Seizure Onset Detection and Prediction using EEG Data

Anirudh R. Shandilya
Student, Electronics and Communication Engineering
BNM Institute of Technology (Aff. to VTU)
Bengaluru, India

Anusha A. N.
Student, Electronics and Communication Engineering
BNM Institute of Technology (Aff. to VTU)
Bengaluru, India

Mrs. Sumathi A.
Associate Professor, Dept. of ECE
BNM Intstitute of Technology (Aff. to VTU)
Bengaluru, India

***Abstract*—Artificial Intelligence (AI) deals with programming systems to take decisions on their own. The use of Machine Learning (ML) concepts in giving systems to learn an algorithm and make corrections by training itself is an important tool for human beings, as it removes the possibility of writing programs manually for every probable iteration of any functionality. This ability is used as an essential way of predicting and detecting the onset of seizure with the help of Electroencephalogram (EEG) data. The system is provided with the data and, as it gains experience, the model can easily detect and give an accurate prediction of the onset of seizure before the actual seizure attack and displays on the LCD screen.This system is patient-independent.**

***Index Terms*—EEG, neural network, seizure, epilepsy**

I. Introduction

Seizure attack is the phenomenon when the brain undergoes a sudden surge of impulses, which disorders the brain functionality. This attack can occur on any circumstance, and it also depends on innumerable factors that the patient would have had before or during the attack. Due to this unpredictable nature of seizure attacks, doctors have found it difficult to provide proper medication and pre-attack measures that the patient can undergo in order to be prepared better for further such attacks in the future. Although some seizure attacks are very minor and sometimes would not be perceived during normal activity, there might be cases where it could become fatal. One of the most brutal causes associated with seizures is epilepsy.

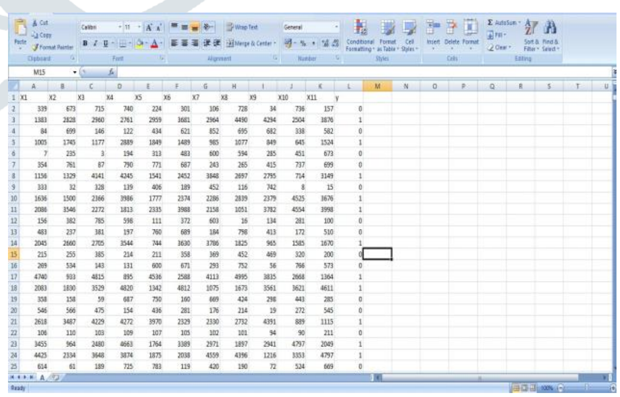
When an epileptic attack occurs, the patient loses control over his body, and his entire nervous system is disrupted. This would sometimes lead to unconsciousness and retarded behaviour. A large proportion of the population that exists suffer seizures due to unhealthy live styles, but no proper method has been devised for their control. The major causes of epilepsy is the existence of some injury that a person would have suffered with his brain(tumors, head injuries, strokes etc.)

Electroencephalogram (EEG) is a report of all the brain activity. It contains vital information that can aid in detection and prediction of epileptic seizures. Many different components exist in the EEG graph, for instance the ictal and preictal stage. By experimentation, it can be shown that the preictal stage contains information, using which the detection of the seizure prior to the attack can be done efficiently. This model is aimed to create a system that accurately detects the presence of this information in the preictal state, from which the successful prediction of an epilepsy attack can be done.

II. Literature Survey

A. Data set

The data represents 23.6 seconds clips of preictal and interictal training and test readings for 3 electrodes on 250 epileptic humans and 250 non-epileptic humans to give a single voltage. Each row represents a single person. So, 500 rows are present in the dataset. The columns contain electrode readings at a particular time segment. The names of the EEG electrodes are given in a vector as X1,X2,..X11. The last row 'Y' shows the classification.



