closed, the subsets C and D are recorded in normal intervals from five patients in the epileptogenic zone (subset C) and from the hippocampal formation of the opposite hemisphere of the brain (subset D). The typical EEG signal samples of normal and seizure are shown in Fig. 1.

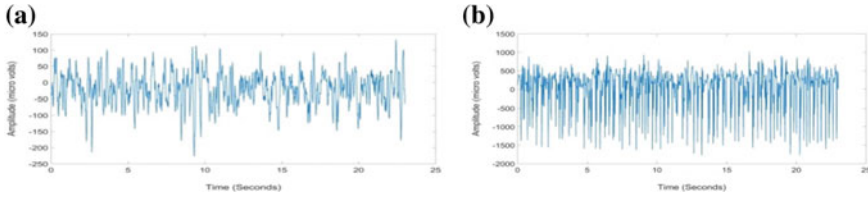


Fig. 1 EEG signals **a** normal and **b** seizure

2.2 Empirical-Mode Decomposition

EMD is a data-driven technique that splits a signal into a finite set of amplitude and frequency modulated (AM-FM) oscillating components. These components are called as IMFs. An IMF is a function which satisfies the conditions as follows: (1) number of maxima and minima are either equal or vary almost by one. (2) The mean value of envelopes, characterized by local maxima and minima, is zero. This signal-dependent decomposition is adaptively done for predefined levels. Moreover, the decomposition of long duration signals is done without a stationary and linearity condition of the signal. The EMD algorithm is used to obtain number of band-limited functions from a signal $x(t)$. The EMD algorithm for the signal $x(t)$ is depicted below [5]:

1. Original signal is set as $x(t) = P_1(t)$.
2. Both the maximum and minimum values for $P_1(t)$ are estimated.
3. The cubic spline interpolation is used to establish the upper and lower envelopes represented as $env_{upper}(t)$ and $env_{lower}(t)$, respectively. To achieve this, the maximum and minimum points are joined.
4. The local mean is computed as $m(t) = (env_{lower}(t) + env_{upper}(t))/2$.
5. The signal $m(t)$ is subtracted from the given signal as $P_1(t) = P_1(t) - m(t)$ means IMF should have zero local mean.
6. To $P_1(t)$ is tested for above defined conditions to check if it is an IMF or not.
7. The steps (2)–(6) are repeated and when an IMFs are computed, the process is stopped.

The sifting is continued until the last IMF is generated. The computation functions and the residues are given by:

$$r_1(t) - P_2(t) = r_2(t), \dots, r_{IMF-1}(t) - P_{IMF}(t) = r_{IMF}(t) \quad (1)$$

where $r_{IMF}(t)$ is called the final signal remained. The final stage of decomposition results is the signal $x(t)$ given by:

$$x(t) = \sum_{p=1}^M P_p(t) + r_{IMF}(t) \quad (2)$$

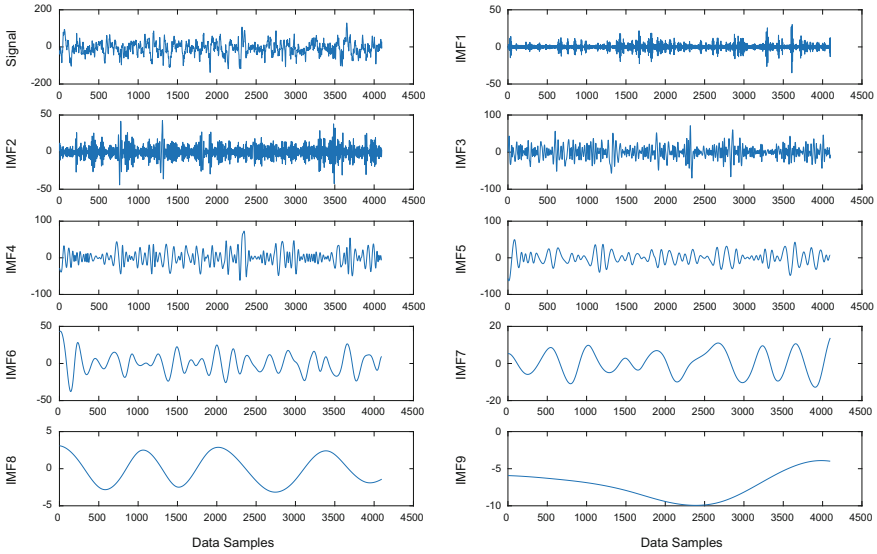


Fig. 2 Intrinsic-mode functions of normal EEG signal

where the number of IMFs is represented by M and $r_{IMF}(t)$ is the final signal remained.

