Proceedings of the 8th World Congress on Electrical Engineering and Computer Systems and Sciences (EECSS'22)
Prague, Czech Republic - July 28- 30, 2022
Paper No. ICBES 139
DOI: 10.11159/icbes22.139

Use of Electronic Seizure Diaries and Decision Trees to Predict Seizure Outcome for Patients with Epilepsy

Dominique L. Tanner, MS[1], Michael Privitera, MD[2], MB Rao, PhD[3], Ishita Basu, PhD[3]
[1]University of Cincinnati College of Engineering and Applied Science
2901 Woodside Drive, Cincinnati, OH, USA
tannerde@mail.uc.edu; privitmd@ucmail.uc.edu
[2]University of Cincinnati Gardner Neuroscience Institute
3113 Bellevue Ave, Cincinnati, OH, USA
raomb@ucmail.uc.edu
basuia@ucmail.uc.edu

Abstract **-** Epilepsy is a neurological disorder that causes unpredictable recurrent seizures. Most people with epilepsy dwell in fear of having unpredictable seizures. In attempts to predict future seizure occurrences, investigators have used data from electronic seizure diaries and machine-learning methods, like decision trees. Using individual patient e-diary data, the purpose of this study is to build patient specific decision trees to 1) determine decision trees overall accuracy in predicting seizures and depicting seizure predictors that influence seizure outcome, and 2) identify seizure predictors that have the most influence on seizure outcome. Patients (n=64) were examined, and their e-diary data was used to build patient specific decision trees. Using a 5-point Likert scale, patients e-diaries entailed information on how they rated the probability of experiencing subsequent seizures and rated their mood, predictive symptoms, stress, and seizure counts. Since e-diaries were recorded in the morning and in the evening, seizures for each patient were assessed by half days. R Programming software was used to generate the decision trees and depict seizure predictors that had the most influence on patient's seizure outcome. A confusion matrix was performed to obtain the decision trees performance accuracy. Patients were categorized into groups based on certain seizure predictors that they shared. The results showed that for decision trees overall accuracy in predicting seizures and depicting seizure predictors that influenced seizure outcome, 49% of decision trees had an accuracy of 100%; 37% of decision trees had an accuracy ranging between 90-99%; and 13% of decision trees had an accuracy of <90%. Additionally, the results showed that there were more seizure predictors that had influence on patient's seizure outcome in the morning than in the evening. This work introduces non-invasive precision medicine, with intentions to develop more personalized and reliable health care treatments for people with epilepsy.

Keywords**:** Epilepsy; Seizure Diaries; Seizure Prediction; Decision Trees

1. Introduction

Epilepsy is complex neurological disorder that affects nearly 60 million people worldwide and causes spontaneous recurrent seizures. Those seizures can cause a person to experience transient neurological impairments that vary from brief alterations of awareness to focal movements, to full convulsions [1,2]. One of the most difficult things about living with epilepsy is the unpredictability of seizures. Many people living with epilepsy reside in fear of having unpredictable seizures, which in turn serves as a major disability [2,3]. Seizure unpredictability can lead to interference with school, employment, relationships, and social interactions [2,3]. Seizure prediction could improve the quality of life of people with epilepsy. In predicting the onset of seizures, immediate intervention can be taken to prevent a seizure from occurring. Additionally, a person can take precautions to avoid certain actions that might lead to an injury during a seizure. Recent studies have used electronic seizure diary (e-diary) data and machine learning-based models, like decision trees, to attempt the prediction of seizure occurrence [4]. Most seizure prediction studies have focused on long term electroencephalography (EEG) data either from scalp or intracranial EEG electrode monitoring [3-5].

