International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-10 Issue-1, October 2020

Analysis of EEG signals using Machine Learning for the Detection and Diagnosis of Epilepsy

Anubha Nagar, Bidushi, Mimangsha Sarma, Mithra Anand Kumar, J.Valarmathi

Abstract*: Electroencephalogram (EEG) is one of the most commonly used tools for epilepsy detection. In this paper we have presented two methods for the diagnosis of epilepsy using machine learning techniques. EEG waveforms have five different kinds of frequency bands. Out of which only two namely theta and gamma bands carry epileptic seizure information. Our model determines the statistical features like mean, variance, maximum, minimum, kurtosis, and skewness from the raw data set. This reduces the mathematical complexities and time consumption of the feature extraction method. It then uses a Logistic regression model and decision tree model to classify whether a person is epileptic or not. After the implementation of the machine learning models, parameters like accuracy, sensitivity, and recall have been found. The results for the same are analyzed in detail in this paper. Epileptic seizures cause severe damage to the brain which affects the health of a person. Our key objective from this paper is to help in the early prediction and detection of epilepsy so that preventive interventions can be provided and precautionary measures are taken to prevent the patient from suffering any severe damage

***Keywords* : Epilepsy, EEG, Decision Tree model, Logistic regression, seizures.**

I. INTRODUCTION

Epilepsy is a neurological illness in which the activity of the nerve cells is affected in the brain, causing seizures, sensations, periods of unusual behavior and at times loss of awareness[23]. For the diagnosis of epilepsy, the minimum requirement is at least two unprovoked seizures. The symptoms for these seizures vary from person to person (which include staring blankly for a period of time or twitching of arms and legs). According to the International League against Epilepsy (ILAE), Epilepsy is a "chronic condition of the brain characterized by an enduring predisposition to generate epileptic seizures, by the neurobiological, cognitive, psychological, and social consequences of this condition"(Fisher et al., 2014) [15].Around 50 million people suffer from Epilepsy.It is one of the most common brain disorders, can occur at all ages, and have many possible presentations and causes. Although incidence in childhood has fallen over the past three decades in developed countries, this disease is matched by an increase in elderly people[20].As per a study conducted by Ngugi AK et al., Epilepsy is more common in underdeveloped countries than in developed countries[17]. This is because there are low standards of hygiene and nutrition, the prenatal care is poor and there is a higher risk associated with brain injuries, endemic conditions, and other infections related to thecerebral activities[21]. People diagnosed with Epilepsy often stop having seizures within a few years of diagnosis. This condition is known as spontaneous remission[16]. Even for those whose seizures have ceased to occur, the recurrence of seizures and their severity can be reduced with the help of therapy. Mortality rates of people diagnosed with epilepsy is higher than those who do not have epilepsy. Sudden Unexpected Death in Epilepsy (SUDEP) is a concerning problem. Although the main reason for this is not known, SUDEP occurs either after an encounter of a seizure or during the episode of an encounter.

The possible factors include:

1. Breathing:

A person who experiences a seizure may have trouble breathing (pauses during breathing - apnea). There can be a tremendous amount of reduction in the intake of oxygen which can lead to a life-threatening situation. During a convulsive seizure, there can be obstructions that lead to suffocation.

2. Heart Failure/Heart Rhythm:

In rare cases, the heart rhythm is affected, this condition is also known as cardiac arrhythmia and sometimes it leads to heart failures. A combination of the above-mentioned factors can also lead to SUDEP. SUDEP is more likely to occur in people who have seizures while they are asleep [18], or those who experience frequent convulsive seizures, or those who have intellectual disabilities [19] and those who experience symptomatic epilepsy.

EEG (electroencephalogram) is a device used to detect brain waves. It records the electrical impulses of the brain. Continuous changes in brain activity are monitored using the EEG for a short period of time. EEG waveforms amplitude ranges from 10V to 100V and frequency in the range of 1 Hz to about 100 Hz.

Revised Manuscript Received on September 15, 2020.

* Correspondence Author

Anubha Nagar*, Department of Electronics and Communication, Vellore Institute of Technology, Tamil Nadu, India.

Bidushi, Department of Electronics and Communication, Vellore Institute of Technology, Tamil Nadu, India.

Mimangsha Sarma, Department of Electronics and Communication, Vellore Institute of Technology, Tamil Nadu, India.

Mithra Anand Kumar, Department of Electronics and Communication, Vellore Institute of Technology, Tamil Nadu, India.