The implementation of the EMD algorithm is done using MATLAB. The implementation of EMD on the 23.6 s EEG signals for 4097 samples is shown in Fig. 2.

2.3 Second-Order Difference Plot and Calculation of Ellipse Area

A nonlinear signal can be viewed with a different perspective, and a continuous chaotic modelling is an effective tool in understanding the nonlinearity in long data series. To classify normal and seizure EEG signals, the SODP of IMFs of EEG signals present important parameters.

For the signal $x(n)$, the SODP is achieved by plotting $X(n)$ versus $Y(n)$ which are defined as [6]:

$$X(n) = x(n + 1) - x(n) \tag{3}$$

$$Y(n) = x(n + 2) - x(n + 1) \tag{4}$$

The chaotic equations are employed in generating graphs, which are known as Poincare plots. The equation below represents a Poincare equation, which is an example of a chaotic system.

$$a_n = Aa_{n-1}(1 - a_{n-1})2 \leq A \leq 4 \quad (5)$$

where A is a constant. The performance of the function is depending on this value.

Recently, the variability measured from the SODP has been used for analysis of EEG signals [6]. The 95% confidence ellipse area measured from the SODP of IMFs of EEG signal data represents a set of parameters for segregation of normal and seizure EEG signals. Figure 2 shows the IMFs of normal signal patterns. It motivates to work out the ellipse area of ‘‘SODP of IMFs’’ for separation of EEG signals. The method to calculate ‘‘95% confidence ellipse area’’ from the SODP approach is presented as follows [6]: The mean value of $X(n)$ and $Y(n)$ is computed as:

$$S_X = \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} X(n)^2} \quad S_Y = \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} Y(n)^2} \quad S_{XY} = \frac{1}{N} \sqrt{\sum X(n)Y(n)} \quad (6)$$

The D value is computed as:

$$D = \sqrt{(S_X^2 + S_Y^2) - 4(S_X^2 S_Y^2 - S_{XY}^2)} \quad (7)$$

The ellipse area is formulated as:

$$A_{\text{ellipse}} = P_i(ab); \quad a = \sqrt{3 \times (S_X^2 + S_Y^2 + D)}; \quad b = \sqrt{3 \times (S_X^2 + S_Y^2 - D)} \quad (8)$$

2.4 Cosine Similarity Measure Support Vector Machine (CSM-SVM) Classifier

The SVM is a powerful machine learning algorithm for regression and classification which produces very precise classification results. SVMs are considered as an important example of ‘‘kernel methods’’, one of the key areas in machine learning, and the Radial Basis Function (RBF) kernel has been employed in this work. The RBF is given by:

$$K(x_i, x_j) = \exp\left(-\left(\|x_i - x_j\|^2 / 2\rho^2\right)\right) \quad (9)$$

where $\rho > 1$ is the parameter controlling the width of the kernel. The Cosine Similarity Index (CSI) between two data vectors is a standard criterion for finding the distance between two data samples. CSI determines the cosine of the angle between the data samples. To construct the cosine resemblance equation, the equation of the dot product for the $\cos \theta$ is to be solved as:

$$\vec{a} \cdot \vec{b} = \|\vec{a}\| \|\vec{b}\| \cos \theta; \quad \text{CSI} = \cos \theta = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \|\vec{b}\|} \quad (10)$$

where CSI represents the whitened cosine similarity value between vector a and b [9].

