

Efficient Seizure Classification using Multiclass Support Vector Machine with RFE

J. Praveena¹, Dr. S. Thiruvankadam²

ME Department of CSE, P.A College of Engineering and Technology, Pollachi, Tamil Nadu, India¹

Professor, Department of IT, P.A College of Engineering and Technology, Pollachi, Tamil Nadu, India²

ABSTRACT: A new multi classification scheme based on Support Vector Machine called Hierarchical multi-class support vector machine with a new Extreme Learning Machine kernel is introduced. The Recursive Feature Elimination (RFE) is used for ranking the features. Wavelets Transform based features such as Statistical Values, and Approximate Entropy have been used for the encephalogram signal classification with the emphasis on epileptic seizure detection. The feature extraction and classification are performed using the publicly available benchmarked datasets. In general support vector provides better classification accuracy, but takes more time for classification and also there is scope for a new multi classification scheme. Mitigating the problem of support vectors, here ranking of the features is done and features with low ranks are eliminated. This will speed up the classification process. Evaluation of the performance of the kernels, the accuracy of other kernels is compared using the benchmarked datasets. One of the major objectives was to examine the Hierarchical support vector with the sensitivity and specificity of encephalogram signals classification.

KEYWORDS: Support Vector Machine, encephalogram, feature extraction, classification.

I. INTRODUCTION

Automatic seizure detection methods in the diagnosis of epilepsy were developed in the early 1970s. In recent years, many algorithms for the detection of seizures have been proposed and applied, such as frequency domain analysis, time-frequency domain analysis, artificial neural network based analysis, and machine learning based analysis. For non-stationary EEG signals, time-frequency analysis method, such as Discrete Wavelet Transform (DWT), has been proved to be an effective analysis tool and could give quantitative evaluation of ictal EEG in different frequency bands.

Effective feature selection is important in order to represent all the characteristics features of the electroencephalogram signals. SVM classifier is used for epileptic seizure detection.

J.Gotman et al. [1982] proposed automatic detection of epileptic seizures in the EEG by online recording of EEG spikes and analysis of epochs in different signals. The method is aimed at the recognition of seizures from scalp and intracerebral recordings. No attempt was made to detect the classic 3/sec spike-and-wave bursts. It attempts to identify seizures that included, at least at sometime during their development, sustained paroxysmal rhythmic activity with a fundamental frequency anywhere between 3 and 20 c/sec.

T. Kalayci et al. [1995] introduced the wavelet preprocessing for automated neural network detection of EEG spikes. Modeled after properties of natural neurons, ANNs have the basic computing units that learn and recognize patterns in ways similar to humans. ANNs do not need any specific rules but only examples for training. Thus, ANNs provide the best solution to recognize and classify the tasks for the complete rules cannot be written.

C.J.C.Burges [1998] gave a paper consisting of the tutorial on the Support Vector Machine for pattern recognition. The tutorial starts with an overview of the concepts of vector dimension and structural risk minimization. It describes linear SVMs for separable and non-separable data, that works with a non-trivial example in detail.

V. N. Vapnik [1999] gave an overview on the statistical learning theory on that the Support Vector Machine is based upon.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

M. Zavar et al. [2011] gave an evolutionary model selection in a wavelet-based support vector machine for automated seizure detection. In this study, an evolutionary scheme searches for optimal kernel types and parameters for automated seizure detection. The Lyapunov exponent, fractal dimension and wavelet entropy for possible feature extraction are considered. The dataset has the electrocardiographic changes in patients with partial epilepsy that two types cardiographic beats like partial epilepsy and normal.

The multi class SVM is used with the hierarchical architecture using the radial basis kernel. The number of classes is reduced and hence the classification speed can be increases. The classification process uses the one versus rest strategy.

II. PROBLEM FORMULATION

Encephalogram signals are non-stationary signals that clearly depict the brain simulation. Support Vector Machines (SVM) are used to classify the signals if they are epileptic seizures or not. SVM classification is based on the kernel methods when the data sets are non-linearly separable. As the number of datasets increase, the classification becomes more complex and hence accuracy and performance becomes a bottleneck. So, a multi class SVM is introduced where the datasets are divided into different classes.

III. SUPPORT VECTOR MACHINE CLASSIFICATION

SVM classification is based on the kernel functions. Different kernel methods have been used out of which the radial basis kernel function is considered as a better kernel method. SVM is based on the statistical learning theory [15].

SVM maps the input patterns into a higher dimensional feature space through some nonlinear mapping chosen at beforehand. A linear decision surface is then constructed in this high-dimensional-feature space. The SVM is a binary classifier that can be extended by fusing several of its kind into a Multiclass classifier.

