

Diseases Detection from Electroencephalogram Signals Using Support Vector Machine

Md Masud Rana¹, Md Walid Hasan² and Ahmed Abdelhadi³

¹Department of Electrical Engineering

The University of Texas At Arlington, Texas, USA

²Department of Electronics and Communication Engineering

Khulna University of Engineering & Technology, Bangladesh

³Department of Engineering Technology

University of Houston, Texas, USA

Email: mmamaraece28@yahoo.com

Abstract—The need and importance of automatic disease recognition such as electroencephalogram has grown over time. Driven by this motivation, this research demonstrates the electroencephalogram (EEG) signals reconstruction process using the finite impulse response, principle component analysis, feature extraction, and support vector machine (SVM). After mentioning related literature, the EEG signals are taken from biomedical database such as Temple University Hospital, Australia. Applying finite impulse filter to the noisy EEG signals, the motion artifacts have been effectively removed. Generally, EEG signal is a multidimensional so it is quite difficult to find out effective channel for different diseases. Applying principle component analysis over filtered EEG signals, dimensional reduced EEG signals are obtained. For classifying EEG signals, different statistical measured such as standard deviation and mean absolute deviation are applied. Moreover, the SVM is used to classify the EEG signal from the selected features. Finally, the system performance is evaluated by 27 patients EEG database. For each disease, it has taken 9 signals. For different signals, the SVM are trained and evaluate the performance. Simulation results show that the SVM provides better performance for higher number of signals.

Index Terms—EEG signals, filter, principle component analysis, and support vector machine.

I. INTRODUCTION

Stroke is one of the main causes of death all over the world [1]. There are two types of strokes named ischemic stroke and hemorrhagic stroke. The majority type of stroke is the ischemic which is caused by blockage of the small blood vessels around the brain [2]. A hemorrhagic stroke happens when an artery in the brain leaks blood or ruptures. This blood puts too much pressure on brain cells, which damages them. High blood pressure and aneurysms balloon-like bulges in an artery that can stretch and burst cause a hemorrhagic stroke [3]. According to the report published by American Stroke Association in 2015, the ischemic stroke is 87% of all strokes cases in the United States of America in 2015 [3].

Generally speaking, the computerized tomography (CT) scanner and MRI are used to accurately diagnosis stroke in Bangladesh. The use of CT-scan and MRI procedure still quite risky because of patient is exposed by X-rays [4]. Especially

for developing countries like Bangladesh, the CT-scan and MRI are not still available as expected. Unfortunately, the cost of CT-scan and MRI are very expensive. Generally, CT scan or MRI may take several hours to obtain examination reports, while the initial aid is quite limited period for stroke patients [5]. Another import potential tool to be used as an alternative is the electroencephalography (EEG) which is not available almost all clinics in Bangladesh. Driven by this motivation, this paper proposes a support vector machine based EEG signal deionising approach for developing smart automatic disease recognition systems.

A. Literature Review:

There are many algorithms for removing noise from the EEG signals are discussed in the literature. Among them, the extreme learning machine, support vector machine, K-nearest neighbour, Bayesian, and Parzen window classifiers are the main schemes used for deionising the EEG signals. In summary, Table I shows the details advantages and disadvantages of these methods. Moreover, the least mean squared and Kalman filter for EEG signal reconstruction is proposed in [13], [14]. Unfortunately, the computational complexity of these iterative approaches are very high [15]. The state-of-the-art for motion capture systems with different approaches are described in [15]. It can be seen that there is a little effort for developing noise reduction algorithm from machine learning point of view.

B. Main Contributions:

This paper investigates EEG signals reconstruction process using the finite impulse response, principle component analysis, feature extraction, and SVM classifier. The main contributions of this paper are as follows:

- Applying principle component analysis over filtered EEG signals, dimensional reduced EEG signals are obtained. Afterwards, the proposed SVM is applied to classify EEG signals from the selected features.
- The performance of the proposed scheme is evaluated by 27 patients EEG signals from the Temple Univer-

TABLE I: Literature review.