1.1. An Overview of Continuous EGG Monitoring, Decision Tree Support Systems, and Electronic Seizure Diary Studies for Seizure Prediction

Seizure prediction can be clinically useful for patients with epilepsy. In a study conducted by Cook et al [6], investigators assessed the safety and efficacy of a long-term implanted seizure advisory system. Moreover, they sought to design this system to predict seizure likelihood and quantify seizures in adults with drug-resistant focal seizures [6]. This study was the first to record long-term EEGs in an ambulatory setting in people and successfully demonstrated prospective seizure prediction [6]. Moreover, it showed that intracranial EEG monitoring is feasible in ambulatory patients with drug-resistant epilepsy. Though some intracranial EEG studies have shown success in predicting seizure likelihood, they are invasive for patients with epilepsy. Hence, the idea of using non-invasive methods, along with patient reported data, including patients' self-predicting when patients will experience a seizure, has become a major focus in seizure prediction research.

Using decision trees as a method for seizure prediction, a study conducted by Gifu [7] analysed medical reports on patients with epilepsy from a publicly available dataset called ProTrack and the ProTrack tool [7]. The ProTrack tool was used to classify data for text analysis and data correlation [7]. To construct the decision trees, the investigator used the Iterative Dichotomiser (ID3) algorithm to select the ideal features to create decision trees based on information from the ProTrack dataset. The results from this study provided useful interpretation for which seizure analysis and regiments can be used to treat patients with epilepsy. Another study done by Neamtu et al [8] utilized a decision tree approach that was based on clinical risk factors for new-borns with neonatal encephalopathy and seizures. Additionally, investigators observed new-borns with encephalopathy and seizures in the perinatal period. Etiology and abnormal outcomes were assessed through correlations with the risk factors. The results from the Neamtu et al [8] study showed that decision-tree approaches could provide a first-step tool for the prognosis of the abnormal outcome in new-borns with encephalopathy.

To date, many seizure diary studies have explored how patients with epilepsy self-predict the likelihood of having subsequent seizures within a 24-hour time span. Studies done by Haut et. al. [9-12] used seizure e-diaries to examine seizure self-prediction in patients with epilepsy. These studies collected data about seizure self-prediction, potential seizure precipitants, and predictive symptoms of seizures in the morning and evening. Data on hours of sleep, menstruation, alcohol use and whether patients took their medication were also collected [9-12]. The results from [9-12] study showed that seizure self-prediction was a good indicator for predicting the onset of seizures and was very strong within the first 6 hours of e-diary entry from patients. A similar study conducted by Privitera et. al. [13] also used e-diaries to examine the correlation between self-prediction and seizure occurrence and examined the correlation of factors that induce seizure events, as well as the circadian influences on seizure self-prediction [9-12,13]. Moreover, this study sought to determine the effect of a stress reduction intervention on seizure occurrence. The results from this study showed that the connection between self-prediction and increased risk of seizures within 24 hours was significant [13].

There are different methods to seizure prediction and evaluating factors that influence the outcome of seizures. This research seeks to use individual patient reports and implement a machine learning approach to predict subsequent seizure events and provide valuable insight on features that impact their outcome. Thus, the purpose of this research is to use e-diary data to build patient specific decision trees, establish decision trees overall accuracy in predicting seizures occurrences, and illustrate seizure predictors that induce seizure outcome. Furthermore, this research seeks to assess seizure predictors and their level of influence on seizure outcome that are shared amongst individual patients. The hypothesis of this work asserts that by using optimizing individual e-diary data, it is possible to build decision trees with high precision in predicting seizure occurrences and depict certain seizure predictors that directly affect seizure outcome in patients with epilepsy.

2. Materials and Methodology

2.1. Patient Dataset

For this work, patient-specific e-diary data was obtained from the Stress Management Intervention for Living with Epilepsy Study (SMILE) study, under the supervision of the primary investigator Michael Privitera [13]. A total of 64 patients from the SMILE study dataset were examined. Using a 5-point Likert scale, patients e-diaries entailed information on how they rated the probability of experiencing subsequent seizures and rated their seizure predictors like their mood, premonitory symptoms, stress, and rated their seizure counts. Since e-diaries were recorded in the morning (AM) and in the evening (PM), seizures for each patient were assessed by half days. Each patient's seizure occurrence was classified as a binary variable: 1 = Yes, a patient did experience a seizure; and 0 = No, a patient did not experience a seizure. All seizure predictors from the SMILE study dataset were analysed, each with a corresponding question from the SMILE Protocol [13].