Dr. Valarmathi J., Professor, Vellore Institute of Technology, Tamil Nadu, India.

EEG frequency pattern as follows: beta (above 13 Hz), alpha (7-13 Hz), theta (4-7 Hz) and delta (0.5-4 Hz)[25].Traditionally neurologists would read the EEG reading visually and identify any abnormalities in the waveform. However, this method may sometimes result as inaccurate and erroneous. This happens because there are no fixed criteria based on which they detect the seizures or abnormalities. Moreover, such human error can result in serious issues of diagnosis of the disease which can at times be fatal. Thus, researches are going on to automate these diagnoses through upcoming techniques.

II. RELATED WORK

There are four different states of epileptic seizures. The preictal state appears before the seizure begins, the ictal state begins with the onset of the seizure and ends with an attack, the postictal state that starts after the ictal state. Interictal state starts after the postictal state of 1st seizure and ends before the start of the preictal state of consecutive seizure. By detecting the beginning of the preictal state can help predict seizures.[9].

Usman et al.[9] have predicted epileptic seizure by detecting the initial stage of the preictal state. They used the CHB-MIT dataset that was recorded by placing electrodes on the scalp of subjects to predict epileptic seizures. Empirical mode decomposition was used in this paper. In empirical mode decomposition (EMD)[14], a time-domain signal is broken into several oscillatory functions known as Intrinsic Mode Functions (IMFs). This process of decomposition of a signal into multiple IMFs while remaining in the time domain is comparable with wavelet decomposition and Fourier transform. EMD is a very useful process for analyzing signals that are non-stationary and not linear. After applying the proposed model on the dataset, on an average, they have predicted the duration of epileptic seizures to be 23.48 minutes.

J. Yooet al., in [12] have used machine learning for the early detection of epileptic seizures by monitoring the ictal patterns collected from the EEG waveforms collected from the hospital. According to them the traditional support vector machines (SVM)has better accuracy in prediction than deep learning. They have compared three types of hardware namely Linear SVM, nonlinear SVM, and dual detector Architecture. The LSVM is the most effective among the three with acceptable sensitivity of 87% however when focused on increasing the sensitivity the false positives increases. Hence to maximize the sensitivity, they used NL-SVM however it requires a lot of data training set and has more computational complexity. Thus they used two LSVM, one is trained for high sensitivity while the other is trained for high specificity. This system also helps to reduce the cost of the hardware.

A. Kumar *et al.,in*[1],have used wavelet transformation and SVM for the classification and the diagnosis of epilepsy. They use discrete wavelet transformation (DWT) to analyse the signal in time-frequency domain and to decompose the waves into alpha, beta, theta, gamma and delta waves from the signal. They then extract features like Energy, Variance, Zero Crossing Rate and Fractal Dimension. These features are used in SVM which for binary classification. In this case it is used to classify the test EEG signals into signalswhich have seizures and which do not. This improved memory efficient model showed 98% sensitivity as compared to the other works from the previously existing model. However the complexity of DWT for feature extraction can be time consuming. Moreover SVM does not perform well when the dataset has more noise, that is, target classes are overlapping.

In paper [2], unlike the usual manual feature extraction, here the authors Zhou M, Tian C, Cao R, *et al.*, have used a convoluted neural network (CNN) based model on EEG signals which in turn was used to detect the different stages in the epilepsy. The databases that were used are intracranial Freiburg and scalp CHB-MIT databases. The authors compared the time and frequency based performances in detection of epilepsy. They created a CNN with three layers. They trained and tested for each patient's dataset. Three experiments were conducted which include

i. preictal vs. interictal

ii. ictal vs. interictal and

iii. preictal vs. interictal vs. ictal.

In the frequency domain : The average accuracies for Freiburg database for three experiments were 96.7%, 95.4% and 94.3% and the average accuracies for CHB-MIT database for the three experiments were 95.6%, 97.5% and 93%. In the time domain: The average accuracies for Freiburg database for three experiments were 91.1%, 83.8 and 85.1% and the average accuracies for CHB-MIT database for the three experiments were 59.5%, 62.3% and 47.9%. This showed that the performance of the frequency domain signals were better than that of the time domain signals. But it is time consuming process. Similarly, CNN has a problem of over fitting and it's mostly computationally expensive because it has to take a large database for training [13].In this paper we have used Machine Learning to detect epilepsy using EEG signals. First we have extracted statistical features from the EEG signals. These features can be used as an attributes for the Machine Learning algorithm. Two algorithms namely linear regression model and decision tree model are used to classify the detection of epilepsy. We have used these two models to compare the accuracy when statistical features in the time domain were estimated from raw dataset. In our proposed methodology we have achieved an accuracy of 87.5% and 85% for logistic regression and decision tree model respectively. From literature survey we have noted that decision tree algorithms have not been implemented so far for this application. This paper uses decision tree algorithms to get better and accurate results in the time domain itself.