Method	Advantage	Disadvantage
Extreme learning machine [6].	(1) The algorithm can properly detect the normal objects. (2) The algorithm has good generalization ability, quick to perform learning, and a simple structure.	(1) The performance of proposed method is not good to detect ischemic stroke patient. (2) Moment of scaling exponent can not distinguish between normal subject and ischemic stroke patient.
Quantitative EEG [7].	To support the localization of portion of nerve on clinical examination.	The analysis process requires more time/storage, so it is very difficult to implement.
Cutting-edge signal processing techniques [8].	Accurately detection of acute ischemic stroke and sizes.	There were some irregularities of the EEG signals in comparisons with CT-scan.
Neural network [3].	Batch normalization can reduce the number of iterations towards convergence, hence accelerate the training process.	Needs small value of learning rate and much more time to training.
Support vector machine [9].	Effective in high dimensional sample spaces.	If the number of features is much greater than the number of samples, the method is likely to give poor performances.
Pattern recognition approach [10].	This approach provides high classification accuracy.	Difficult to implement.
K-nearest neighbour [11].	Provide minimum prediction error.	High complexity.
Bayesian, K-nearest neighbor, and Parzen window classifiers [12].	Fast, reliable and low computational complexity classification method.	The method is useful on the data sets that have few numbers of features.

sity Hospital database. Simulation results show that the proposed SVM provides better performance considering higher number of signals.

C. Organisation:

The rest of the paper is organized as follows. Section II describes the proposed method which follows the simulation results. Finally, conclusions are made at the end of the paper.

II. PROPOSED ALGORITHM

Fig. 1 shows the generalised EEG signal reconstruction process [16], [14]. It can be seen that there are five steps of

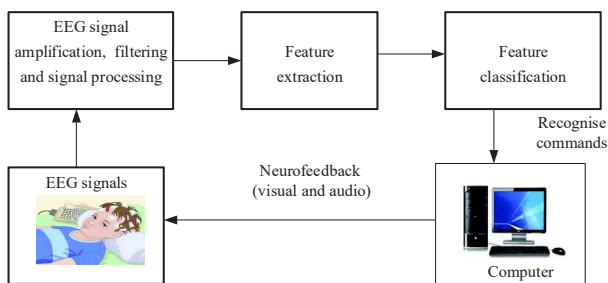


Fig. 1: Typical Neurofeedback systems [16], [14].

this process as follows:

- 1) **Brain activity measurement:** This step consists of various types of sensors in order to obtain signals reflecting to the human brain activity.
- 2) **Pre-processing:** This step consists of amplification and de-noising of input EEG signals in order to enhance the relevant information embedded in the signals.
- 3) **Feature extraction:** Feature extraction aims at describing the signals by a few relevant values called features.
- 4) **Feature classification:** The classification step assigns a set of features extracted from the EEG signals. This class corresponds to the kind of mental state identification.
- 5) **Translation into a command:** In this step, the diseases of subject are identified with their names.