2.2. Equations

A decision tree is a machine learning technique that uses a tree-like model to illustrate decisions and their possible outcomes [14]. Moreover, a decision tree consists of a root node, sub-nodes, and terminal nodes [14]. To calculate the probability of a specific feature that is classified incorrectly when randomly selected, the Gini Impurity method is used [14].

$$\text{Gini} = \sum_{i=1}^{n} p_i^2 \tag{1}$$

n = the total class

P(i) = indicates the probability of an element being classified for a distinct class

$$\text{Gini Impurity} = 1 - \sum_{i=1}^{n} p_i^2 \tag{2}$$

For this work, the SMILE e-diary data [13] was optimized to generate patient-specific decision trees. Information on days when patients were observed was incorporated into the root node of the decision trees. Since the e-diary data is classified as binary, the Gini impurity method was used to split the nodes of the decision trees. Gini Impurity varies between values 0 (indicating that all elements within a node belong to a specified class) and 1 (indicating the random distribution of elements across various classes) [14]. The value of 0.5 of Gini Impurity shows an equal distribution of elements over some classes [14]. To split nodes within a decision tree, Gini Impurity is calculated. The node that has the lowest value of Gini Impurity is selected and is split into a sub-node. This process is repeated until the nodes in the decision trees are homogeneous [14].

2.3. Developing Patient-Specific Decision Trees

The SMILE e-diary data [13] was optimized to generate patient-specific decision trees (Figure 1). R programming software was used to generate decision trees and obtain their accuracy and overall performance. Patients rated their probability of experiencing seizure predictors and future seizure occurrences using a 5-point Likert scale. Information on days when patients were observed was incorporated into the root node of the decision trees. Next, a seizure predictor and a cut point were chosen within the root node and was split into sub-nodes**.** Patients with days of observation where a seizure predictor was greater than or equal to the cut point were positioned into the left sub-node. Patients with days of observation where a seizure predictor was less than a cut point were positioned into the right sub-node. Following, a seizure predictor and a cut point were chosen within the sub-nodes and was split into terminal nodes, which could no longer be split. The terminal nodes in the decision tree show the final predictions of seizure outcome based on the influence of seizure predictors. To split the root node into sub-nodes and terminal nodes, the Gini Impurity was calculated. A confusion matrix was used to test the decision tree’s accuracy and overall performance for predicting seizure outcome based on the influence of certain seizure predictors.

