Avestia Publishing
Journal of Biomedical Engineering and Biosciences (JBEB)
Volume 9, Year 2022
ISSN: 2564-4998
DOI: 10.11159/jbeb.2022.007

Decision Trees as a Method for Forecasting Seizure Precipitants and Identifying Their Influences on Seizure Outcome

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
[3]University of Cincinnati College of Medicine
3230 Eden Avenue, Cincinnati, OH, USA
raomb@ucmail.uc.edu
basuia@ucmail.uc.edu

Abstract - *Epilepsy is a complex disease that causes unpredictable seizures, which can lead to severe neurological impairments. Not knowing when a seizure will occur, many people with epilepsy often experience feelings such as anxiety, fear, and stress. In an effort to predict when seizures might occur, investigators have used data from patients' electronic seizure diaries, as well as machine-learning methods, like decision trees. The objective of this work is to create patient-specific decision trees to 1) forecast seizure occurrence and identify seizure precipitants that influence seizure occurrences, and 2) determine seizure precipitants' level of influence on seizure occurrences. Patients' (n=64) seizure diaries were examined individually. Diaries contained data on how patients rated mood, predictive symptoms, stress, seizure occurrences, and seizure likelihood using a 5-point Likert scale. Diaries were recorded in the morning and in the evening, thereby evaluating seizures by half days. R Programming software was used for data analysis and decision tree development, and a confusion matrix was used for predictive accuracy. Results showed that precipitants' influence on patient's seizure outcome was greater in the morning than in the evening. Patients were also categorized in groups based on shared seizure precipitants. This work introduced non-invasive, personalized healthcare regimen for people with epilepsy.*

***Keywords*:** Epilepsy, Seizure diaries, Seizure predictions, Machine learning, Decision trees.

1. Introduction

Epilepsy is a complex disease that causes spontaneous, persistent seizures, and for many, the uncertainty of seizure occurrences often disrupts activities of daily living. These seizures can cause people to experience neurological, physiological, and cognitive impairments [1,2]. Though it has gained wide recognition from many researchers and medical professionals, epileptic seizure prediction remains a challenge [2,3]. Over the years there have been significant advancements in seizure prediction research. Most seizure prediction studies have focused on long term electroencephalography (EEG) data from intracranial EEG electrode monitoring [3,4]. Though continuous EEG monitoring have shown positive results for seizure predictive abilities, they are not feasible for many patients with epilepsy [4,5]. Continuous EEG recordings with intracranial electrodes are invasive and can be uncomfortable, time consuming, expensive, and have some potential risk for patients. [6].

Date Received: 2022-10-20
Date Accepted: 2022-11-18
Date Published: 2022-12-08

The idea of patients self-predicting when they will experience subsequent seizures has rapidly become a key focus in seizure prediction research [6]. To date, many electronic seizure diary (e-diary) studies investigate how patients with epilepsy self-predict the likelihood of having seizures within a 24-hour time span [7-11]. To further understand how to characterize the pre-ictal phase of seizures, researchers have concentrated on the feasibility of self-prediction by providing an inventory of possible seizure triggers, predictive symptoms (i.e., circadian rhythms, hours of sleep, experience of stress), and measurements of mood [8-10]. By assessing patients' mood, as well as the correlation between self-prediction, seizure occurrence, and predictive symptoms, many e-diary studies have demonstrated success in displaying associations between self-prediction and increased risk of seizures [10,11]. Furthermore, these studies have shown that patients with epilepsy have the capacity to self-predict when they will experience subsequent seizures [7-11].

Decision trees are a machine learning tool that, upon evaluating data, uses tree-like models to illustrate choices and their probable outcomes [12]. In utilizing decision trees to predict seizures, one study [13] analysed patients with epilepsy medical reports from publicly available datasets. These tools were used to classify data and generate decision trees based on ideal features associated with epileptic seizures [13]. Results from this study presented useful clarification for which seizure analysis and regimen could be applied to treat patients with epilepsy. In another study [14], investigators generated decision trees based on clinical risk factors for new-borns with neonatal encephalopathy and seizures. Investigators also assessed the etiology and abnormal outcomes that were associated with said risk factors. The results from this study [14] study revealed that implementing decision-tree could serve as a major tool for the prognosis of the abnormal outcome in new-borns with encephalopathy and seizures.

