A Review on Classification of EEG Signals of Different States

Rahul Ingle, Digambar Puri,Vikram Kehri, R. N. Awale
Department of Electrical Engineering, Veermata Jijabai Technological Institute, Mumbai, India

Abstract— Artificial intelligence techniques are being used effectively in medical diagnostic support tools to increase the diagnostic accuracy and provide some additional knowledge that can help in diagnosis. Classification of electroencephalogram (EEG) signals is an important basis for a brain-computer interface (BCI) systems. In this paper an effort is made to review some of these techniques which are applied on standard five EEG datasets (A-E) and give a brief comparison for the same in terms of classification accuracy of EEG data. Feature extraction techniques and neural network is used to classify the data. Wavelet Transform (WT), Principle Component Analysis (PCA), Independent Component Analysis (ICA), etc. are feature extraction techniques. Literature presents different techniques to classify data such as probabilistic neural network (PNN), Support Vector Machine (SVM), Artificial Neural Networks (ANN), etc. Combination of WT and SVM improved the classification accuracy than other combinations such as DWT with ANN, ICA with MLPN, PCA with ANN and DWT with PNN.

Index Terms-- Electro-encephalogram (EEG), Wavelet Transform (WT), Principle Component Analysis (PCA), Support Vector Machine (SVM), Multilayer Perceptron Neural Network (MLPNN), Probabilistic Neural Network (PNN).

I. INTRODUCTION

The Electro-encephalogram (EEG) is highly complex signal, widely used clinically to investigate brain disorders [1]. EEG signals are highly non-linear, aperiodic, time varying responses characterized with small amplitude and very low frequency [4]. EEG signals involve a great deal of information about the function of the brain. It is a record of electrical potentials generated by cerebral cortex nerve cells. The changes in the voltage difference between electrodes are sensed and amplified before being transmitted to a computer program to display the tracing of the voltage potential recordings [5], [16].

In order to extract relevant information from recordings of brain electrical activity, a variety of computerized-analysis methods have been developed. In early days Fast Fourier Transform (FFT) was used for analysis of EEG signals, but FFT suffers from large noise sensitivity. Parametric power spectrum methods such as autoregressive (AR), reduces the spectral loss problems and gives better frequency resolution [6]. Since EEG signals are non-stationary, the parametric methods are not suitable for frequency decomposition of these signals [18]. Another method is Short Time Fourier Transform (STFT) which provides resolution in short window of time for all frequencies. FFT, AR, STFT do not have time and frequency resolution at same time [3], [7].

To extract the features of EEG signals and to overcome the problems of STFT, AR, FFT, a power-full tool that is Wavelet Transform can be applied to extract the wavelet coefficients of discrete time signals. This procedure makes use of multi-rate signal-processing techniques [1], [7].

Artificial Neural Networks have been used in great number of medical diagnostic decision support system applications because of the belief that these have great predictive power. Many authors have shown that combining the prediction of several models often results in a prediction accuracy that is higher than that of individual models [8]. In this paper, we discuss various methods proposed for classification of same EEG datasets.

Rest of the paper is organized as follows: Section II describes EEG Data Acquisition. Section III includes, various feature Extraction Techniques proposed by different peoples for same EEG dataset. Section IV is about Classification Methods. Section V is Comparison of results, last section concludes this work.

II. EEG DATA ACQUISITION

This paper is review of all papers that have used the data taken from which is publicly available [4]. The EEG dataset consists of five sets (denoted A-E), with each set containing 100 single-channels EEG signals of 23.6 S [4]. Each signal has been selected after visual inspection for artifacts and has passed a weak stationary criterion [9]. The sampling rate of the data is 173.61s. The time series have the spectral bandwidth of the acquisition system, which is 0.5 Hz to 85 Hz. The application of a low-pass filter of 40 Hz is regarded as the first step of analysis. Sets A and B have been taken from surface EEG recordings of five healthy volunteers with eyes open and closed, respectively, the other two sets 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), set E containing seizure activity, and it is selected from all recording sites exhibiting ictal activity. Datasets A and B are recorded extra cranially, whereas sets C, D, and E are recorded intracranially. Apart from the different recording electrodes, the recording parameters were fixed.