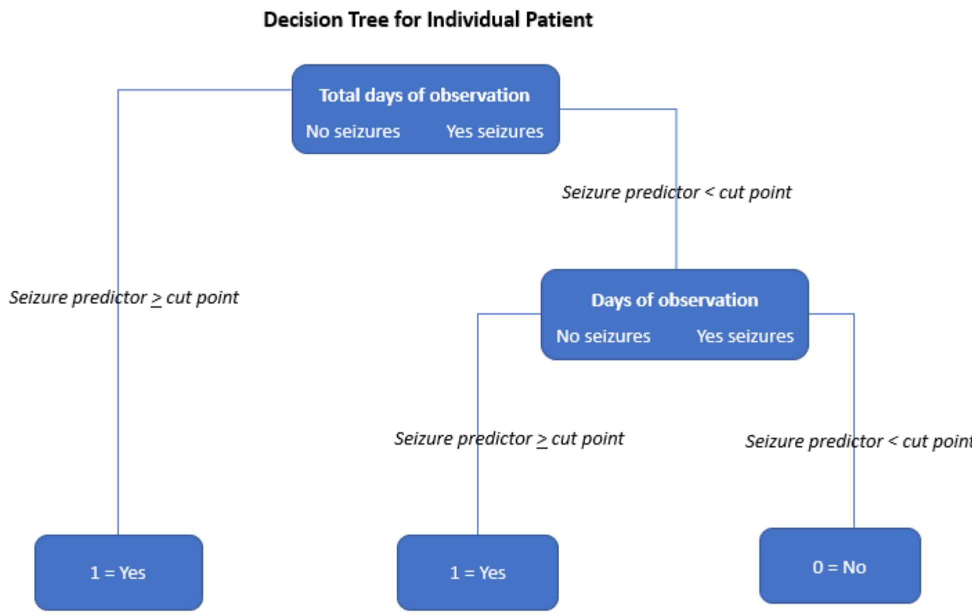

Figure. 1: An example of a patient specific decision tree using the SMILE study dataset [13]. Data from the e-diaries was analysed and used to generate decision trees. The Gini Impurity was calculated to split nodes within the trees. The decision tree shows outcome of seizures and seizure predictors.

3. Results

There were a total of 64 patients with epilepsy from the SMILE study e-dairy data [13]; 42 patients experienced seizures ranging from 4 to 18; 15 patients experienced seizures ranging from 20 to 46; and 5 patients experienced seizures ranging from 60 to 199. Individual e-diaries were recorded in the morning and in the evening, thus, seizure occurrences and seizure predictors were evaluated by half days (i.e., AM and PM). A total of 128 patient-specific decision trees were generated (i.e., 64 decision trees for AM; 64 decision trees for PM).

The decision trees approach was used to obtain the overall accuracy in predicting seizures and depict seizure predictors that affected the outcome of seizures in the morning and evening for patients with epilepsy. For patients with seizures ranging from 4 to 18 that completed their e-diaries in the morning, 31 decision trees had an accuracy of 100%; 11 decision trees had an accuracy varying between 90-99%; and 1 tree had an accuracy less than 90%. For one particular patient, no decision tree was generated for the morning, because all their seizures occurred in the evening. For completing their e-diaries in the evening, 18 decision trees had an accuracy of 100%; 20 decision trees had an accuracy varying between 90-99%; and 2 decision trees had an accuracy less than 90%. For two particular patients, no decision trees were generated for the evening, because all their seizures occurred in morning.

For patients with seizures ranging from 20 to 46 that completed their e-diaries in the morning, 4 decision trees had an accuracy of 100%; 9 decision trees had an accuracy varying between 90-99%; and 2 decision trees had an accuracy less than 90%. For completing their e-diaries in the evening, 2 decision trees had an accuracy of 100%; 4 decision trees had an accuracy varying between 90-99%; and 9 decision trees had an accuracy less than 90%.

For patients with seizures ranging from 60 to 199 that completed their e-diaries in the morning, 2 decision trees had an accuracy of 100%; 1 decision tree had an accuracy varying between 90-99%; and 2 decision trees had an accuracy less than 90%. For completing their e-diaries in the evening, 4 decision trees had an accuracy of 100% and 1 decision tree had an accuracy varying between 90-99%. There were no decision trees that had an accuracy of less than 90%. The prediction accuracies for the decision trees for patients can been seen in Table 1.

Table 1: Patients from the SMILE study dataset [13], the number of seizures occurrences they experienced, and their decision tree prediction accuracies for seizure diaries in the morning (AM) and evening (PM). Seizure occurrences and decision tree prediction accuracies vary among patients.

Patients with 4-18 Seizures	AM Trees with 100% Prediction Accuracy 31	AM Trees with 90-99% Prediction Accuracy 11	AM Trees with < 90% Prediction Accuracy 1	Number of AM Trees Not Produced 1
	PM Trees with 100% Prediction Accuracy 18	PM Trees with 90-99% Prediction Accuracy 20	PM Trees with < 90% Prediction Accuracy 2	Number of PM Trees Not Produced 2
Patients with 20-46 Seizures	AM Trees with 100% Prediction Accuracy 4	AM Trees with 90-99% Prediction Accuracy 9	AM Trees with < 90% Prediction Accuracy 2	Number of AM Trees Not Produced 0
	PM Trees with 100% Prediction Accuracy 2	PM Trees with 90-99% Prediction Accuracy 4	PM Trees with < 90% Prediction Accuracy 9	Number of PM Trees Not Produced 0
Patients with 60-199 Seizures	AM Trees with 100% Prediction Accuracy 2	AM Trees with 90-99% Prediction Accuracy 1	AM Trees with < 90% Prediction Accuracy 2	Number of AM Trees Not Produced 0
	PM Trees with 100% Prediction Accuracy 4	PM Trees with 90-99% Prediction Accuracy 1	PM Trees with < 90% Prediction Accuracy 0	Number of PM Trees Not Produced 0