Notwithstanding the various methods for seizure prediction in previous studies, the purpose of this study is to 1) optimize patient-specific e-diary data, 2) employ a decision tree approach to forecast seizures in advance, and 3) identify seizure precipitants that influence seizure outcome in patients with epilepsy. Additionally, this work seeks to categorize patients into groups based in certain seizure precipitants that are commonly shared. The hypothesis for this study affirms that by using data from individual patient reports, it is possible to create decision trees with significant accuracy in forecasting seizure precipitants that that directly affect seizure outcome in patients with epilepsy.

2. Materials and Methodology

2.1. Patient Data

E-diary data for this study was obtained from the Stress Management Intervention for Living with Epilepsy Study (SMILE) study [11], which was conducted between University of Cincinnati, Ohio, Montefiore Medical Center, Bronx, New York, and University of California San Francisco. From the dataset, there were 64 patients that were examined; using a 5-point Likert scale, patients rated the likelihood of experiencing seizures. A visual analogue scale was used to record early signs of an incoming seizure (i.e., mood, predictive symptoms, stress). Additionally, patients rated the number of times they experienced seizures (i.e., seizure counts). Patients e-diaries were recorded in the morning (AM) and in the evening (PM), thus, seizures were assessed by half days. One particular patient did not record data in their e-diaries; thus, they were removed from the patient dataset.

2.2. Seizure Classification

Patients' seizure counts were classified as a binary variable: 1 = Yes, a patient did experience a seizure; and 0 = No, a patient did not experience a seizure.

2.3. Decision Tree, Gini Impurity

R programming software (version 4.2.1) was used to evaluate e-diary data and generate patient-specific decision trees; each tree consisted of a root node, sub-nodes, and terminal nodes [15]. Given that seizure counts were classified as binary, the Gini Impurity method was used to split the nodes within the trees [15,16]:

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

where n represents the total class and P_i represents the probability of a feature being classified for a specific class.

Gini Impurity is a measure of how chaotic the distribution of the classes is in the outcome [15]; 0 indicates that all elements within a node belong to a specified class, 0.5 indicates equal distribution of elements over some classes [15,16]. The goal for Gini Impurity is to minimize impurity for splitting the nodes (i.e., selecting a seizure precipitant and cut point for a

Gini Impurity that is minimum). Nodes that had the lowest value of Gini Impurity were selected and split; this process was repeated until nodes within the decision trees were homogeneous.

2.4. Generating Patient-Specific Decision Trees

Given that seizure counts were evaluated by half days, decision trees were generated for the morning and evening. Illustrated in figure 1, information about patients' observation period during the SMILE study [11] were placed in the root node of the decision tree. From the root node, a seizure predictor and a cut point were selected, thereby splitting the root node into sub-nodes; observed days where precipitants were greater than or equal to the cut point were positioned into right sub-nodes. Consequently, observed days where precipitants were less than a cut point were positioned into left sub-nodes. Following the pruning conventions, this splitting process was repeated until sub-nodes split into terminal nodes (which could no longer be split). Each terminal node was identified with a majority outcome.

For patients, decision trees forecasted certain precipitants that influenced seizure onset based on patients' scores in their diaries from scale and mood circumplex and circadian patterns.