A. Kernel

A function that returns the value of the dot product between the images of the two arguments $K(x_1, x_2) = f(x_1) \cdot f(x_2)$. Given a function K , we can verify that if it is a kernel. One can use LLMs in a feature space by simply rewriting it in dual representation and replacing dot products with kernels.

The Kernel Matrix is the central structure in kernel machine and also contains all necessary information for the learning algorithm. It also fuses information about the data and kernel. The kernel matrix is Symmetric Positive Definite. Any symmetric positive definite matrix can be regarded as a kernel matrix that is as an inner product matrix in some space. Every semi positive definite, symmetric function is a kernel (i.e.) there exists a mapping between the input (x) and the output points (y) .

A radial basis function network kernel is an artificial neural network kernel type that uses radial basis functions as activation functions. It is a linear combination of radial basis functions. They are used in function approximation, time series prediction, and control. They are considered best suited kernel function for the signal classification. Sums of radial basis function are typically used to approximate given functions. This process can also be interpreted as a kind of neural network with more simplicity. Using radial basis function in this manner yields a reasonable interpolation approach provided that the fitting set has been chosen such that it covers the entire range systematically, equidistant data points are ideal. Without a polynomial term that is orthogonal to the radial basis functions, the estimates outside the fitting set don not tend to perform properly.

The ELM kernel is described with SVM that maps the features to higher dimensional space and then uses an optimal hyperplane in the mapped space. This implies that though the original features extracted have the adequate information for good classification, the mapping is done to a higher dimensional feature space could potentially provide better discriminatory clues that are not present in the original feature space. The projection from a low-dimensional feature space into a high-dimensional feature space is achieved by kernels. In practice, SVMs are often used jointly with kernels.

They can be computed by any valid kernel. The SVMs and kernels used together open the way for using SVMs in a wide range of domains: there exist kernels to deal with real numbers, sequences, graphs etc. Given two data points x and z and an ELM with p neurons defining a mapping from R^d to R^p , the corresponding ELM kernel function is defined in equation (1).

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

$$k(x, z) = \frac{1}{p} \phi(x) \cdot \phi(z) \quad (1)$$

where, $k(x, z)$ = kernel function

p = feature space dimension

x, z = input vector values

B. Method Description

The Multilayer Perceptron Neural Network (MLPNN) [9] is the most commonly used neural-network architectures since they have features such as the ability to learn and recognize, smaller training-set requirements better operation, and ease of implementation. One major property of these networks is their ability to find nonlinear surfaces separating the underlying patterns, which is generally considered as an improvement on conventional methods. The MLPNN is a nonparametric technique for performing a wide variety of detection and estimation tasks.

The Radial Basis Function Neural Network (RBFNN) system consists of three layers namely input, hidden, output. The activation of a hidden neuron is determined in two steps: The first is computing the distance usually by using the Euclidean norm between the input vector and a centre c^i that represents the i^{th} hidden neuron. Second, a function h that is usually bell-shaped is applied, using the obtained distance to get the final activation of the hidden neuron. An RBFNN with one output neuron implements the input-output relation which is indeed an assemble of the nonlinear mapping realized by the hidden layer and the linear mapping realized by the output layer.

In this case the Gaussian function $G(x) = \exp\left(-\frac{x^2}{\sigma^2}\right)$ is used. The parameter σ is called unit width and is determined using the heuristic rule global first nearest neighbor. The activation of a neuron in the output layer is determined by a linear combination of the fixed nonlinear basis functions given by equation (2),

$$F(x) = \sum_{i=1}^M \omega_i \phi(x) \quad (2)$$

where, $\phi(x) = G(\|x - c_i\|)$ and ω_i are the adjustable weights that link the output nodes with the appropriate hidden neurons.

There are no weights on the lines from the input nodes to the hidden nodes. The input vector is fed to each m^{th} hidden node where it is put through that nodes radial basis function. The drawbacks are:

1. Robustness.
2. Low accuracy level.
3. Output will not be constant.

IV. HSVM CLASSIFICATION METHODOLOGY

A. PROJECT DESCRIPTION

The classification of EEG signal that is extracted using the popular features namely min, max, mean and standard deviation is performed. The classification is performed using most popular classifier named Support Vector Machines that uses the hierarchical approach.

The raw EEG signal is given as input to the feature extraction phase and the wavelet decomposition is performed and the signal is divided into five sub bands with the wavelet coefficients namely Mean, Standard Deviation and Variance.