X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	y
339	673	715	740	224	301	106	728	34	736	157	0
1383	2828	2960	2761	2959	3681	2964	4490	4294	2504	3876	1
84	699	146	122	434	621	852	695	682	338	582	0
1005	1745	1177	2889	1849	1489	985	1077	849	645	1524	1
7	235	3	194	313	483	600	594	285	451	673	0
354	761	87	790	771	687	243	265	415	737	699	0
1156	1329	4141	4245	1541	2452	3848	2697	2795	714	3149	1
333	32	328	139	406	189	452	116	742	8	15	0
1636	1500	2366	3986	1777	2374	2286	2839	2379	4525	3676	1
2086	3546	2272	1813	2335	3988	2158	1051	3782	4554	3998	1
156	382	785	598	111	372	603	16	134	281	100	0
483	237	381	197	760	689	184	798	413	172	510	0
2045	2660	2705	3544	744	3630	3786	1825	965	1585	1670	1
215	255	385	214	211	358	369	452	469	320	200	0
269	534	143	131	600	671	293	752	56	766	573	0
4740	933	4815	895	4536	2588	4113	4995	3835	2668	1364	1
2083	1830	3529	4820	1342	4812	1075	1673	3561	3621	4611	1
358	158	59	687	750	160	669	424	298	443	285	0
546	566	475	154	436	281	176	214	19	272	545	0
2618	3487	4229	4272	3970	2329	2330	2732	4391	889	1115	1
106	110	103	109	107	105	102	101	94	90	211	0
3455	964	2480	4663	1764	3389	2971	1897	2941	4797	2049	1
4425	2334	3648	3874	1875	2038	4559	4396	1216	3353	4797	1
614	61	189	725	783	119	420	190	72	524	669	0

Fig. 1. Data set

There is no guarantee that the EEG sensors were placed on the exact same location on the patient's brains or arms.

The original data set had of 5 different folders. It had 100 files, with each file that represented a single person. Each file was a recording of brain activity for 23.6 seconds. The corresponding time-series is sampled into 4097 data points, which is divided and shuffled into 23 chunks, each chunk contains 178 data points for 1 second, and each data point was the value of the EEG recording at a different point in time. Each information contains 178 data points for 1 second(column), the last column represented the label y 1,2,3,4,5. 1 represents epileptic group, the rest were non-epileptic groups, as cited by the reference paper. Here, the data set was created by correlating with the original data set that could be measure by the EEG sensors available.

B. Data Preprocessing

The data preprocessing consists of a data mining technique. EEG data was taken directly through the sensors attached. This data was raw which cannot be interpreted by the algorithm. Therefore it had to be transformed into an understandable format. These tasks were done before using the data set: 1. Data cleansing 2. Data editing. 3. Data reduction 4. Data wrangling

Another important process involved was feature scaling. Feature scaling was done to remove additional variables that are available in the data proclaimed. By doing so, the data could be used for comparing using a common ground.

III. METHODOLOGIES

The data set as cited in the reference paper[1], II-A, is sampled into 178 data points one second. 250 epileptic people and 250 non-epileptic peoples' data was created and considered.For the created data set, '2', '3', '4', and '5' were grouped as '0' and '1' was labeled '1', with reference to the original data set. The data set created, has 11 data points for 23.6 seconds and the label "y", which had the classification. This label has 2 different numbers assigned, and each number represents the conditions that was put in place during the recording session of the EEG. This data set was shuffled for proper training.

0 - Non-epileptic, Normal person

1 - Epileptic Activity of the person

The entire decision making algorithm comprised many algorithms that combined together to provide the maximum efficiency. The four component functionalities that were present in the algorithm were:

1. Feature extraction
2. Normalization
3. Dimensional reduction
4. Classification

Short-Time Fourier Transform (STFT) was mainly used to extract the essential features that is required for the seizure detection. Since it was domain-specific, STFT also provides vital information when the EEG signals were read. The EEG data was a short clip, usually 8-10 minutes long, and it covered both the preictal and interictal stages[1]. However, there was a lot of information that was interfering in the feature extraction, and had to be either reduced or removed completely. The candidates were either going to be Principal Component Analysis (PCA) or Discrete Wavelength Transform (DWT)[2]. On experimentation, it was found that DWT actually removed key parts of the signals, which hindered the accuracy of STFT, whereas PCA enhanced the accuracy, making STFT perform better with PCA. PCA also helped in two other important functions, normalizing the data, and most importantly, reduced the dimensions. This made STFT to perform feature extraction without the interruption of the non-essential parts of the data[3][7]. The final step involved the simple binary classification. The two algorithms considered for this was the Logistic Regression (LR) and the Support Vector Machine (SVM). It was seen that LR had the upper hand in the model, due to its higher classification rate than SVM when the training data was provided. The reason for this could have been the way LR functions. LR uses a probabilistic approach during classification, as opposed to SVM's more strict approach. Due to the noise present in the training data, SVM provided a low classification accuracy due its interference[4]. K-Nearest Neighbour (KNN) could have also been used, but the noise hindrance made it impossible to distinguish the data points accurately, and thus the classification got affected as cited[5].

