2020 IEEE Region 10 Symposium (TENSYMP), 5-7 June 2020, Dhaka, Bangladesh

Efficient Approach to Detect Epileptic Seizure using Machine Learning Models for Modern Healthcare System

Tanbin Islam Rohan, Md. Salah Uddin Yusuf, Monira Islam and Shidhartho Roy
Department of Electrical and Electronic Engginerring
Khulna University of Engineering & Technology (KUET)
Khulna-9203, Bangladesh
tanbinislam009@gmail.com, suyusuf@eee.kuet.ac.bd, monira_kuet08@yahoo.com, swapno15roy@gmail.com

***Abstract*- Epileptic seizure is one of the common neurological disorder now a day. But this is curable if it can be detected in the early stage. So, this research become a necessity in the early prediction of epileptic seizure. A complete and reliable system can classify the epileptic seizure patients and the states of epileptic seizure. This research explores a supervised machine learning and deep learning model for the classification of epileptic seizure patients from the Epileptic Seizure dataset of UCI machine learning repository. The dataset has 11,500 instances; every information contains 178 attributes. XGBoost is used for the Machine learning approach and ANN is used for Deep learning approach. The proposed ANN algorithm has the improved accuracy and accurately classified the epileptic seizure class patients. 10-fold cross validation is used for the validation purpose. XGBoost acquires 96.6% test accuracy and ANN acquires 98.26% test accuracy. The proposed Deep Learning approach has out-performed the conventional epileptic seizure classifier algorithms. Additionally, the Deep learning model enhances the performance of epileptic seizure detection.**

Keywords: Epileptic Seizure, XGboost, ANN, 10-fold cross-validation, Healthcare.

I. INTRODUCTION

Epilepsy is one of the world's most well-known neurological diseases affecting 50 million people [1]. Spontaneous sensory blasts within the brain are collected with the Electroencephalogram (EEG), a non-invasive scalp electrode, which can be used to diagnose multiple neurological conditions. Seizures result from non-coordinated neuronal electrical discharge, primarily epilepsy, and recurring seizures in the cerebral cortex [2]. This may cause serious damage, including collapses and sudden unforeseen epilepsy death. If seizures are controlled and medical assistance is given when seizures occur, the likelihood of mortality will be decreased. About 30% of patients with epilepsy do not react to medicines. To enhance the quality of their lives, extracting new knowledge from their physiological signals is very critical. [3].

It is not simple and still a challenge to automatically classify the EEG signal. There is no obvious distinction between non-epileptic seizures in EEG signals in individuals closing their eyes to patients with epileptic seizures. Also, it is hard to acknowledge EEG signals in patients with tumor region between epileptic seizures to non-epileptic seizures. Many scientists published their work on automatic epileptic seizure, but most of it used only two categories, the EEG seizure signal epileptic and the EEG seizure signal non-epileptic. It is harder to identify five categories than two classes in the machine learning domain. It is very helpful for neurologist and medical practice to diagnose their clients if we are able to properly identify EEG information in five categories annotated in the information set.

In an attempt to reduce the computational complexity of seizure detection systems, the authors of [4] used the low-complex support vector (SVM) classifier along with the traditional wavelet features. The average classification accuracy of this technique was 95.33 percent. In [5], a similar rating precision of 95.61% was accomplished with a new EEG extraction technique, while sensor side characteristics were obtained and a server-side seizure analysis was conducted. In order to identify the presence of epileptic activities [6] a feature extraction technique based on Hilbert transformation was also applied alongside an SVM Classifier. It accomplished 97.00% classification accuracy. The autonomous seizure-associated characteristics were also assessed using independent component analysis (ICA) [7]. With the SVM index, 96.00 percent, 94.00 percent and 95.00 percent respectively, were used to attain awareness, specificity and classification exactness. In [8], an average accuracy 97.19% with a brief computation time of 0.065 seconds was shown in a threat tracking system relying on some statistical characteristics and a least-square SVM (LSSVM). In [9], the performance of machine learning algorithms is used to classify the epileptic seizure class and SVM obtained the highest 97.31% accuracy. In the proposed method of [10], CHB-MIT Scalp EEG database was used and 98% of sensitivity was obtained. In [11], SVM classifier was used and the performance parameters were AUC, sensitivity, specificity and the values were 0.7622, 0.7252, and 0.7252 correspondingly. Using DNN the F-measure, precision and sensitivity were 95.13, 94.21, and 96.27 respectively in [12].