Patients that experienced certain seizure predictors were categorized into specific groups. Using a 5-point Likert scale, patients from the SMILE study e-diaries [13] provided information on how they rated the probability of experiencing future seizure occurrences, and rated their mood, premonitory symptoms, stress, and seizure counts. For seizure predictors that influenced seizure outcome in the morning (Figure 2A), patients that rated their feelings about unpleasantness/pleasantness, relaxed/stressed, and seizure occurrences since their last seizure diary entry to have the highest influence; feelings about quietness/alertness, depression/excitement, and being worried were the second highest; and feelings about the likelihood of experiencing seizures within the next 24 hours, happiness, sadness, nervousness, tense, focused, thinking about how stressful an event will be if said event would occur, hours of sleep, and cold/flu symptoms were the third highest. Feelings about patients thinking whether a stressful event will occur and experiencing symptoms that may indicate the onset of a seizure had no influence as seizure predictors on seizure outcome.

For seizure predictors that influenced seizure outcome in the evening (Figure 2B), patients that rated their feelings about unpleasantness/pleasantness, sleepiness/alertness, relaxed/stressed, and a seizure occurrence since their last diary entry to have the highest influence; feelings about depression/excitement, happiness, sadness, nervousness, being worried, tense, and how stressful an event was that actually occurred were the second highest; and feelings about the likelihood of experiencing seizures within the next 24 hours, how often patients felt like they were unable to control the important things in their life, how often patients felt that difficulties were piling up so high that they could not overcome them, along with patients taking their medication was the third highest. Feelings about patients being focused, stating if something stressful actually did occur, how often patients felt confident about their ability to handle their personal problems, and how often patients felt that things were going their way had no influence as seizure predictors on seizure outcome.