In this paper, we are using Artificial Neural Networks (ANN) was used so that any noise could be controlled and the human interference in the training process is very less. ANN could be also one of the method that gained highest accuracy also.

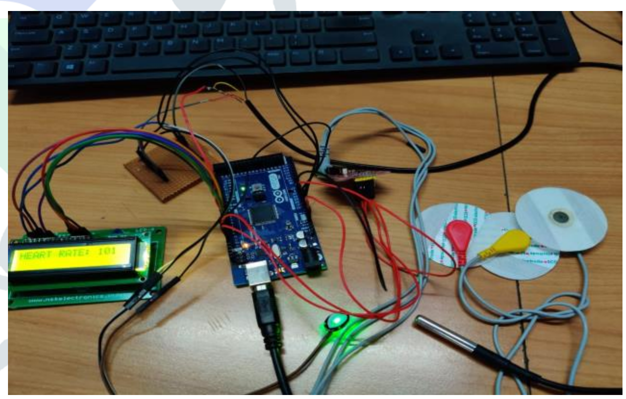


Fig. 2. Hardware Connection

IV. RESULTS

Due to the lockdown, the data that would be provided by a recognised medical institute could not be provided, the sufficient training of the model could not be done. The data was scarce to begin with, due to the kind of data that was needed for the training process to take place, the algorithm had to be suffice with the little amount of information that was in hand. Another problem that was encountered was the overall information that can be derived from the existing data.

Due to lack of background information regarding the processes that were exercised during the recording, some of the aspects that may have made an influence to the subsequent results, the algorithm could not categorically and precisely predict.

With a data set that was small in size and predicted results already known, the algorithm showed an accuracy percentage of 100%. A resolve was made to create a new data set by physically taking EEG recordings from the sensors connected to the system. This method is a tedious task, as the process involves recording of data without the intervention of noise, and mapping the information with the data set that was in possession. Only a small subset of the data was created. The epochs are uploaded and the accuracy can be seen as almost 100%. The output is given in terms of correlation to the data

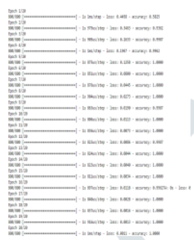


Fig. 3. Epochs

set, hence it is a decimal number. In order to normalize it, a threshold is specified. Here, 0.2 is specified as the threshold to the correlation. It is taken threshold in order to save a patient from any minor epilepsy too, so that the person can necessary precautions. Any correlation below 0.2 indicates non-epileptic stage. All those above 0.2 has been into epileptic stage. The output is accurately specified and the output is predicted successfully.

Fig. 4. Output

With that data set in hand, the accuracy shown by the algorithm was significantly reduced. Repeated testing showed a varied range of percentages. These were, however, got fit into a range of around 50 to 70 %. The algorithm had to be suffice in giving these accuracy levels until proper training of it could take place.

V. Conclusion

The aim of the project was to detect and predict onset of seizure attacks before their occurrence, such that the said patient can take relevant steps for their safety. A neural network was designed that was subjected to training by feeding EEG data. The prediction is done by the detection of the preictal stage of EEG. This stage contains information and shows certain characteristics when an epileptic attack would occur. The task of the neural network is to detect the preictal stage, study the information it possesses, and perform a binary classification. The data set for training is recorded with an EEG sensor, along with a pulse reader and a temperature sensor, as they also aid in detection of the preictal stage. The recording procedure follows a detailed method, by which the EEG data is recorded with as minimum noise as possible. Before feeding this data to the network for classification, certain preprocessing steps are taken in order to remove the recorded parts which are not required and also has a negative effect on the classification. The data is then given to the network, which performs the classification, and predicts whether the patient is going to have a seizure attack. As stated previously, the unseen events of the current year put us in a tight situation. The pandemic prevented us from getting the right amount of data that was of utmost importance in efficiently providing the necessary training of the neural network. But it was impossible as the neural network was heavily dependent on the data provided by a reputed medical institute, and also given the fact that such information was considered confidential, and hence was not found on the internet. If provided with sufficient amount of data for training, the network can give a prediction accuracy of about 80 to 90 %.