The rest of this paper is organized as follows. Section II introduces the methodology of the proposed research work. Experimental results and analyses have been carried out in Section III. Finally, Section IV concludes the result.

II. METHODOLOGY

In case of epileptic seizure detection, the classification shows a very important role. For the case of multi class classification we generally use Machine Learning algorithms. But for some data sets the Machine Learning algorithms are not applicable in all cases. This type of data sets shows nonlinear characteristics as well. So we can get better accuracy by using neural networks. The Block diagram of the proposed research is shown in Fig.1.

978-1-7281-7366-5/20/$31.00 ©2020 IEEE

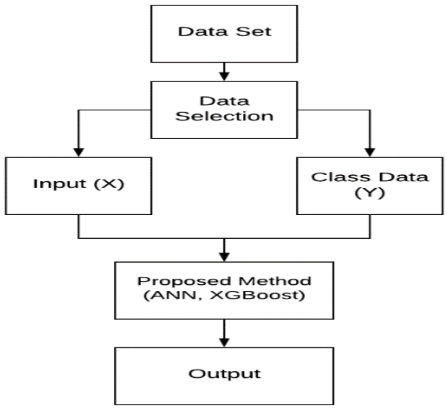

Fig. 1. Block diagram of the proposed approaches.

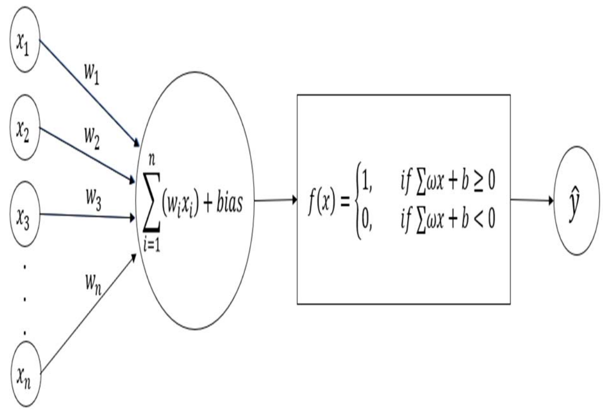

Fig. 2. ANN structure for multi-class classification.

A. Data Collection and Provision

The data set called Epileptic Seizure is retrieved from the UCI machine learning repository [13]. The data set consists of 11500 instances, 179 attributes and there are 5 cases. The data set was shuffled in every 4097 data points into 23 parts, every part consists of 178 data points for 1 second. Every data point is the value of the EEG recording at a diverse point in time. So, the data set consists of 11500 rows, 178 data points for 1 second as column. The last column signifies the label y {1, 2, 3, 4, 5}. The rejoinder variable is y in column 179, the Explanatory variables X1, X2, ………………....., X178 and y comprises the group of the 178-dimensional input vector. Precisely, y in {1, 2, 3, 4, 5}:

Case 5 is (Eyes open, means when they were recording the EEG signal of the brain the patient had their eyes open). Case 4 is (Eyes closed, means when they were recording the EEG signal the patient had their eyes closed). Case 3 is (They could identify where the region of the tumor was in the brain and recording the EEG activity from the healthy brain area). Case 2 is (They recorded the EEG from the area where the tumor was located). Case 1 is (Recording of seizure activity).

All subjects falling in classes 2, 3, 4, and 5 are subjects who didn't have an epileptic seizure. Just subjects in class 1 have an epileptic seizure. Our inspiration for making this form of the information was to disentangle access to the information through the production of a .csv rendition of it. Despite the fact that there are 5 classes most creators have done double order, to be specific class 1 (Epileptic seizure) against the rest.

B. ANN

The ANN structure for multi-class classification is shown in Fig.2.Unsupervised learning is a part of AI that gains from test information that has not been named, ordered or classified. Rather than reacting to input, solo learning recognizes shared characteristics in the information and responds dependent on the nearness or nonattendance of such shared characteristics in each new bit of information. Choices incorporate managed learning and fortification learning. The fundamental structure of an ANN comprises of counterfeit neurons (similar to organic neurons in the human mind) that are gathered into layers. The most well-known ANN structure comprises an information layer, at least one shrouded layers and a yield layer.