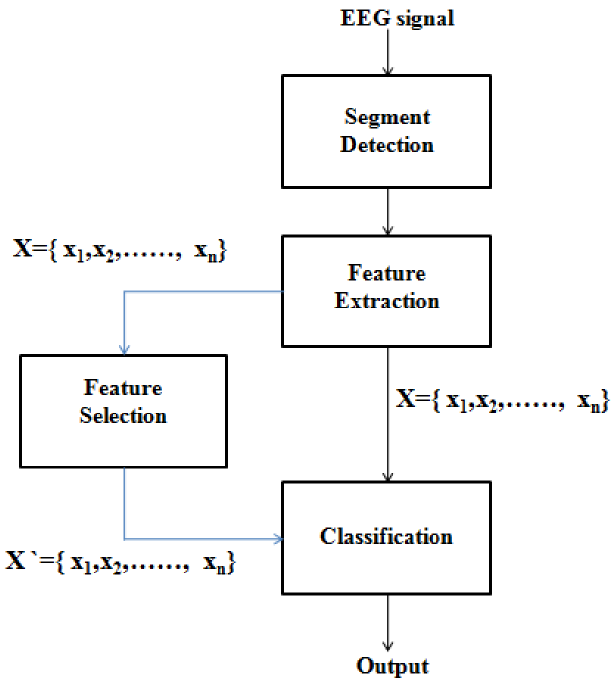

Fig.1. Functional modules in a typical computerized EEG system.

III. FEATURE EXTRACTION TECHNIQUES

There are several feature extraction techniques in the literature, some of them which have been used in the papers as follows:

A. *Wavelet Transform:*

The Wavelet Transform (WT) is designed to address the problem of non stationary signals. It involves representing a time function in terms of simple, fixed building blocks, termed Wavelets [16]. These building blocks are actually a family of functions which are derivedfrom a single generating function called the mother wavelet by translation and dilation operations. The main advantage of WT is that it has a varying window size, being broad at low frequencies and narrow at high frequencies, thus leading to an optimal time frequency resolution in all frequency ranges [14]. One-dimensional WT is defined as,

$$W_f = \frac{1}{\sqrt{d}}\int_{-\infty}^{\infty} f(t)\, \Psi^*\left(\frac{t-m}{d}\right) dt \qquad (1)$$

Where $\psi^*(t)$ is the conjugate function of mother wavelet $\psi(t)$ and d, m are called scale parameter and shift parameter respectively. Major advantage of WT for EEG signals is, itprovides the better time frequency resolution at same time andcompress data by preserving original information.

B. *Principle Component Analysis:*

Principal component analysis (PCA) has been called oneof the most valuable results from applied linear algebra. PCA is used abundantly in all forms of analysis from neuroscience to computer graphics - because it is a simple, non-parametric method of extracting relevant information from confusing data sets. PCA used for for finding similarity in data pattern. For any data set Z = [z1, z2, z3…zn] (n- dimension of data-point). Suppose Z includes m data-points, the new set of data-points is given by [15],

$$Y_i = z_i - \mu \qquad i = 1,2,3, \ldots\ldots m \qquad (2)$$

$\mu$is a mean vector. We have to find the covariance matrix of Yi can be expressed as,

$$C_{n\,x\,n} = cov(z_i, z_j)\,;\; i = 1,2,3 \ldots\ldots n;\; j = 1,2,3 \ldots\ldots n. \,(3)$$

$$C = \sum_{j=1}^{m} (z_i - \mu)\,(z_i - \mu)^T \qquad (4)$$

m is total number of data-points. The eigen value λ_i and eigen

Vectors ψ_i of covariance matrix C satisfy,

$$C_{\psi_i} = \lambda_i \psi_i \qquad (5)$$

$$E = [\psi_1\,, \psi_2, \psi_3, \ldots\ldots, \psi_m] \qquad (6)$$

Project the data set Y into the eigen vector space, we get:

$$P^{m\,x\,n} = E^T Y^T \qquad (7)$$

Where, Y = [Y_1, Y_2…. Yn]; We can easily mapped P back into the original coordinates, where

$$Y^T = EP \qquad (8)$$

E is a orthogonal matrix, and $E^{-1} = E^T$

$$Y^T = E'P' \qquad (9)$$

If λ_i is arranged from the largest to the smallest and $\lambda_i$is arranged according to the value λ_i, we select the first k row of matrix P to form a matrix P’, which represents the principle component of the data set X. P’ can construct m data points with k dimension. We can mapped the principle component back into the original coordinates by,

C. *Independent Component Analysis:*

ICA is a signal processing technique in which observed random data is transformed into components that are statistically independent from each other. Linear ICA was used to separate neural activity from muscle and blind artifacts in spontaneous EEG data. It was verified that the ICA can separate artifactual, stimulus locked, response-locked, and nonevent related background EEG activities into separate components [5]. Furthermore, ICA would appear to be able to separate task-related potentials from other neural and artifactual EEG sources during hand movement imagination in form of independent components. Power spectra of the linear ICA transformations provided feature subsets with higher classification accuracy than the power spectra of the original EEG signals.

IV. CLASSIFICATION METHODS

Artificial neural network techniques are the recent trendsfor classification of 1-D and 2-D signals. In literature classification of EEG signals done by using various techniques such as feed forward network at different layers (FFNN), back propagation neural network (BPNN), principle component analysis (PCA), Support vector machine (SVM), Logistic Regression (LR), multilayer perceptron neural network (MLPNN), probabilistic neural network (PNN), etc.

A. Support Vector Machine (SVM):

SVM is a supervised learning machine based on Statistical theory. It trains a classifier by finding an optimal separating hyperplane which maximizes the margin between two classes of data in the kernel induced feature space. A training sample set $\{x_i, y_i\}$; i =1-N is considered, where N is total number of samples [2]. The hyperplane f(x) = 0 that separates the given data can be obtained as a solution to the following optimization problem,

minimize:

$$\frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{N}\xi_i \tag{10}$$

Subject to,

$$\begin{cases} y_i(\omega^T x_i + b) \geq 1 - \xi_i \\ \xi \geq 0,\ i = 1,2,3,\dots,n \end{cases} \tag{11}$$

Where C is a constant representing error penalty. Rewriting the above optimization problem in terms of Lagrange multiplier(λ), leads to the following problem [3], maximize,

$$W(\lambda) = \sum_{i=1}^{N}\lambda_i - \frac{1}{2}\sum_{i,j=1}^{N} y_i y_j\, x_i x_j \lambda_i \lambda_j \tag{12}$$

Subjectto,

$$0 \leq \lambda_i \leq C \tag{13}$$

$$\sum_{i=1}^{N}\lambda_i y_i = 0, \qquad i = 1,2,3,\dots\dots,N \tag{14}$$

Holding hyperplane vectors are termed as support vectors. In literature SVM has been used to classify multiple datasets of EEG signals.

B. MLPNN:

The MLPNNs are the most commonly used neural-network architectures since they have features such as the ability to learn and generalize, smaller training-set requirements, fast operation, and ease of implementation [17]. 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 [1]. The MLPNN is a nonparametric technique for performing a wide variety of detection and estimation tasks [12].