III. PROPOSED METHODOLOGY

The proposed methodology has been carried out in five different steps as shown in Fig. 1

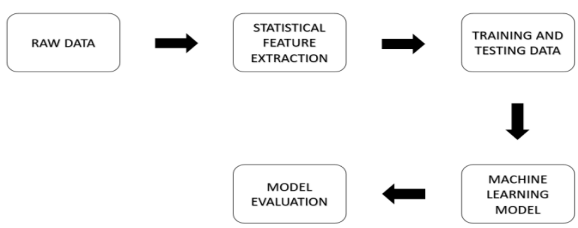

Fig. 1.Block Diagram of proposed methodology

Retrieval Number: 100.1/ijeat.A16851010120
DOI:10.35940/ijeat.A1685.1010120
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication

The steps are described below in detail, respectively.

A. Collection of Raw Data

Even though we have proposed a model to predict whether a person is Epileptic or not, acquiring EEG data through electrodes from the patients are out of the framework of our study. Hence, we have used an online data set available by Phys. Rev. E, 64, 061907 [3]. The EEG signal used has a sampling rate of 173.61 Hz. Set A is a data set is of a Healthy person, with 100 segments, each with 23.6 seconds observation duration. The EEG signal in set A is taken when the person's eyes are in the open state. Set C is the data set for an Epileptic patient. Similar to the normal person data, this data also has a total of 100 segments. The duration of each segment is also 23.6 seconds. The recordings were taken from the hippocampal half-sphere when the person is in the Pre-Seizure state [4]. Each set of readings consist of 4096 samples series in ASCII code [3].

B. Statistical feature extraction

From the raw data the statistical features namely mean, variance, maximum and minimum magnitudes, skewness and kurtosis were extracted. These features explain how the data behaves on an average basis and how the data is spread from the central tendency. The Statistical features are computed using the equations 1 to 6.

$$E(x) = \frac{1}{N}\sum_{i=0}^{N} x_i (1)$$

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - E(x))^2 \quad (2)$$

$$Maximum = \max(x_i) \quad (3)$$

$$Minimum = \min(x_i) \quad (4)$$

$$skewness = \frac{1}{N}\sum_{i=1}^{N}\frac{(x_i - E(x))^3}{\sigma^3} \quad (5)$$

$$kurtosis = \frac{1}{N}\sum_{i=1}^{N}\frac{(x_i - E(x))^4}{\sigma^4} \quad (6)$$

Where E(x) indicates the mean of the data and σ^2indicates the variance of the data.

C. Training of dataset

For selecting the best algorithm for our dataset, we implemented two different machine learning models. The data was split in the ratio of 1:4, where 80% was used for training and 20% was used for testing.

D. Machine learning model implementation

- Logistic Regression

It is one of the machine learning algorithm used for binary classification and is used here to predict if a person is Epileptic or not. In our dataset, we took six different statistical features: mean, variance, maximum, minimum, skewness, kurtosis, and used them to predict the output as 0 or 1, where 0 meant epileptic person whereas 1 meant healthy person. Since we have more than one independent variable; we have used multiple logistic regressions to predict the output. Multiple Logistic regressions can be represented by the following sigmoid equation to predict a value between 0 and 1[22]

$$y = \frac{1}{(1+ e^{-(b_0+b_1x_1+b_2x_2+\cdots+b_nx_n)})} \quad (7)$$

Here, y is the expected probability that the outcome is present, $x_1$through x_n are the distinct independent variables and b_0 through b_n are the regression coefficients. The calculation of regression coefficients is done from the training dataset using Maximum-likelihood estimation. In this estimation, the product of all individual likelihood is taken and the value is noted. Then the plot of these points is shifted and the new individual likelihood of the points on the shifted graph is multiplied to obtain the new likelihood. This process is repeated until the maximum likelihood is achieved. The best coefficients would result in the output value to be close to 1 for the interested class and value close to 0 for the other class.