The arbitrary loads are allocated to all the linkage to begin the calculation. The enactment pace of concealed hubs is found by utilizing the sources of info and the linkages. Utilizing the actuation pace of shrouded nodes and linkages to yield the initiation pace of output nodes are found. The blunder rate is found at the output node and recalibrate all the linkages between hidden hubs and output hubs. The loads and blunder found at yield hub are utilized the mistake to shrouded hubs are fell down. Recalibrate the loads between hidden nodes and input nodes. Rehash the procedure till the union model is met. Utilizing the last linkage loads score the initiation pace of the output nodes.

C. XGBoost

XGBoost is one of the most established and proficient usages of the Gradient Boosted Trees calculation, an administered learning strategy that depends on a capacity guess by enhancing explicit misfortune works just as applying a few regularization strategies.

The loss function in XGBoost has an extra guideline term omega. This loss function is used to calculate maximum gain which can be directly used in tree building process while splitting each node. Along these lines, this is the means by which the two-stage process in GBM is decreased to single step with regularization. Beginning with Loss work that should be limited.

$$Loss = \sum_{i=1}^{n}\left[g_i h_m(x_i) + \frac{1}{2}k_i h_m^2(x_i)\right] + \Omega(h_m T) \quad (1)$$

Loss function is expanded similar to Taylor series. The consistent term created is evacuated in light of the fact that they won't contribute in figuring gain while parting a node in tree.

The Regularization term omega for a tree can be obtained as:

$$\Omega_m(hm_T) = \epsilon T + \frac{1}{2}\lambda \sum_{j=1}^{T} bj_m^2 \quad (2)$$

Substituting the value of hm(x) as Tree equation and dropping m (iteration number) for neatness, Loss function become:

$$Loss = \sum_{j=1}^{T}\left[g_j b_j + \frac{1}{2}(H_j + \lambda)b_j^2\right] + \epsilon T \quad (3)$$

If the 1st term is noticed, the summation term, is an equation of parabola. It can be seen as $ax + bx^2$. The minimum value for a parabola will occur at $x = -a/2b$. In our case, $b_j = -G_j/(H_j + \lambda)$. Substituting b_j in loss function, the optimum loss function becomes:

$$Loss = \frac{1}{2}\sum_{j=1}^{T}\left(\frac{G_j^2}{H_j+\lambda}\right) + \epsilon T \quad (4)$$

The Gain when a leaf is split into two leaves (Left L and Right R) becomes:

$$Gain = -\epsilon + \frac{1}{2}\left[\frac{G_L^2}{H_j+\lambda} + \frac{G_R^2}{H_j+\lambda} - \frac{(G_L+G_R)^2}{H_j+\lambda}\right] \quad (5)$$

III. Results and Discussion

An accurate Epileptic Seizure detection technique is proposed in this work. Two algorithms were used for the detection purpose. XGBoost and ANN were executed on a computer configured with 8GB RAM and Intel Core i5 processor. Scikit-learn has been used which is a open source tool was used. Scikit-learn is built in python language for machine learning library. For this detection process a well-known open-source cross-platform IDE named Spyder is used which is developed in Python language for scientific programming and analysis.

XGBoost algorithm has been used for the machine learning approach and Artificial Neural Network (ANN) is used for the Deep learning approach. Data has been split into 10350 (90%) instances for training and rest of the 1150 (10%) instances for testing purpose for the both machine learning and deep learning approach. 10-fold cross validation was used for the model. In the XGBoost classifier seed through random state was taken 69 for the ultimate gradient boosting for this work. It was the preferable random state for the highest accuracy optimization. In the deep learning approach, the Artificial Neural Network has been used. In the input layer sequential model is used and the activation function was 'relu'. Four hidden layers are used for the expected accuracy as the dataset consists of five classes of dataset. In every hidden layer the activation function was also 'relu'. In the output layer the result was filtered by softmax activation function.

In the Machine learning approach XGBoost gave 96.6% accuracy in the test phase and 100% accuracy in the training phase through in the training phase 11500 out of 11500 instances are correctly identified. In the test phase 1111 out of 1150 instances were correctly identified. In the deep learning approach Artificial Neural Network (ANN) gave 98.26% accuracy in the test phase and 100% accuracy in the training phase. In the training phase 11500 out of 11500 instances are correctly identified. In the test phase 1130 out of 1150 instances are correctly identified. The comparison of the proposed approach result with others are shown in Table1 as well as in Fig. 3.