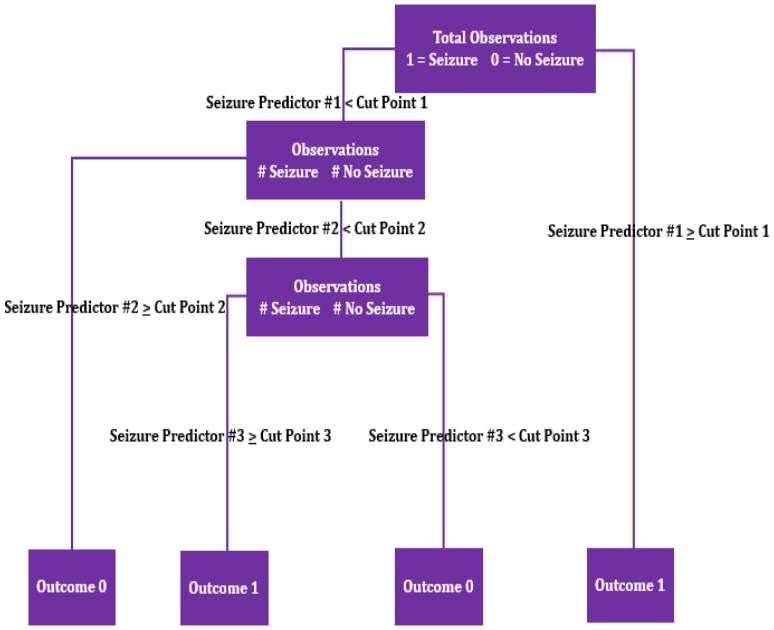

Figure 1: An example of a patient-specific decision tree. Data about patients and their seizure precipitants from e-diaries were analysed. Decision trees were generated to depict certain precipitants that were associated with seizure outcome. The Gini Impurity approach was used to split nodes within the trees.

3. Statistical Analysis

A confusion matrix was used for predictive accuracy. To test and train the forecasting model, patients' data was split; 75% of data was used to train the decision tree while the remaining 25% was used for testing. To determine decision trees' overall performance for forecasting seizure precipitants that influenced seizure outcome, accuracy, sensitivity and specificity, the misclassification rate (i.e., error rate) were calculated independently for morning and evening e-diaries. Error rate was calculated by obtaining number of all incorrect predictions (false positives + false negatives) and dividing them by the total number of the dataset (positives + negatives):

$$Error\ Rate = \frac{FP + FN}{TP + TN + FP + FN} = \frac{FP + FN}{P + N}$$

For this work, error rate was calculated in efforts to fully understand how well the forecasting models were performing in forecasting and to determine whether improvement was needed.

4. Results

All patients from the SMILE study [11] and data from their e-diaries were examined separately. The average number of days of diary recordings that were analysed for patients was calculated – patients had an average of 55.45 days of morning recordings and 51.72 days for evening recordings.

For the observation periods (i.e., both AM and PM), 41 patients had seizures ranging from 4 to 18; 15 patients with seizures ranging from 20 to 46; and 5 patients with seizures ranging from 60 to 199. E-diaries were recorded in the morning and in the evening, meaning that seizure counts and precipitants were assessed by half days. There were 122 patient-specific decision trees that were generated (61 trees for the morning and 61 trees for the evening). For two patients (129 and 308), morning decision trees could not be generated, because all their seizure occurrences happened in the evening. Consequently, for two other patients (101 and 511), evening decision trees could not be generated, because all their seizure occurrences happened in the morning.

Out of 63 patients, 34 had feasible predictive analyses where tree performances, sensitivity and specificity, and misclassification rates were successfully obtained (15 patients from the morning and 19 from the evening). Out of the 34 patients, 25 had sensitivities and

specificities that were greater than 50% (indicating that the model seemed to be helpful in forecasting seizures); 8 patients had sensitivities that were greater than 50% and specificities less than 50% (indicating that the model seemed to be moderately helpful), and 1 patient's sensitivity was less than 50% and specificity greater than 50%.

For the remaining 29 patients, though tree performance and misclassification rates were obtained, sensitivities and/or specificities were calculated as 0% (indicating that the model was not helpful in forecasting seizures). Therefore, the remaining patients' analyses were not included in this study. Though some statistical analyses for some patients were low, this study was aiming for high sensitivities as a means to not overlook forecasting seizures. However, it is noted that a low specificity is tolerable, as this means there may be more false alarms.

4.1. Morning Seizure Diaries and Decision Trees

Along with total seizure counts, the number of seizure counts for the morning, and observation periods were evaluated (table 1). Tree accuracies, sensitivities and specificities, and error rates varied (table 2).