2.5 Performance Evaluation

The assessment of the SVM-based classifier for classification of seizure and non-seizure EEG signals is carried out by calculating the sensitivity, specificity and accuracy. The sensitivity (SE), specificity (SP) and accuracy (AC) can be defined as:

$$SE = \frac{TP}{TP + FN} \times 100\%; \quad SP = \frac{TN}{TN + FP} \times 100\%; \quad (11)$$

$$AC = \frac{TP}{TP + FN + FP + FN} \times 100\% \quad (12)$$

where TP and TN signify the overall number of appropriately detected true positive patterns and true negative patterns, respectively. The FP signifies overall number of erroneously positive patterns, and FN signifies erroneously negative patterns. The positive and negative patterns signify detected seizure and detected normal EEG signals, respectively.

3 Results and Discussion

In the first step, the results are obtained by decomposing EEG signals from data sets B and E by EMD method to obtain nine IMFs as shown in Fig. 2. The first four IMFs are selected to compute SODP, and 95% ellipse area as the maximum frequency variation and nonlinearity of the signal is obtained from first few IMFs. The classification of normal and seizure EEG signals is performed using the ellipse area parameters of SODP for first four IMF's pairs. Figure 3 shows plot of SODP for all the IMFs achieved by EMD process. The compute 95% ellipse area of SODP of combination of initial four IMFs, the six pairs has been selected as IMF12, IMF13, IMF14, IMF23, IMF24, and IMF34.

The plot of 95% confidence ellipse area of these six pairs is shown in Fig. 4. Generally, first pair of IMFs represents frequency variation with more values. The 95% confidence area as a statistical feature is computed from the pair of IMFs as it clearly represents the nature of the underlying EEG signal. The maximum and minimum scaled values of 95% confidence are shown in Table 1 for normal and seizure data samples. The values for IMF12 pair are 19,642.13 and 123.92 respectively. The first IMF and its associated SODP show that the feature set can distinguish between normal and seizure signals. The elliptical shape of SODP of IMFs encourages computation of 95% confidence ellipse area of SODP of IMFs of EEG signals. As with the seizure-free case, similar observation can be drawn for the

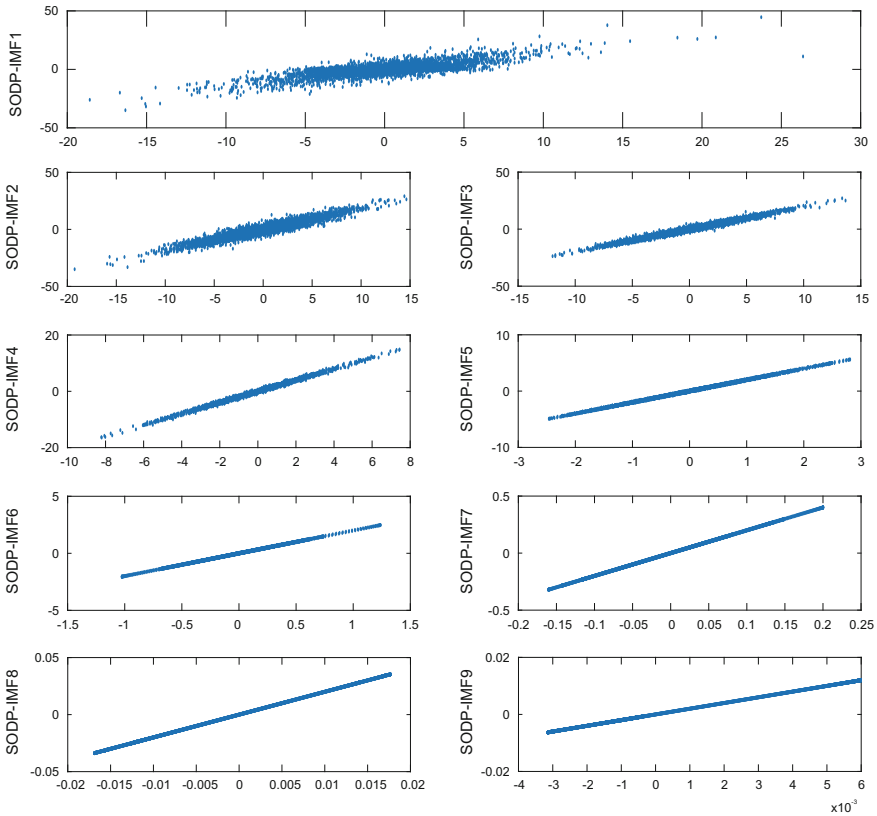


Fig. 3 Second-order difference plot of IMFs of normal EEG signal

case of ictal EEG signal also. The 95% of confidence ellipse area parameters has been computed for ictal and non-ictal classes using SODP of IMFs for a range of data values of the signal (500, 1000, 1500 and 4097 samples). The class discriminating performance of ellipse area feature is verified using box plots.