As the dataset was 5 classes through 5 different cases, the expected outcome of the work is satisfying as the results are obtained very precisely. The stability of the proposed algorithms has been judged in different cases. As the variation of the test and training set have been done, different results have obtained from there. According to the Fig. 4 it's clear that the stability of the Deep Learning method through

TABLE I. Comparison Of Proposed Approach With Others Work

Algorithm	Accuracy (%)
SVM [9]	97.31
Random Forest [9]	97.08
Naïve-Bayes [9]	95.98
Neural Network [9]	93.53
KNN [9]	90.01
Proposed XGBoost	96.60
Proposed ANN	98.26

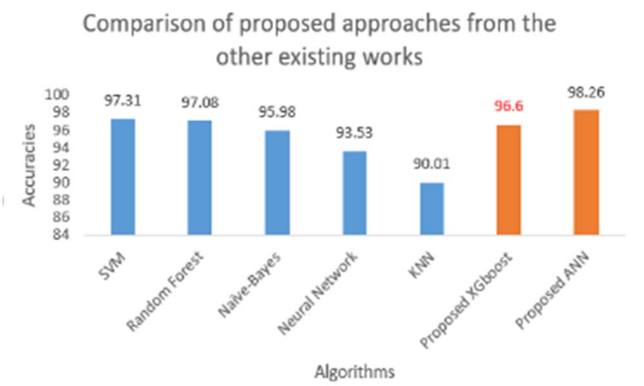

Fig. 3. Comparison of proposed approaches from the other existing works.

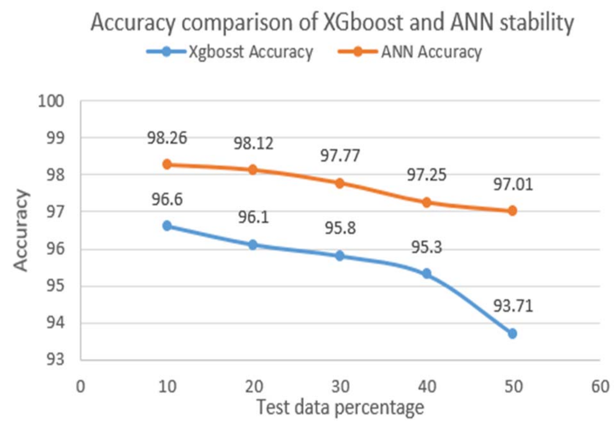

Fig. 4. Accuracy Comparison of XGboost and ANN Stabilility

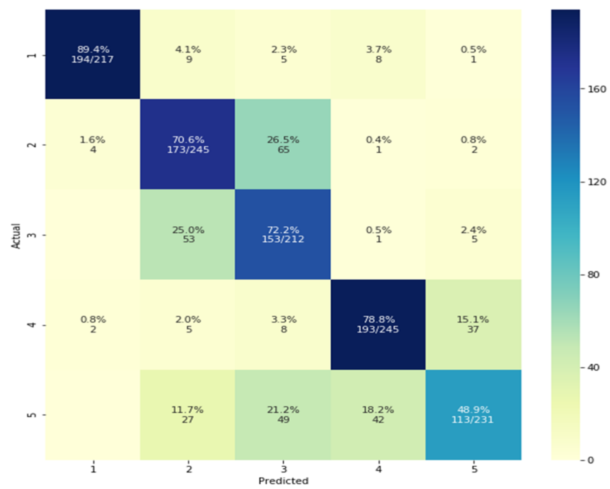

Fig. 5. 5 × 5 Confusion Matrix for the 5 cases of Epileptic Seizure.

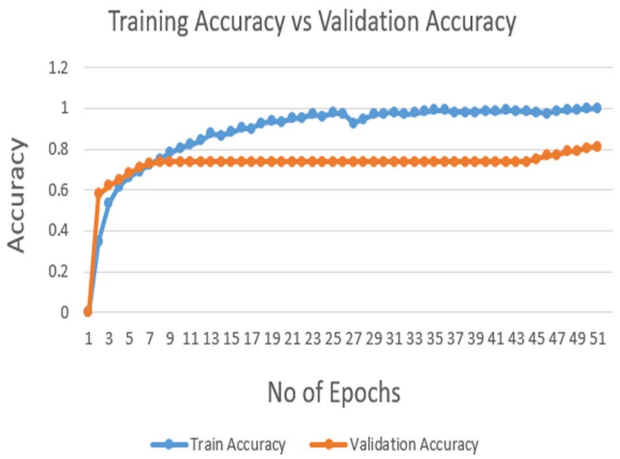

Fig. 6. Training accuracy vs Validation accuracy curve.