The project has been developed with the following phases:

Phase 1: Data Preprocessing from various signals

Phase 2: Feature Extraction and ranking from various signals

Phase 3: Classification of extracted features

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

B. FEATURE EXTRACTION AND RANKING

Transforming the input data into the set of features is called feature extraction. The following are the feature extraction techniques used in extracting the composite features of EEG signals. This technique uses wavelet decomposition to separate the input EEG signal into five sub bands namely Alpha, Beta, Gamma, and Theta with the wavelet coefficients like Mean, Standard Deviation and Variance.

Wavelet Transform (WT) [2] is an effective method of time frequency representation of a signal. The attractive feature of WT is that it provides accurate frequency information at low frequencies and accurate time information at high frequencies. This property is important in bio medical applications; because most signals in this field always contain high frequency components with short duration and low frequency components with long time duration.

Abnormalities in the EEG in serious psychiatric disorders are many times too subtle to be detected using conventional techniques, such as Fourier Transform (FT). Wavelets are specifically appropriate for analysis of non-stationary signals. It is well suited for locating transient events that always occur during epileptic seizure.

Wavelet feature extraction and representation properties can be used to analyze various transient in biological signals [12]. The WT makes use of multi resolution signal analysis technique to decompose EEG signals into a number of frequency bands. The wavelet analysis relies on the introduction of an appropriate basis function and the characterization of signal is reflected through the distribution of amplitude in the basis function.

The features extracted from huge datasets form larger subsets. This will slow down the classification process and also decrease the accuracy. The Recursive Feature Elimination algorithm is used to rank the features and eliminate the low ranking features. RFE is also called as the Backward Propagation algorithm. RFE is an iterative process which consists of the following steps.

1. Training of the classifier
2. Computing the ranking criteria for the features
3. Eliminating the low ranking features

The weights of the features are calculated using the equation (3).

$$\omega = \sum \alpha_i x_i y_i \quad (3)$$

where α_i = lagrange's multiplier

x_i = input vectors

y_i = class labels

The ranking criteria is found by equation (4)

$$c = (\omega_i)^2 \quad (4)$$

The feature with smallest ranking criteria is found using the equation (5)

$$f = \text{argmin}(c) \quad (5)$$

The feature with the value $c = f$ gets eliminated. Through this RFE algorithm the huge subset of the features get reduced involving small classes of features. The classifier now uses the testing dataset and classification is done with the remaining features with the highest ranking. The dataset size gets reduced using this algorithm and therefore high accuracy can be achieved. The iteration process is continued till the minimum weight vector set becomes empty.

C. CLASSIFICATION

EEG Signal classification is the categorization of Signal for its most effective and efficient use. The feature extracted EEG signal is classified using the Multiclass Support Vector Machines classifier [5] extended from the binary classifier. The improved method has the advantage of less time consumption and usage of less system resources, increasing the efficiency of the classification process.

In Figure 1 the steps of a new classifier named Hierarchical Support Vector Machines (HSVM) extended from the binary classifier for various signal classification is shown.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

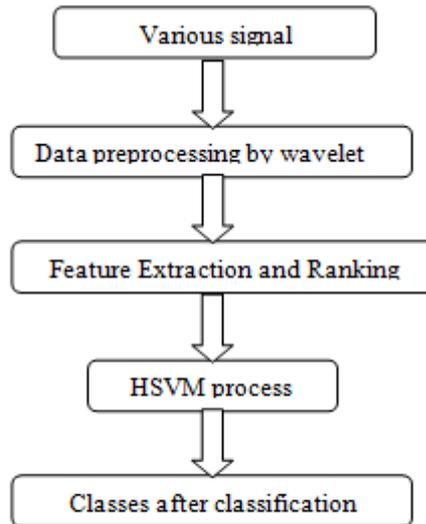


Fig 1.HSVM Process

The hierarchical classification scheme performs better than the multiclass classification problem as it uses less system resources especially system time. The system is because complex multiclass problem is divided into several binary classes problem. In general n class problem is segmented as n-1 binary classes. In this case the five class problem is converted into four binary classes. The inter-related classes are combined to form a group. As there exists many inter dependencies among the classes they are combined to form groups. The binary class SVMs is trained to classify those groups. The trained classes are tested for their accuracy.

The classification involves training and testing. In training phase the data are given to the SVM along with their labels [14]. The labels indicate to that class or group it belongs to. Then in testing phase the data are tested for that class it comes under. Since the binary classifiers are used for classification, only small amount of system resources are utilized. HSVM gives higher accuracy rate and consumes less time.