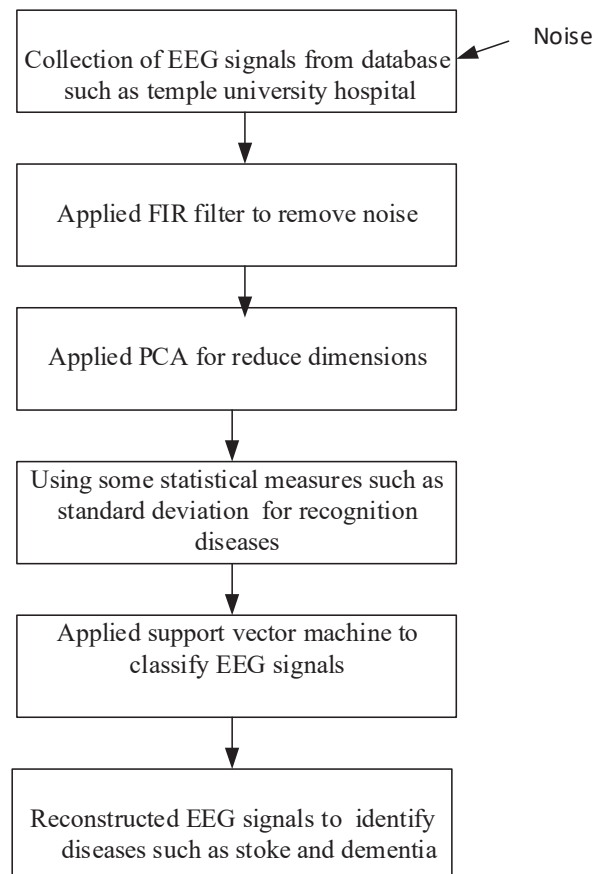


Fig. 2: Complete flow diagram of the proposed method.

The proposed method for diseases detection is mainly based on the features extraction from the noisy EEG signals. The complete flow diagram of the proposed method is described in Fig. 2. First of all, 27 patients EEG signals from the Temple University Hospital, Australia are collected from the noisy

TABLE II: Feature analysis of pre-processed 27 patients EEG signal.

Disease detection	Standard deviation	Mean absolute deviation	Median absolute deviation	Max norm	L_1 norm	L_2 norm	Skewness	Kurtosis
Stroke	169.68	112.08	54.37	985.77	3.38×10^7	93017.33	-2.54	439.95
Normal	161.73	24.29	13.41	6547.15	8.49×10^6	95610.42	35.76	65.82
Dementia	203.33	100.2	62.15	4246.92	2.68×10^7	105113.7	4.82	29.78
Stroke	701.56	350.18	162.77	9289.17	1.02×10^8	386301	2.62	28.64
Normal	88.11	62.97	42	568.08	2.38×10^7	56498.18	-0.66	4.89
Dementia	2827.17	2480.48	2594.29	5366.22	7.85×10^8	1.59×10^6	-0.17	1.79
Stroke	174.06	78.85	31.77	1316.28	2.50×10^7	98001.27	-4.43	25.02
Normal	181.19	44.44	19.54	5242.93	1.60×10^5	110725.1	15.88	320.42
Dementia	307.09	29.99	14.79	12152.22	8.89×10^4	169141.7	25.68	988.14
Stroke	314.08	54.5	35.49	20373.72	1.81×10^5	180982.6	38.07	1747.34
Normal	121.05	70.81	43.3	907.75	2.15×10^7	66633.46	-2.38	15.63
Dementia	163.92	51.76	30.03	4896.2	1.58×10^5	90426.02	13.2	316.19
Stroke	1142.81	623.02	246.95	5214.01	2.19×10^7	721078.3	1.94	10.92
Normal	120.58	72.27	37.29	854.12	2.64×10^7	72872.5	1.42	11.26
Dementia	379.15	98.69	24.02	4999.41	2.11×10^7	187052.8	6.73	92.58
Stroke	231.24	98.13	50.9	4994.15	2.83×10^7	125635.5	7.25	117.73
Normal	210.72	188.17	199.63	1608.44	6.40×10^6	122915.8	0.15	1.63
Dementia	275.11	63.1	19.22	4871.92	2.27×10^5	165181.5	4.56	120.29
Stroke	156.93	28.73	14.4	5000	1.04×10^5	94556.99	13.66	566.96
Normal	239.09	114.8	58.29	4827.32	3.21×10^5	128253.5	-7.67	106.52
Dementia	113.59	69.43	35.67	1138.13	2.04×10^5	61696.54	3.82	24.56
Stroke	145.69	64.44	23.46	1162.74	2.00×10^5	81212.73	4.65	31.49
Normal	110.56	52.44	18.54	679.13	1.55×10^7	60792.22	-3.86	18.35
Dementia	273.88	82.47	40.21	13051.48	3.07×10^7	166217.1	-25.05	1007.29
Stroke	159.8	102.92	57.57	1058.05	4.56×10^7	106357.1	-0.71	7.06
Normal	155.77	114.17	53.46	818.85	3.40×10^7	84960.44	1.89	6.8
Dementia	439.96	65.82	29.79	13469.96	2.05×10^7	246381.1	20.59	0