VI. References

[1] Alex Greaves, Arushi Raghuvanshi, Kai-Yuan Neo. Predicting Seizure Onset with Intracranial Electroencephalogram (EEG) Data - Project Report, December 2014.

[2] A signal processing based analysis and prediction of seizure onset in patients with epilepsy by Hamidreza Namazi,Vladimir V. Kulish, Jamal Hussaini, Jalal Hussaini, Ali Delaviz, Fatemeh Delaviz, Shaghayegh Habibi and Sara Ramezanpoor, January 5, 2016.

[3] Human Intracranial EEG Quantitative Analysis and Automatic Feature Learning for Epileptic Seizure Prediction by Ramy Hussein, Mohamed Osama Ahmed, Rabab Ward, Z. Jane Wang, Levin Kuhlmann, Yi Guo, 7 April, 2019.

[4] Comparison Machine Learning Algorithms for Recognition of Epileptic Seizures in EEG by Bekir Karlık and Şengül Bayrak Hayta, 9 April, 2014.

[5] Forecasting Seizures Using Intracranial EEG Measures and SVM in Naturally Occurring Canine Epilepsy by Benjamin H. Brinkmann, Edward E. Patterson,Charles Vite,Vincent M. Vasoli,Daniel Crepeau,Matt Stead,J. Jeffry Howbert,Vladimir Cherkassky,Joost B. Wagenaar,Brian Litt,Gregory A. Worrell, 4 August, 2015.

[6] Yaylali I, Koçak H, Jayakar P. Detection of seizures from small samples using nonlinear dynamic system theory. IEEE Trans Biomed Eng. 1996; 43:743–751. [PubMed] [Google Scholar]

[7] Subasi A, Gursoy MI. EEG signal classification using PCA, ICA, LDA and support vector machines. Expert Syst Appl. 2010; 37:8659–8666. [Google Scholar]

[8] Ghaderyan P, Abbasi A, Sedaaghi MH. Linear features, principal component analysis, and support vector machine for epileptic seizure prediction progress. 21st Iranian Conference on Electrical Engineering (ICEE); Mashhad, Iran. 2013. pp. 14–16. [Google Scholar]

[9] Ghaderyan P, Abbasi A, Sedaaghi MH. An efficient seizure prediction method using KNN-based undersampling and linear frequency measures. J Neurosci Meth. 2014; 232:134–142. [PubMed] [Google Scholar]

[10] Khamis H, Mohamed A, Simpson S. Seizure state detection of temporal lobe seizures by autoregressive spectral analysis of scalp EEG. Clin Neurophysiol. 2009; 120:1479–1488. [PubMed] [Google Scholar]

[11] Świderski B, Osowski S, Cichocki A, Rysz A. Epileptic Seizure Prediction Using Lyapunov Exponents and Support Vector Machine. Spring Comp Sci. 2007; 4432:373–381. [Google Scholar]

[12] Lai YC, Harrison MA, Frei MG, Osorio I. Controlled test for predictive power of Lyapunov exponents: their inability to predict epileptic seizures. Chaos. 2004; 14:630–642. [PubMed] [Google Scholar]

[13] Swiderski B, Osowski S, Rysz A. Lyapunov exponent of EEG signal for epileptic seizure characterization. Proceedings of the 2005 European Conference on Circuit Theory and Design. 2005 doi: 10.1109/ECCTD.2005.1523016. [CrossRef] [Google Scholar]

[14] Li J, Yan J, Liu X, Ouyang G. Using Permutation Entropy to Measure the Changes in EEG Signals During Absence Seizures. Entropy. 2014; 16:3049–3061. [Google Scholar]

[15] Li X, Ouyang G, Richards DA. Predictability analysis of absence seizures with permutation entropy. Epilepsy Res. 2007; 77:70–74. [PubMed] [Google Scholar]