- Decision Tree Model

In this model, we have used the CART(Classification and Regression Tree)algorithm for binary classification. The input variables and split points on the variable are shown by each root node. The output variables values are stored in the leaf nodes, which is used for the predictions. To make predictions the model starts evaluating from the root node of the decision tree and is filtered through the tree. For this classification problem, we used the Gini Index function as the cost function to determine the most important feature, from where the split begins. The output predicted is then compared to the original values to find accuracy. If the count is not too large the splitting will not be accepted and the node is taken as a final leaf node. Once all the nodes become final leaf nodes then the training stops. This helps the computations to happen faster and in a more efficient way.

E. Model Evaluation

Extensive analysis has been performed to obtain parameters such as Accuracy, Precision, and Recall from the confusion matrix using equations 8 to 10.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+TN+FN}} \quad (8)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (9)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (10)$$

TP = true positive (the epilepsy of a person is predicted correctly)
TN = true negative(Healthy person is predicted correctly)
FP = false positive(Epileptic person is wrongly predicted as healthy)
FN = false negative (Healthy person is wrongly predicted as Epileptic)

IV. RESULT AND DISCUSSION

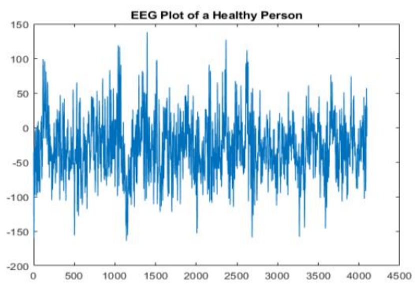

Fig. 2.EEG plot of a healthy person

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication

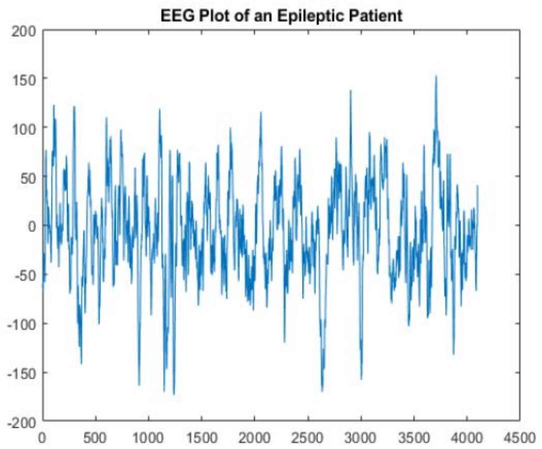

Fig. 3.EEG plot of an Epileptic patient

Table- I: Name of the Table that justify the values

	Decision Tree	Logistic Regression
Accuracy	85%	87.5%
Precision	89.5%	77.27%
Recall	80.9%	100%

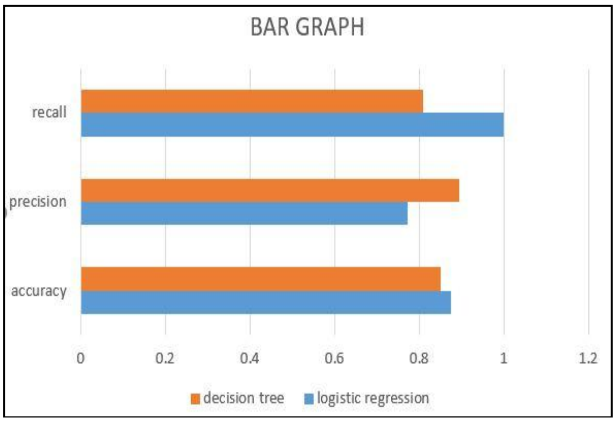

Fig. 4. Bar Graph representation of Performance matrix

On the raw EEG data set the feature extraction was implemented. In this process the statistical features were extracted. The statistical features were given as an input to our machine learning models- Decision Tree and Logistic regression. In the logistic regression model the accuracy was observed to be 87.5% whereas in the decision tree model it was observed to be 85%. While estimating the precision values for both the models, it was noticed that the precision for the decision tree was 89.5% whereas for logistic regression it dropped down to 77.27%. We also calculated the recall value for both the cases and it was 80.9% for the decision tree and 100% for the logistic regression model. The recall value of logistic regression model is 100% whichimplies that out of the total relevant result(77.27%), 100% were classified correctly.