environments. Using filter, the noise is removed which follows PCA, and SVM steps for diseases classification such as stroke, normal, and dementia [17], [18], [19]. Basically, there are different statistical measurements such as standard deviation are used to train the support vector machine. The detail process is described as follows:

- 1) **Step:** After collecting EEG signals, first step of the diseases detection method is to filter the EEG signals.
- 2) **Step:** Apply PCA to the filtered EEG signal for the purpose of channel reduction and finding out the effective channel from 62 to 1 channel for EEG signals.
- 3) **Step:** After channel reduction, the feature extraction process is applied to extract the previously mention features. Features are the values which define some relevant properties of the considered signals. Feature extraction is an operation which converts one or several signals into a feature vector.
- 4) **Step:** From these feature data, two feature matrices are formed. The first matrix contains the feature of stroke, dementia and normal EEG signals while other contains the feature stroke and dementia of EEG signals. The classification algorithm uses feature extracted from EEG signals to classify the EEG signals or detect diseases.
- 5) **Step:** As the data set has been labeled, so this group matrix corresponding to the feature matrix can easily form. It is required for SVM.
- 6) **Step:** By using previous feature and group matrix, the SVM is trained and tested.

- 7) **Step:** SVM training and testing procedure will be executed in two way. First of all, it is the person dependent approach. Those persons who is achieved best performance on the basis of accuracy will be selected, while second approach is person independent. Given a set of training examples, each marked as belonging to one or the other of two categories. An SVM training algorithm constructs a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier. A SVM model is a representation of the samples as points in space and mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New samples are then mapped into that same space and predicted belong to a category based on which side of the gap they fall.

The performance of the proposed algorithm is described in the following section.

III. SIMULATION RESULTS AND DISCUSSIONS

The brain is the most wonderful part in human body. The electroencephalogram measures the electrical activity of the brain from the scalp. The electrical activity of the brain is picked by the electrodes located on the scalp as shown in Fig. 3. It controls all functions and senses of the body. In order to de-noise the EEG signal, the complete simulation process is illustrated in Fig. 3. The filtering approach is applied to remove noise from the EEG signal which follows the PCA, feature classification, and extraction process. The considered

noise is uniformly distributed random noise which is due to external influences and electrode polarization. The filtered

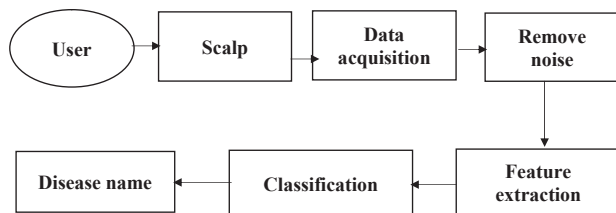


Fig. 3: The simulation setup for diseases identification.

EEG signals are demonstrated in Fig. 4. It can be seen that the designed filter can properly remove the noise. Features are the

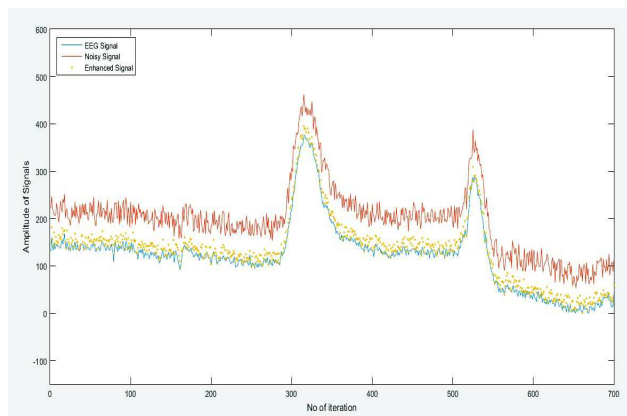


Fig. 4: Temporal filtered signals.