V. DATASET DESCRIPTION

Here we have used the dataset (digitized various signals) for both healthy and epileptic subjects made Table 1. Five classes of dataset

Class	Description
A	EEG recordings of five healthy volunteers, with eyes open
B	EEG recordings of five healthy volunteers, with eyes closed
C	Seizure-free intervals, from the hippocampal formation of the opposite hemisphere of the brain
D	Seizure-free intervals, from five patients in the epileptogenic zone
E	Seizure activity, selected from all the sites with ictal activity

available online by Dr. Ralph Andrzejak of the Epilepsy Center at the University of Bonn, German (<http://www.meb.unibonn.de/epileptologie/physik/eegdata.html>). The dataset includes five subsets (denoted as A, B, C, D, and E) each containing 100 single-channel EEG segments of 23.6 s duration. Table 1 lists each class and its

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

description. Sets A and B have been acquired from surface EEG recordings of five healthy volunteers, with eyes open and closed, respectively. Signals in subsets C and D have been measured in seizure-free intervals, from five patients in the epileptogenic zone (D) and from the hippocampal formation of the opposite hemisphere of the brain (C). The zone E contains seizure activity, selected from all recording sites exhibiting ictal activity. We used the data described in, which is publicly available. The results are shown below for the proposed method used.

VI. RESULTS AND DISCUSSION

The result obtained by using the proposed classification technique called HSVM with RFE gives higher accuracy when compared with other classification techniques. The comparison is made on the available benchmark datasets as in Table 1. The detail wavelet coefficients at the first-decomposition level of the five types of EEG signals are different from each other and, therefore, they can serve as useful parameters in discriminating the EEG signals.

The combination of the min, max, mean, standard deviation features have the higher accuracy when compared to the individual features. Hence the combination can be taken as the feature set for training the support vector machine and also for classification.

The Table 2 shows the classification rate of the proposed classification technique over various kernels such as linear, Polynomial, Radial Basis Function and Enhanced Radial Basis Function. The classification rate is taken based on the number of true positives which denote the number of seizures detected when compared to the classification done by the experts.

The classification rate achieved is considered to be highest when compared to other kernel methods used.

Table 2. Classification Rate for Various Kernels

KERNEL	CLASSIFICATION RATE
--------	---------------------

The specificity can be represented in the form of graph using the true positive and false positive rates for the dataset. The graph is given in figure 2.

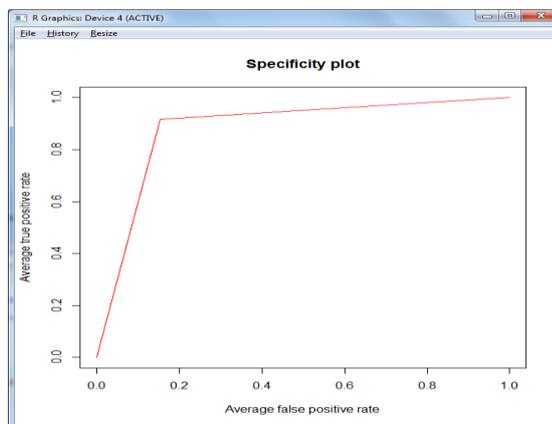


Fig 2. Specificity plot

The error rate is also decreased with the decrease in the feature set. The features are ranked and the number of features is reduced optimizing the bias parameter for the radial basis kernel. The optimized parameter in turn reduces the error rate. The graph representing the error rate is given in figure 3.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

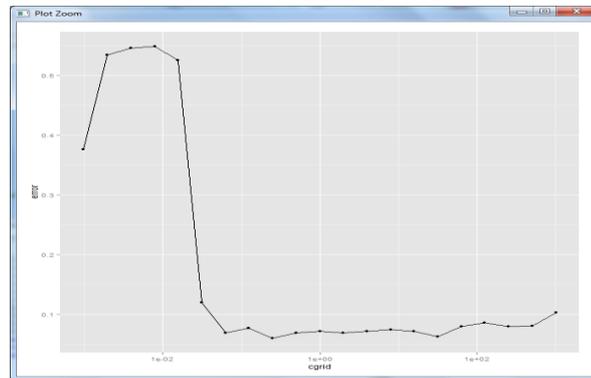


Fig 3. Plot for error rate

VII. CONCLUSION

The improved approach has successfully classified complete range of encephalogram data sets multi-classes A to E with the emphasis on epileptic seizure detection. Compared with the other classification schemes, our method is efficient in terms of classification accuracy and computation complexity. The hierarchical tree structure generated by our approach indicates the interclass relationships among different classes and dataset. The project explains an approach merging both the support vector machine and Extreme Learning Machine framework. Using a kernel defined using the first layer, experiments show that the accuracy of classifiers compares to the accuracy obtained with standard Radial Basis Function kernels. It appears that the only parameter of this. The results from the work will be expanded to include a more complete range of pathologies. Possible directions for further work include optimizing the features and kernel parameters using Particle Swarm Optimization and developing real time epileptic seizure detection and monitoring system.