The IMF’s ellipse area parameters of SODP of IMFs are considerably diversified for the two classes of EEG signals (normal and seizure). The MATLAB function “*tcd*” is used to compute the p -value of the data classified with defined degree of freedom. It is found that $p \leq 0.01$ indicates the 1%, i.e. less than 1 in 100 chance of being wrong (refer to Fig. 4).

The comparison of classifier performance reveals that the division of EEG time series into four sub-bands results in varied performance parameters. The performance for 1500 and 4097 data sample is depicted in Table 2. The ellipse area features obtained from shape of the SODP of IMFs result in more suitable features for a data segment of 1500. Figure 5 shows the box plot of ellipse area of SODP of IMFs for

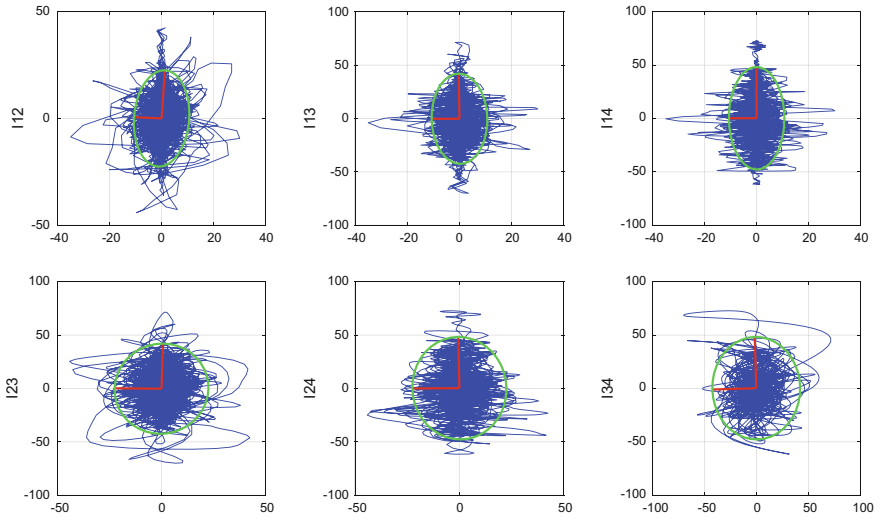


Fig. 4 Ellipse area of IMF pairs for initial four IMFs of normal signal

Table 1 95% confidence area for six IMF pairs for normal and seizure EEG data (80 samples of normal EEG)

EEG data	IMF pairs	IMF12	IMF13	IMF14	IMF23	IMF24	IMF34
Normal data (80 samples)	Min.	123.92	246.12	241.89	450.45	580.36	1280.46
	Max.	19,642.13	26,486.63	29,694.69	46,342.25	51,957.78	69,503.2
Seizure data (80 samples)	Min.	7933.89	8723.08	11,605.05	10,254.90	16,020.05	16,305.7
	Max.	3,884,243.3	3,231,613.7	1,783,224.9	3,346,937.3	1,547,209.4	1,795,701.2

normal and seizure EEG signals. The experimental results are found noteworthy for inequity among normal and ictal signals.

The CSM-SVM is an optimized form of SVM which uses wrapper method to select features suitable for classification and implemented in MATLAB. The toolbox is constructed around a fast LS-SVM training and simulation algorithm [10]. The LS-SVM functions have been used for classification as well as for plotting the performance. A fast LS-SVM training and simulation algorithm is present in the MATLAB toolbox. The LS-SVM functions have been used for categorization in addition to that for plotting the performance.