values which define some relevant properties of the acquired signals. Feature extraction is an operation which converts one or several signals into a feature vector. The classification algorithm uses feature extracted from EEG signals to classify or detect EEG diseases. Here, the standard deviation, mean absolute deviation, median absolute deviation, max norm, L1-norm, L2-norm, skewness, kurtosis are extracted from the 27 patients EEG signals for recognition diseases. Table II shows the diseases classification outcome of the proposed approach. It represents that the SVM can properly detect diseases of EEG signals.

Finally, we have taken different number of EEG signals each time for measuring accuracy of the proposed method. Table III demonstrates the performance of the proposed method. From the table it can be concluded that more number of EEG signals provide more detection accuracy. The lowest number of used EEG signals are 3 and it respective accuracy is 51.85 %. The lowest number of EEG signals are 27, and it respective accuracy is 81.48%. In summary, the accuracy of the system is increased with the increasing the number of EEG signals in training session.

IV. CONCLUSION AND FUTURE WORK

This paper presents EEG signals reconstruction process. Basically the proposed SVM algorithm can detect the normality

TABLE III: Performance analysis of the proposed method.

Number of EEGs	Accuracy	Number of EEGs	Accuracy
3	37.04%	6	59.26%
9	59.26%	12	77.78%
15	70.37%	18	77.78%
21	81.48%	24	85.19%
27	88.89%		

and abnormality of the patient EEG signals. After collecting EEG signals from temple university hospital, the principle component analysis and support vector machine are applied for EEG signal reconstruction. Using statistical measured such as standard deviation, mean absolute deviation, median absolute deviation, L_1 -norm, L_2 -norm, skewness and kurtosis, the EEG signals are reconstructed with reasonable accuracy. Its shows that the proposed method can effectively reduce noise from the EEG signals. Therefore, this work is valuable for signal processing and biomedical research communities. It can be seen that current system only classify three diseases, so this research will need to extend by increasing the number of diseases.

REFERENCES

- [1] S. R. Suhas, K. Behbehani, S. Vijendra, J. R. Burk, and E. A. Lucas, "ECG biomarkers for simultaneous detection of obstructive sleep apnea and cheyne-stokes breathing," in *Annual International Conference on Engineering in Medicine and Biology Society*, 2007, pp. 1047–1050.
- [2] K. Sudharani, T. Sarma, and K. S. Prasad, "Brain stroke detection using K-Nearest neighbor and minimum mean distance technique," in *International Conference on Control, Instrumentation, Communication and Computational Technologies*, 2015, pp. 770–776.
- [3] E. P. Giri, M. I. Fanany, A. M. Arymurthy, and S. K. Wijaya, "Ischemic stroke identification based on EEG and EOG using ID convolutional neural network and batch normalization," in *International Conference on Advanced Computer Science and Information Systems*, 2016, pp. 484–491.
- [4] I. Q. Grunwald, K. Fassbender, and A. K. Wakhloo, *How to set up an Acute Stroke Service*. Springer Science and Business Media, 2011.
- [5] R. C. Hwa and T. C. Ferree, "Stroke detection based on the scaling properties of human EEG," *Physica A: Statistical Mechanics and its Applications*, vol. 338, no. 1-2, pp. 246–254, 2004.
- [6] H. Adhi, S. Wijaya, C. Badri, and M. Rezal, "Automatic detection of ischemic stroke based on scaling exponent electroencephalogram using extreme learning machine," in *Journal of Physics: Conference Series*, vol. 820, no. 1, 2017, p. 012005.
- [7] L. Murri, S. Gori, R. Massetani, E. Bonanni, F. Marcella, and S. Milani, "Evaluation of acute ischemic stroke using quantitative EEG: A comparison with conventional EEG and CT scan," *Neurophysiologie Clinique/Clinical Neurophysiology*, vol. 28, no. 3, pp. 249–257, 1998.
- [8] K. G. Jordan, "Emergency EEG and continuous EEG monitoring in acute ischemic stroke," *Journal of Clinical Neurophysiology*, vol. 21, no. 5, pp. 341–352, 2004.
- [9] S. Shalev-Shwartz, Y. Singer, N. Srebro, and A. Cotter, "Pegasos: Primal estimated sub-gradient solver for SVM," *Mathematical programming*, vol. 127, no. 1, pp. 3–30, 2011.
- [10] H. U. Amin, W. Mumtaz, A. R. Subhani, M. N. M. Saad, and A. S. Malik, "Classification of EEG signals based on pattern recognition approach," *Frontiers in computational neuroscience*, vol. 11, p. 103, 2017.
- [11] S. Y. Adam, A. Yousif, and M. B. Bashir, "Classification of ischemic stroke using machine learning algorithms," *International Journal of Computer Applications*, vol. 149, no. 10, 2016.
- [12] N. Panahi, M. G. Shayesteh, S. Mihandoost, and B. Z. Varghahan, "Recognition of different datasets using PCA, LDA, and various classifiers," in *International Conference on Application of Information and Communication Technologies*, 2011, pp. 1–5.