V. CONCLUSION

In our paper, we have used the raw EEG signal and extracted statistical attributes to be sent as inputs in the machine learning model. The model computational time is less also better in terms of both precision and accuracy. In the future these models can be used to find the real time recorded EEG signals. In the logistic regression model, the accuracy was observed to be 87.5% whereas in the decision tree model it was observed to be 85%.From these values, we infer that the decision tree model gives better precision than the logistic regression model but its accuracy is slightly lesser than the latter. This model can be useful for the prediction of Epilepsy, which otherwise can cause extreme damage to the patients.

REFERENCES

1. Kumar and M. H. Kolekar, "Machine learning approach for epileptic seizure detection using wavelet analysis of EEG signals," 2014 International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom), Greater Noida, 2014, pp. 412-416, doi: 10.1109/MedCom.2014.7006043.
2. Zhou M, Tian C, Cao R, et al. Epileptic Seizure Detection Based on EEG Signals and CNN. Front Neuroinform. 2018;12:95. Published 2018 Dec 10. doi:10.3389/fninf.2018.00095
3. Bonn Dataset. Available online: http://www.meb.unibonn.de/epileptologie/science/physik/eegdata.html (accessed on 17 December 2018).
4. Türk Ö, Özerdem MS. Epilepsy Detection by Using Scalogram Based Convolutional Neural Network from EEG Signals. Brain Sci. 2019;9(5):115. Published 2019 May 17. doi:10.3390/brainsci9050115
5. Explorable.com. 2020. Parameters And Statistics. [online] Available at: <https://explorable.com/parameters-and-statistics> [Accessed 21 May 2020]
6. K. ManasviBhat, P. P. Anchalia, S. Yashashree, R. Sanjeetha and A. Kanavalli, "Detection and Prediction of the Preictal State of an Epileptic Seizure using Machine Learning Techniques on EEG Data," 2019 IEEE Bombay Section Signature Conference (IBSSC), Mumbai, India, 2019, pp. 1-5, doi: 10.1109/IBSSC47189.2019.8972992.
7. Zhou, Mengni et al. "Epileptic Seizure Detection Based on EEG Signals and CNN." Frontiers in neuroinformatics vol. 12 95. 10 Dec. 2018, doi:10.3389/fninf.2018.00095
8. Wang Xiashuang, Gong Guanghong, Li Ni, Qiu Shi "Detection Analysis of Epileptic EEG Using a Novel Random Forest Model Combined With Grid Search Optimization" 2019, doi=10.3389/fnhum.2019.00052
9. 9.Syed Muhammad Usman, Muhammad Usman,and Simon Fong, "Epileptic Seizures Prediction Using Machine Learning Methods." Computational and Mathematical Methods in Medicine, vol. 2017, 19 Dec. 2017, pp. 1–10., doi:10.1155/2017/9074759.
10. M. R. N. Kousarrizi, A. A. Ghanbari, M. Teshnehlab, M. A. Shorehdeli and A. Gharaviri, "Feature Extraction and Classification of EEG Signals Using Wavelet Transform, SVM and Artificial Neural Networks for Brain Computer Interfaces," 2009 International Joint Conference on Bioinformatics, Systems Biology and Intelligent Computing, Shanghai, 2009, pp. 352-355, doi: 10.1109/IJCBS.2009.100.
11. Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods. 2004;134(1):9-21. doi:10.1016/j.jneumeth.2003.10.009
12. 12.J. Yoo, "On-chip epilepsy detection: Where machine learning meets patient-specific healthcare," 2017 International SoC Design Conference (ISOCC), Seoul, 2017, pp. 146-147, doi: 10.1109/ISOCC.2017.8368839.
13. Olanrewaju, Oyaniyi. (2019). Re: What are Convolutional Neural Networks (CNN) weakness?.Retrieved from: https://www.researchgate.net/post/What_are_Convolutional_Neural_Networks_CNN_weakness/5d38cf69a4714bae36020ba8/citation/download.