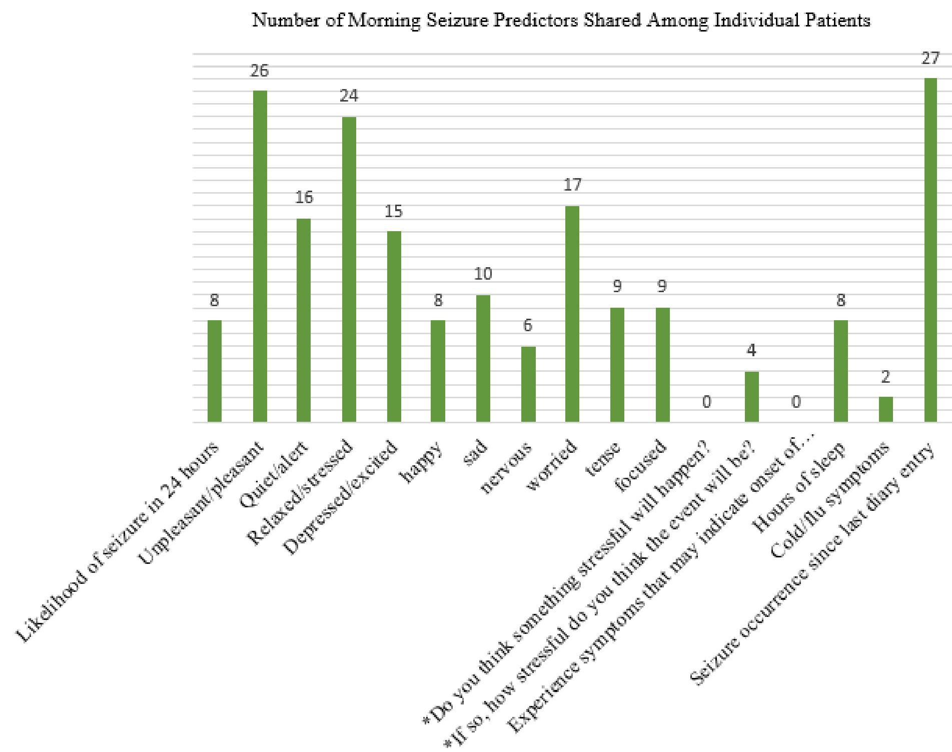

Figure. 2A: Shared seizure predictors that had influence on seizure outcome in the morning among patients from the SMILE study e-dairy dataset [13].

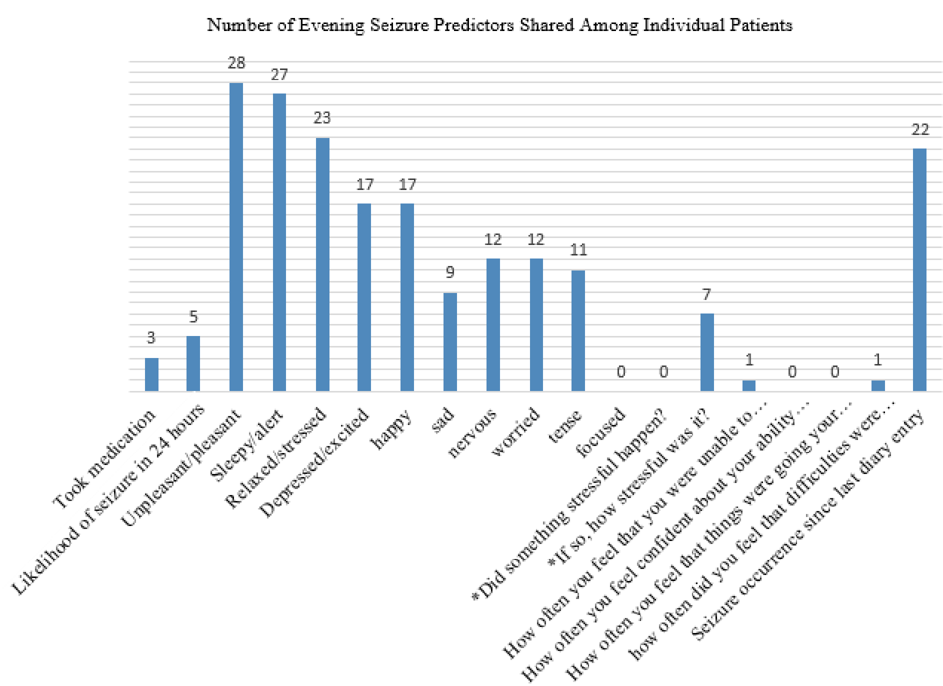

Figure. 2B: Shared seizure predictors that had influence on seizure outcome in the evening among patients from the SMILE study e-dairy dataset [13].

4. Conclusion

Epilepsy is a complex neurological disorder that affects millions of people globally. One of the biggest concerns is not knowing when another seizure will occur. Having e-diary data and utilizing decision trees can serve as major tools in building non-invasive seizure prediction models. There are different factors that influence the outcome of seizures, which may not be the same for many patients. Moreover, having a "one-size-fits-all" approach to epilepsy specific medicine can be limiting due to individual patients and their health being affected differently by epilepsy. This work optimized electronic seizure diaries and decision trees to build an algorithm that could produce personalized decision trees that have high, accurate performance in predicting seizure occurrences and identify seizure predictors that directly impact the outcome of seizures for patients with epilepsy. Additionally, this research introduces personalized medicine, with the intention to develop more precise, predictable, and reliable health care treatments for people who suffer from epilepsy.