Table 1: Fifteen patients from morning e-diary dataset, along with their total seizure counts, seizure counts specifically for the morning, and days of observation.

Morning Seizure Diary Data			
Patient	Total Seizures	Only AM Seizures	Total AM Observation (days)
101	8	8	56
102	30	14	88
104	44	26	55
110	33	21	53
116	9	6	57
119	14	4	47
120	33	14	55
122	71	21	56
128	36	19	53
133	31	24	40
340	29	24	38
341	55	12	48
352	20	10	57
505	14	9	41
511	7	7	52

Table 2: Morning patient-specific decision tree performances. Fifteen patients had successful independent performance of trees. Sensitivity and specificity, and error rates were also calculated.

Morning Seizure Diary Data Statistical Analyses				
Patient	Tree Accuracy	Sens.	Spec.	Error Rate
101	98.21%	87.50%	100%	17.86%
102	86.36%	100%	25%	9%
104	75%	100%	50%	11.50%
110	100%	100%	100%	11.10%
116	92.98%	83.33%	94.11%	5%
119	93.61%	100%	93%	6%
120	71.43%	90%	25%	9%
122	50%	66.67%	20%	14%
128	73.58%	78.94%	70.58%	22%
133	50%	50%	50%	10%
340	77.77%	100%	66.66%	16.60%
341	75%	50%	89%	16%
352	85.70%	83.30%	100%	8.70%
505	40%	37.50%	50%	9%
511	88.46%	57.14%	93.33%	5%

4.1.1 Morning Seizure Precipitants

Specific seizure precipitants were classified into groups based on how often they were shared amongst patients' decision trees. As shown in table 3, precipitants on pleasantness, being relaxed and/or stressed, and seizure occurrences since last e-diary entries had the highest influence on seizure outcome. Precipitants on quietness and/or alertness, feeling depressed and/or excited, and worried had moderate influence. Precipitants on the probability of having a seizure within the next half-day, feeling happy, sad, nervous, tense, focused, considering the idea of how stressful an event would be if one occurred, hours of sleep, and experiencing symptoms that may be associated to a cold/flu had low influence. Contemplating the idea of a stressful event occurring and experiencing premonitory symptoms that may indicate seizure onset were not considered precipitants for patients.

Table 3. Morning seizure precipitants, predictive symptoms, and measurements of mood that were depicted in decision trees. The first column are precipitants identified from trees; the second column are the number of patients out of 63 who experienced each precipitant.

Seizure Precipitants Identified	Number of Patients That Experienced Precipitants
Seizure occurrence since last e-diary entry	27
Unpleasantness/pleasantness	26
Relaxed/stress	24
Worried	17
Quietness/alertness	16
Depression/excitement	15
Sadness	10
Tense	9
Focused	9
Probably of seizure in 24 hrs	8
Happiness	8
Hours of sleep	8
Nervousness	6
Thinking how stressful the evet will be if it occurred	4
Symptoms of cold/flu	2
Thinking that a stressful event will occur	0
Premonitory symptoms that may indicate seizure onset	0

4.2. Evening Seizure Diaries and Decision Trees

A total of 19 patients had feasible statistical analyses. Total seizure counts, the number of seizure counts for the evening, and observation periods were evaluated (table 4); tree accuracies, sensitivities and specificities, and error rates varied (table 5).

Table 4: Nineteen patients from evening e-diary dataset, along with their total seizure counts, seizure counts specifically for the evening, and days of observation.

Evening Seizure Diary Data			
Patient	Total Seizures	Only PM Seizures	Total AM Observation (days)
102	30	16	83
104	44	18	51
105	10	9	57
108	9	8	45
110	33	12	48
119	14	10	53
120	33	19	56
125	11	10	56
126	22	14	54
128	36	17	56
312	26	17	52
317	23	17	5
321	34	30	76
323	11	7	54
337	17	12	53
345	21	15	55
351	35	29	57
352	20	10	58
354	12	8	62

Table 5: Evening patient-specific decision tree performances. Nineteen patients had successful independent performance of trees. Sensitivity and specificity, and error rates were also calculated.