14. N. E. Huang, Z. Shen, S. R. Long et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. In of the Royal Society of London A: mathematical," in Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences, The Royal Society, vol. 454, pp. 903–995, 1998.
15. Fisher, R.S., Acevedo, C., Arzimanoglou, A., Bogacz, A., Cross, J.H., Elger, J.H., Engel Jr., J., Forsgren, L., French, J.A., Glynn, M., Hesdorffer, D.C., Lee, B.I., Mathern, G.W., Moshé, S.L., Perucca, E., Scheffer, I.E., Tomson, T., Watanabe, M., Wiebe, S., 2014. ILAE official report: a practical clinical definition of epilepsy. Epilepsia 55 (4), 475–482.
16. "Facts and Statistics." Epilepsy Society, 8 Jan. 2020, www.epilepsysociety.org.uk/facts-and-statistics#.Xv8j_ZMzY1I.
17. Ngugi AK, Bottomley C, Kleinschmidt I, Sander JW, Newton CR. Estimation of the burden of active and life-time epilepsy: a meta-analytic approach. Epilepsia. 2010;51:883–90.
18. Ali, Ahmer, et al. "Association of Sleep with Sudden Unexpected Death in Epilepsy." Epilepsy & Behavior, vol. 76, 2017, pp. 1–6., doi:10.1016/j.yebeh.2017.08.021.
19. Kiani, R., et al. "Mortality from Sudden Unexpected Death in Epilepsy (SUDEP) in a Cohort of Adults with Intellectual Disability." Journal of Intellectual Disability Research, vol. 58, no. 6, 2013, pp. 508–520., doi:10.1111/jir.12047.
20. John S Duncan, Josemir W Sander, Sanjay M Sisodiya, Matthew C Walker,Adult epilepsy,
21. The Lancet,Volume 367, Issue 9516,2006,
22. Pages 1087-1100,ISSN 0140-6736,
23. https://doi.org/10.1016/S0140-6736(06)68477-8.
24. (http://www.sciencedirect.com/science/article/pii/S0140673606684778)
25. 21. "Epilepsy." World Health Organization, World Health Organization, www.who.int/en/news-room/fact-sheets/detail/epilepsy.
26. 22. Ari, Erkan. (2016). Using Multinomial Logistic Regression to Examine the Relationship Between Children's Work Status and Demographic Characteristics. Research Journal of Politics, Economics and Management. 4. 77-93.
27. 23. "Epilepsy." Mayo Clinic, Mayo Foundation for Medical Education and Research, 5 May 2020, www.mayoclinic.org/diseases-conditions/epilepsy/symptoms-causes/syc-20350093#:~:text=Epilepsy is a central nervous,races, ethnic backgrounds and ages.
28. 24. Türk, Ömer, and Mehmet SiraçÖzerdem. "Epilepsy Detection by Using Scalogram Based Convolutional Neural Network from EEG Signals." Brain sciences vol. 9,5 115. 17 May. 2019, doi:10.3390/brainsci9050115
29. 25. B. S. Zainuddin, Z. Hussain and I. S. Isa, "Alpha and beta EEG brainwave signal classification technique: A conceptual study," 2014 IEEE 10th International Colloquium on Signal Processing and its Applications, Kuala Lumpur, 2014, pp. 233-237, doi: 10.1109/CSPA.2014.6805755.

AUTHORS PROFILE

Anubha Nagar is pursuing Bachelors of Technology in Electronics and Communication Engineering from Vellore Institute of Technology, Her research interests include data science, machine learning, statistics, deep learning and neural networks.

Bidushi is pursuing Bachelors of Technology in Electronics and Communications Engineering from Vellore Institute of Technology, Vellore India. Her research interests include the fields of Machine Learning, Deep Learning and Neural Networks and Statistics.

Mithra Anand Kumar is pursuing Bachelors of Technology in Biomedical Engineering from Vellore Institute of Technology, Vellore, India. Her research interests include data science, machine learning, medical informatics and embedded systems.

Mimangsha Sarma Mimangsha Sarma is pursuing Bachelors of Technology in Biomedical Engineering from Vellore Institute of Technology, Vellore India. Her research include in the field of Bio-signal processing, Biomechanics, Image processing. She had also presented a paper on contemporary trends on lead free composite material for radiation shielding in ICAICTAEE-2019

Dr. Valarmathi J. born in Vellore, Tamil Nadu, 1968. Received B. Tech Electronics from MIT, Anna University in 1992 and completed her M.Tech and PhD from VIT Vellore in 2004 and 2013. Currently working as a Professor in VIT, Vellore. She has 25 years of teaching experience with more than 50 publications in Journals and conferences. She also completed 3 projects funded by Ministry of Defense. Her research interest includes multi-sensor data fusion in radar signals and Lidar signal processing. Currently, focusing the research on the Deep Reinforcement Learning (DRL) in target tracking.