Acknowledgements

This is work is based upon data provided by Michael Privitera, MD, who served as the primary investigator for the Stress Management Intervention for Living with Epilepsy (SMILE) Study. Research guidance was provided by Michael Privitera, MD, MB Rao, PhD, and Ishita Basu, PhD. Financial support was provided by the University of Cincinnati Institute for Research in Sensing Fellowship.

References

[1] Z. Zhang, G. Lu, Y. Zhong, Q. Tan, W. Liao, Z. Chen, J. Shi, Y. Liu., "Impaired Perceptual Networks in Temporal Lobe Epilepsy Revealed by Resting fMRI," *Journal of Neurology*, vol 256, pp. 1705-171, 2009.

[2] Center for Disease Control and Prevention (2015). Epilepsy Data and Statistics. [Online]. Available: https://www.cdc.gov/epilepsy/data/index.html.

[3] T. Dissanayake, T. Fernando, S. Denman, S. Sridharan and C. Fookes., "Deep Learning for Patient-Independent Epileptic Seizure Prediction Using Scalp EEG Signals," *IEEE Sensors Journal*, vol. 21, no. 7, pp. 9377-9388, 2021.

[4] K. Das, D. Daschakladar, P. P. Roy, A. Chatterjee, S. P. Saha., "Epileptic seizure prediction by the detection of seizure waveform from the pre-ictal phase of EEG signal", *Biomedical Signal Processing and Control*, vol. 57, 2020.

[5] K. M. Tsiouris, V. C. Pezoulas, D. D. Koutsouris, M. Zervakis and D. I. Fotiadis., "Discrimination of Preictal and Interictal Brain States from Long-Term EEG Data", *IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS)*, pp. 318-323, 2017.

[6] M. J. Cook, T. J. O'Brien, S. F. Berkovic, M. Murphy, A. Morokoff, G. Fabinyi, W. D'Souza, R. Yerra, J. Archer, L. Litewka, S. Hosking, P. Lightfoot, V. Ruedebusch, W. D. Sheffield, D. Snyder, K. Leyde, D. Himes., "Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: a first-in-man study", *Lancet Neurol*. pp. 563-71, 2013.

[7] D. Gifu, "The Use of Decision Trees for Analysis of the Epilepsy," *Procedia Computer Science,* vol 192, pp. 2844-2853, 2021.

[8] B. M. Neamțu, G. Visa, I. Maniu, M. L. Ognean, R. Pérez-Elvira, A. Dragomir, M. Agudo, C. R. Șofariu, M. Gheonea, A. Pitic, R. Brad, C. Matei, M. Teodoru, C. Băcilă., "A Decision-Tree Approach to Assist in Forecasting the Outcomes of the Neonatal Brain Injury", *Int. J. Environ. Res. Public Health,* vol 19, no. 9, pp. 1-19, 2021.

[9] S. R. Haut, C. B. Hall, T. Borkowski, H. Tennen, R. B. Lipton., "Modeling seizure self-prediction: an e-diary study". Epilepsia. vol 54(11), pp 1960-7, 2013.

[10] S. R. Haut, C. B. Hall, T. Borkowski, H. Tennen, R. B. Lipton., "Clinical features of the preictal state: Mood changes and premonitory symptoms." *Epilepsy Behav*., vol 23, pp 415–421, 2012.

[11] S. R. Haut, C. B. Hall, A. J. LeValley, R. B. Lipton., "Can patients with epilepsy predict their seizures?" *Neurology,* vol 68, pp 262–266, 2007a.

[12] S. R. Haut, C. B. Hall, J. Masur, R. B. Lipton., "Seizure occurrence: precipitants and prediction". *Neurology.* Vol 69(20), pp 1905–1910, 2007b.

[13] M. Privitera, S. R. Haut, R. B. Lipton, J. S. McGinley, S. Cornes., "Seizure self-prediction in a randomized controlled trial of stress management", *Neurology.*, vol 93(22), pp e2021-e2031, 2019.

[14] F. A. Farris, "The Gini Index and Measures of Inequality", *The American Mathematical Monthly*, vol 117, no. 10, pp. 851-864, 2010.