C. Probabilistic Neural Network (PNN):

A single PNN is capable of handling multiclass problems. This is opposite to the so-called one-against-the rest or one-per-class approach taken by some classifiers, such as the SVM, which decompose a multiclass classification problem into dichotomies and each chotomizer, has to separate a single class from all others [3]. The PNN architecture is composed of many interconnected processing units or neurons organized in Successive layers. The input layer unit does not perform any computation and simply distributes the input to the neurons in the pattern layer. On receiving a pattern x from the input layer, the neuron xij of the pattern layer computes its output is given by [10],[17],

$$\phi_{ij} = \frac{1}{(2\pi)^{\frac{d}{2}}\sigma^d} exp\left[\frac{-(x-x_{ij})^T(x-x_{ij})}{2\sigma^2}\right] \tag{15}$$

Where d denotes the dimension of the pattern vector x, is the smoothing parameter, and xij is the neuron vector [11], [13].

V. COMPARATIVE DISCUSSION AND FUTURE WORK

In literature, [6] has done two types of experiments, In the first experiment the raw datacontaining the 4096 samples for each series is used directly for feature extraction using discrete wavelet transform. In the second experiment a rectangular window which is formed by256 discrete data was selected. After down-sampling, for the 100 series set, a total of 1600 vectors is obtained from each set. The obtained data is used for feature extraction usingPCA [6]. PCA+ANN gives 95.2% where as DWT+ANN gives 90.4% [6].

In [1] shown that the different features are extracted using DWT+ICA, classified using SVM, PNN, MLPNN [1]. The total-classification accuracies of SVM, PNN, and MLPNN obtained in the first experiment (training and testing sets consisted of raw data) were 98.23%, 98.05%, and 93.63%, respectively, as shown in table I.

The high classification accuracies of the multiclass SVM and PNN classifiers give insights into the features used for defining the EEG signals. The conclusions drawn in the applications demonstrated that the wavelet coefficients are the features, which represent the EEG signals, and by the usage of these features a good distinction between classes can be obtained. The applications of the SVM to the EEG signals presented in the literature consisted of two-class EEG-signals classification problem [7], [8]. In literature [9] used decision-directed acyclic graphs (DDAGs) for application of the SVMs to the multiclass EEG (spontaneous EEG during five mental tasks) signals classification problem. A given DDAG is evaluated much like a binary decision tree, where each internal node implements a decision between two of the k classes of the classification problem. The multiclass SVM and error-correcting output codes (ECOC) algorithm used in our study to classify the EEG signals indicated higher performance than that of the multiclass SVM and the DDAG algorithm presented by [9]. Multiclass SVM trained on composite feature vectors obtains the highest accuracies as compared to the other classifiers. The accuracies obtained by the PNN are slightly lower than the accuracies of the multiclass SVM. The performance of these two classifiers is then compared with that of the MLPNN. Thelowest accuracies are obtained by the MLPNN.

CLASSIFICATION TECHNIQUES	ACCURACY (%)
DWT + ANN	90.4
ICA + MLPN	93.63
PCA + ANN	95.4
DWT + PNN	98.05
DWT + SVM	99.28

V. CONCLUSION

Classification of EEG signals is a challenging task for researchers. The paper described few feature extraction and classification techniques which have been used for classification of five datasets of EEG signals. WT reduced data and provided better features as compared with PCA and ICA.

The multiclass SVM showed great performance since it maps the features to a higher dimensional space. Besides this, the PNN provided encouraging results, which could have originated from the architecture of the PNN. The performance of the MLPNN was not as high as the multiclass SVM and PNN. This may be attributed to several factors including the training algorithms, estimation of the network parameters, and the scattered and mixed nature of the features. Study of the present paper demonstrated that the multiclass SVM and PNN can be used in the classification of the EEG signals by taking into consideration the misclassification rates. Performance of classifier further improved by using combination of different machine learning and artificial neural techniques with number of hidden layers.

REFERENCES

[1] Fatma Guler, N.; Ubeyli, E.D., "Multiclass Support Vector Machines for EEG-Signals Classification," *Information Technology in Biomedicine, IEEE Trans.*, vol.11, no.2, pp.117-126, March 2007.

[2] Minfen Shen; Lanxin Lin; Jialiang Chen; Chang, C.Q., "A Prediction Approach for Multichannel EEG Signals Modeling Using Local Wavelet SVM," *Instrumentation and Measurement, IEEE Trans.*, vol.59, no.5, pp.1485-1492, May 2010.