The future scope of the work is to achieve still better accuracy by developing a custom, user-defined kernel. Also the work will be extended to realtime implementation of a non clinical environment sensor of the given dataset.

REFERENCES

1. Abe S. (2003), 'Analysis of multiclass support vector machines', In Proceedings of International Conference on Computational Intelligence for Modelling Control and Automation (CIMCA'2003), pp. 385-396.
2. Adeli H., Zhou H., and Dadmehr H. (2001), 'Analysis of EEG records in an epileptic patient using wavelet transform', J. Neurosci. Methods, vol.123, no. 1, pp. 69-87.
3. Andrzejak R.G., K. Lehnertz., F. Mormann, C. Rieke P., David C.E., Elger, Indications of nonlinear deterministic and finite-dimensional structures in timeseries of brain electrical activity: dependence on recording region and brainstate, Physical Review E 64 (6) (2001), Article ID 061907.
4. Altunay S., Telatar Z., Eroglu O. (2010), 'Epileptic EEG detection using the linear prediction error energy, Expert Systems with Applications', pp.5661-5665.
5. Boashash B., Mesbah M., P. Colditz, Time frequency detection of EEG abnormalities, in Time-Frequency Signal Analysis and Processing: A Comprehensive Reference, Elsevier, Oxford, UK, 2003, pp. 663-670, ch.15, article 15.5.
6. Critchley H. D. (2002), 'Electro dermal Responses: What Happens in the Brain', in The Neuroscientist.
7. Dogan U, Glasmachers T., Igel C. (2011), 'Fast Training of Multi-class Support Vector Machines', Technical Report no.03/2011, Faculty of science, university of Copenhagen.
8. Ferreira A. J. M. (2009), 'MATLAB Codes for Finite Element Analysis', Springer. ISBN 978-1-4020-9199-5.
9. Gardner A. B., Krieger A. M., Vachtsevanos G., and Litt B. (2006), 'One-class novelty detection for seizure analysis from intracranial EEG', J. Mach. Learn. Res., vol. 7, pp. 1025-1044.
10. Gotman J. and Gloor P. (1976), 'Automatic recognition and quantification of the interictal epileptic activity present in the human scalp EEG', Electroencephalogr. Clin. Neurophysiol., vol. 41, pp. 513-529.
11. Guo L., Rivero D., Pazos A. (2010), 'Epileptic seizure detection using multiwavelet transform based approximate entropy and artificial neural networks', Journal of Neuroscience Methods pp.156-163.
12. Liu Y., Wang R. and Zeng Y. (2007), 'An Improvement of One-against-all Method for Multiclass Support Vector Machine', 4th International Conference: Sciences of Electronic, Technologies of Information and telecommunications, March 25-29, TUNISIA.
13. Majumdar K. and Vardhan P. (2011), 'Automatic seizure detection in ECoG by differential operator and windowed variance', IEEE Trans. Neural Syst. Rehabil. Eng., vol. 19, no. 4, pp. 356-365.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

14. Martis R.J., U.R. Acharya, J.H. Tan, A. Petznick, R. Yanti, C.K. Chua, E.Y. Ng, L. Tong, Application of empirical mode decomposition (EMD) for automated detection of epilepsy using EEG signals, *International Journal of Neural Systems* 22(December (6)) (2012).
15. Polat K., Günes M., , Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform, *Applied Mathematics and Computation* 187 (April (2)) (2007) 1017–1026.
16. Sanei S. and Chambers J. A. (2007), 'EEG Signal Processing', John Wiley & Sons Ltd, New York.
17. Santaniello S., Burns S. P., Singer J. M., Anderson W., and Sarma S. V. (2011), 'Quickest detection of drug-resistant seizures: An optimal control approach', *Epilepsy Behav.*, vol.22, pp. 49–60.
18. Tang Y. and Durand D. M. (2012), 'A tunable support vector machine assembly classifier for epileptic seizure detection', *Expert Syst. Appl.*, vol. 39, pp. 3925–3938.
19. Vapnik V. N. (1999), 'An overview of statistical learning theory', *IEEE Trans. Neural Networks*, vol. 10, no. 5, pp. 988–999.
20. Zavar M., Rahati S., Akbarzadeh T M. R., and Ghasemifard H. (2011), 'Evolutionary model selection in a wavelet-based support vector machine for automated seizure detection', *Expert Syst. Appl.*, vol. 38, pp. 10751–10758.