Table 2 The performance of LS-SVM classifier for 1500 and 4097 data samples

EEG data	Performance parameters	IMF12	IMF13	IMF14	IMF23	IMF24	IMF34
1500 data samples	SE	95.2	94.2	96.3	95.4	92.1	88.3
	SP	91.5	93.2	84.2	87.3	79.5	64.3
	AC	92.1	98.5	91.2	90.1	50.2	65.5
	AVG	96.2	95.5	75.5	88.5	91.5	90.3
4097 data samples	SE	96.3	95.1	90.2	89.3	85.5	79.3
	SP	95.1	90.5	95.3	94.1	94.4	91.7
	AC	91.2	90.3	40.5	82.5	92.5	50.1
	AVG	90.3	88.2	39.2	81.3	89.5	54.3

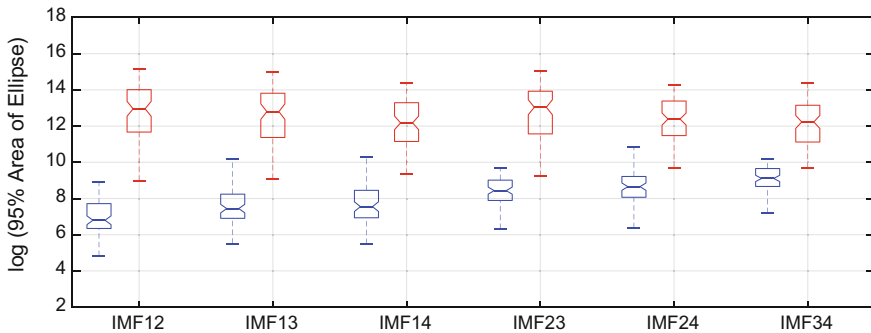


Fig. 5 Box plots of ellipse area of SODP of IMFs for normal and seizure EEG signals

4 Conclusion

Epilepsy is a chronic disorder, and EEG-based automated seizure detection system is evolving as an alternative diagnosis option where the main requirement is optimum signal processing algorithms. This paper presents the performance of classifier using ellipse area of SODP of IMFs as a feature set. For categorization of ictal and non-ictal EEG signals, the ellipse area is obtained using EMD algorithm to generate optimum feature space. The ellipse area parameters of first four IMFs are used in generating six pairs. The ellipse areas of first and second IMFs have provided better classification accuracy. The seizure EEG signals have significant ellipse area as compared to the normal EEG signals. The increased variation in seizure signals results into the desired variation in ellipse area of SODP of IMF pairs. The CSM-SVM provided highest classification accuracy of 96.4% for first pair of IMFs. The application of this method in automated seizure detection system will be tested on out-of-sample data sets. The system may also assist caretakers of epileptic patients by detecting occurrence of seizure signal.

References

1. Pati S, Alexopoulos AV (2010) Pharmacoresistant epilepsy: from pathogenesis to current and emerging therapies. *Cleve Clin J Med* 77(7):457–467
2. Sharma M, Pachori RB, Rajendra Acharya U (2017) A new approach to characterize epileptic seizures using analytic time–frequency flexible wavelet transform and fractal dimension. *Pattern Recogn Lett*
3. Acharya UR, Sree SV, Swapna G, Martis RJ, Suri JS (2013) Automated EEG analysis of epilepsy: a review. *Knowl Based Syst* 45:147–165
4. Torse DA, Desai VV (2016) Design of adaptive EEG preprocessing algorithm for neuro-feedback system. In: 2016 international conference on communication and signal processing (ICCSP). IEEE, pp 0392–0395
5. Rilling G, Flandrin P, Goncalves P (2003) On empirical mode decomposition and its algorithms. In: IEEE-EURASIP workshop on nonlinear signal and image processing, vol 3. IEEE, Grado, Italy
6. Pachori RB, Patidar S (2014) Epileptic seizure classification in EEG signals using second-order difference plot of intrinsic mode functions. *Comput Methods Programs Biomed* 113(2):494–502
7. Chen Gang, Chen Jin (2015) A novel wrapper method for feature selection and its applications. *Neurocomputing* 159:219–226
8. Andrzejak RG, Lehnertz K et al (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
9. Machine Learning: Cosine Similarity for Vector Space Models (Part III): <http://blog.christianperone.com/2013/09/machine-learning-cosine-similarity-for-vector-space-models-part-iii/>
10. Suykens JAK, Vandewalle J (1999) Least squares support vector machine classifiers. *Neural Process Lett* 9:293–300