- [13] P. Khanna, N. C. Swann, C. de Hemptinne, S. Miocinovic, A. Miller, P. A. Starr, and J. M. Carmena, "Neurofeedback control in parkinsonian patients using electrocorticography signals accessed wirelessly with a chronic, fully implanted device," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 10, pp. 1715–1724, 2017.
- [14] W. C. Ung, T. Funane, T. Katura, H. Sato, T. B. Tang, A. F. M. Hani, and M. Kiguchi, "Effectiveness evaluation of real-time scalp signal separating algorithm on near-infrared spectroscopy neurofeedback," *IEEE journal of biomedical and health informatics*, vol. 22, no. 4, pp. 1148–1156, 2018.
- [15] M. M. Rana and N. Halim, "Motion capture systems using optimal signal processing algorithm: A state-of-the-art literature," *Universal Journal of Communications and Network*, vol. 6(1), no. 1, pp. 1–5, 2018.
- [16] D. A. Torse and V. V. Desai, "Design of adaptive EEG preprocessing algorithm for neurofeedback system," in *International Conference on Communication and Signal Processing*, 2016, pp. 0392–0395.
- [17] Q. Wang and O. Sourina, "Real-time mental arithmetic task recognition from EEG signals," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 21, no. 2, pp. 225–232, 2013.
- [18] M. Islam, M. M. Rana, and M. M. Rahman, "Application of principal components analysis (PCA) for image processing," in *International Conference on Electrical Computer and Telecommunication Engineering*, 2012, pp. 1–5.
- [19] M. M. Rahman, A. A. Hossain, M. M. Rana, and K. Mitobe, "Hand motion capture system in piano playing," in *Proc. International Conference on Informatics, Electronics and Vision*, 2013, pp. 1–5.



Ahmed Abdelhadi (SM'16) is an Assistant Professor at the University of Houston. Before joining UH, he was a Research Assistant Professor at Virginia Tech. He received his Ph.D. in Electrical and Computer Engineering from the University of Texas at Austin in 2011. He was a member in Wireless Networking and

Communications Group (WNCG) and Laboratory of Informatics, Networks and Communications (LINC) group during his Ph.D. In 2012, he joined Bradley Department of Electrical and Computer Engineering and Hume Center for National Security and Technology at Virginia Tech. He was a faculty member of Wireless @ Virginia Tech. His research interests are in the areas of wireless communications and networks, spectrum sharing, cyber physical systems, and security. Dr. Abdelhadi coauthored more than 70 journal and conference papers, and 5 books in these research topics.