[3] Ismail Babaoglu, Og˘uz Fındık, Mehmet Bayrak, "Effects of principle component analysis on assessment of coronary artery diseases using support vector machine," *Expert Systems with Applications*, vol.37, pp.2182–2185, 2010.

[4] Ralph G. Andrzejak, Klaus Lehnertz,Florian Mormann, Christoph Rieke, Peter David, Christian E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity dependence on recording region and brain state", Physical Review E, vol. 64, 2001.

[5] T. Jung, S. makeig, M. Wasterfield, J. Townsend, "Analysis and visualisation of single-trial event-related potential",Human Brain Mapping, vol.no. 14,166-185,2001.

[6] S.S.Lekshmi; V.Selvam;M.Pallikonda Rajasekaran, "EEG signal Classification using Principle Component Analysis and Wavelet Transform with Neural Network" International conference on communication and signal processing, IEEE conference, vol.1, pp.456-459, April 3-5, 2014.

[7] F. Miwakeichi, R. Ramirez-Padron, P. A. Valdes-Sosa, and T. Ozaki, “A comparison of non-linear non-parametric models for epilepsy data,” Comput.Biol. Med., vol. 31, no. 1, pp. 41–57, 2001.

[8] N. Acir, I. O¨ ztura, M. Kuntalp, B. Baklan, and C. Gu¨zelis¸, “Automaticdetection of epileptiform events in EEG by a three-stage procedure basedon artificial neural networks,” *IEEE Trans.* Biomed. Eng., vol. 52, no. 1,pp. 30–40, Jan. 2005.

[9] D. Garrett, D. A. Peterson, C. W. Anderson, and M. H. Thaut, “Comparisonof linear, nonlinear, and feature selection methods for EEG signal classification,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2,pp. 141–144, Jun. 2003.

[10] D. F. Specht, “Probabilistic neural networks, “*Neural Netw*., vol. 3, no. 1,pp. 109–118, 1990.

[11] P. Burrascano, “Learning vector quantization for the probabilistic neural network,”*IEEE Trans. Neural Netw*., vol. 2, no. 4, pp. 458–461, Jul. 1991.

[12] M. T. Hagan and M. B. Menhaj, “Training feedforward networks with the Marquardt algorithm,” *IEEE Trans. Neural Netw*., vol. 5, no. 6, pp. 989–993, Nov. 1994.

[13] Kavita Mahajan, M. R. Vargantwar, Sangita M. Rajput, "Classification of EEG using PCA, ICA and Neural Network",

International Journal of Engineering and Advanced Technology (lJEAT)ISSN: 2249 - 8958, vol.l, oct .2011.

[14] S.S.Lekshmi; V.Selvam;M.Pallikonda Rajasekaran, "EEG signal Classification using Principle Component Analysis and Wavelet Transform with Neural Network"*International conference on communication and signal processing, IEEE conference*, vol.1, pp.456-459, April 3-5, 2014.

[15] Li Ke; Rui Li, "Classification of EEG signals by multi-scale filtering and PCA,"*International Conference on Intelligent Computing and Intelligent Systems, ICIS. IEEE conference*, vol.1, pp.362,366, 20-22 Nov., 2009.

[16] Elif Derya Übeyli, "Combined neural network model employing wavelet coefficients for EE signals classification" *Digital Signal Processing*, vol.19, pp. 297–308.

[17] Kashtiban, A.M.; Razmi, H.; Kozehkonan, M.K., "Combined LVQ neural network and multivariate statistical method employing wavelet coefficient for EEG signal classification," *Mechatronics (ICM), IEEE International Conference on* , vol., no., pp.809,814, 13-15 April 2011.

[18] Abdulhamid Subasi, Ergun Ercelebi, "Classification of EEG signals using neural network and logistic regression", *Computer methods and Program in Biomedicine*, Vol.78, pp.87-99, 2005.