ANN is showing the better stability than the Machine Learning Approach through XGBoost algorithm.

Fig. 5 shows a 5 × 5 confusion matrix. To extract the accuracy from this confusion matrix label binarization is done. This is a problem with a great computational complexity because there are five different cases. Train accuracy and validation accuracy curves from ANN approach are shown in Fig 6. Validation test is done for checking the robustness of the algorithm.

IV. Conclusion

Medical condition prediction is troublesome work and most of the time it goes wrong in case of automated prediction. In this matters machine learning and deep learning models give an accurate prediction approximately. This paper offers an approach for the multi class classification of the epileptic seizure. By early detection the problem can be cured. XGBoost and ANN have given promising results in accordance of accuracy but ANN is recommendable for the precise multiclass classification. As the Machine learning and Deep learning models are preferable for the accurate detection. This results strongly suggested that the proposed Artificial Neural Network can play a significant role in the field of epileptic seizure detection.

References

[1] U.R. Acharya, et al., "Automated EEG analysis of epilepsy: a review," Knowledge-Based Systems. Vol. 45, p. 147-165, 2013.

[2] P. Artameeyanant, W. Chiracharit and K. Chamnongthai, "Spike and Epileptic Seizure Detection Using Wavelet Packet Transform Based on Approximate Entropy and Energy with Artificial Neural Network", 2012 Biomedical Engineering International Conference (BMEiCON-2012)

[3] E. Foundation, "www.epilepsy.com/learn/impact/mortality," 2015.

[4] Y. Liu, W. Zhou, Q. Yuan, and S. Chen, "Automatic seizure detection using wavelet transform and SVM in long-term intracranial EEG," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 20, no. 6, pp. 749-755, Nov. 2012.

[5] J. Chiang, and R. Ward, "Energy-efficient data reduction techniques for wireless seizure detection systems," *Sensors (Basel, Switzerland), PMC. Web.*, vol. 14, no. 2, pp. 2036-2051, July 2014.

[6] Rahul Kumar Chaurasiya, K. Jain, S. Goutam, and Manisha, "Epileptic seizure detection using HHT and SVM," *International Conference on Electrical, Electronics, Signals, Communication and Optimization (EESCO)*, pp. 1-6, Jan. 2015.

[7] M. P. Hosseini, A. Hajisami, and D. Pompili, "Real-time epileptic seizure detection from EEG signals via random subspace ensemble learning," *IEEE International Conference on Autonomic Computing*, Wurzburg, pp. 209-218, 2016.

[8] D. S. T. Behara, A. Kumar, P. Swami, B. K. Panigrahi, and T. K. Gandhi, "Detection of epileptic seizure patterns in EEG through fragmented feature extraction," *International Conference on Computing for Sustainable Global Development*, New Delhi, pp. 2539-2542, 2016.

[9] Resque, P., Barros, A., Rosário, D., & Cerqueira, E. (2019, June). "An Investigation of Different Machine Learning Approaches for Epileptic Seizure Detection", 2019 *15th International Wireless Communications & Mobile Computing Conference* (IWCMC) (pp. 301-306). IEEE.

[10] Fan, M., & Chou, C. A. (2018). "Detecting abnormal pattern of epileptic seizures via temporal synchronization of EEG signals", *IEEE Transactions on Biomedical Engineering*, 66(3), 601-608.

[11] Popov, A., Panichev, O., Karplyuk, Y., Smirnov, Y., Zaunseder, S., & Kharytonov, V. (2017, September). "Heart beat-to-beat intervals classification for epileptic seizure prediction", *Signal Processing Symposium* (SPSympo) (pp. 1-4). IEEE, 2017

[12] Birjandtalab, J., Heydarzadeh, M., & Nourani, M. (2017, August). "Automated EEG-based epileptic seizure detection using deep neural networks" *IEEE International Conference on Healthcare Informatics* (ICHI) (pp. 552-555). IEEE, 2017

[13] Andrzejak, R. G., Lehnertz, K., Mormann, F., Rieke, C., David, P., & Elger, C. E. (2001). "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, 64(6), 061907", Available in: https://archive.ics.uci.edu/ml/datasets/Epileptic+Seizure+Recognition