Evening Seizure Diary Data Statistical Analyses				
Patient	Tree Accuracy	Sens.	Spec.	Error Rate
102	71.43%	82.35%	25%	9%
104	41.67%	50%	25%	13%
105	85.71%	66.66%	93.75%	10%
108	84.44%	62.50%	89.18%	8%
110	91.60%	100%	66%	12.70%
119	83%	50%	90.69%	9%
120	64.29%	77.78%	40%	25%
125	92.80%	100%	50%	5.30%
126	78.57%	100%	25%	14%
128	57.10%	60%	50%	21.80%
312	83%	75%	100%	14%
317	64.29%	80%	25%	10%
321	50%	50%	50%	15%
323	98.14%	100%	97.87%	1%
337	84.90%	75%	87.80%	15%
345	74.54%	46.66%	85%	16%
351	78.94%	93.10%	64.28%	15%
352	91.37%	70%	95.83%	5%
354	88.70%	50%	94.44%	11%

4.2.1 Evening Seizure Precipitants

Shown in table 6, precipitants on pleasantness, feeling sleepy and/or alert, being relaxed and/or stressed, and seizure occurrences since last e-diary entries had the highest influence on seizure outcome. Precipitants on feeling depressed and/or excited, happy,

sad, nervous, worried, tense, and recounting how stressful an event was had moderate influence. Precipitants on the probability of having a seizure within the next half-day, feeling unable to control the important things in life, difficulties were increasing to where they could not be overcome, and medication compliance had low influence.

Though feelings around being focused was considered a seizure predictor for some patients in the morning, it was not considered a predictor for patients in the evening. Patients stating whether a stressful occurred, felt confident about handling personal problems, and felt that things were going their way were also not considered as precipitants and had no influence on seizure outcome.

Table 6. Evening seizure precipitants, predictive symptoms, and measurements of mood that were depicted in decision trees. The first column are precipitants identified from trees; the second column are the number of patients out of 63 who experienced each precipitant.

Seizure Precipitants Identified	Number of Patients That Experienced Precipitants
Unpleasantness/pleasantness	28
Sleepiness/alertness	27
Relaxed/stressed	23
Seizure occurrence since last e-diary entry	22
Depression/excitement	17
Happiness	17
Worried	12
Nervousness	12
Tense	11
Sadness	9
What was stressful event that occurred	7
Probably of seizure in 24 hrs	5
Medication compliance	3
Feeling unable to control important things	1
Feeling that difficult things were piling up	1
Focused	0
Stating whether a stressful event occurred	0
Feeling confident about handling their problems	0
Feeling that things are going patients' way	0

4.3. Limitations

From the morning and evening dataset, tree accuracies, sensitivity and specificity, and error rates were obtained for all 63 patients. However, 29 of these patients had low sensitivities and/or specificities (0%). Since having a test with low sensitivity and specificity is not ideal (as it may provide a high number of false positives), the statistical analyses for these particular patients were not deemed feasible.

5. Discussion

Seizure diaries and the decision tree approach was successfully used to identify seizure precipitants that were correlated with seizure onset for individual patients. The level of influence that these precipitants had on outcome was also identified. Based on how patients' responses from the Likert scale, mood circumplex, circadian patterns, and recorded seizure counts, all decision trees were different.

Decision trees provided valuable insight regarding patients' experiencing subsequent seizure, as well as probable seizure triggers, predictive symptoms, and mood. This study showed that by optimizing medical reports and machine learning techniques, creating personalized regimen for patients with epilepsy is ideal.

This study seeks to expand its work by further refining precision medicine and develop more dependable epilepsy-based healthcare treatments. Moreover, this work seeks to transform the way clinicians, who specialize in epilepsy-related medicine, understand how people respond to patient-specific regimen by 1) shifting the emphasis in epilepsy from reaction to prevention, 2) forecast potential seizure onsets in advance, 3) customize seizure-prevention strategies, and 4) improve health outcomes and quality of life.

6. Conclusion

People with epilepsy report the unpredictability of seizures as a major impediment to quality of life. Quantitative and machine learning strategies as applied in this study can help people with epilepsy understand the specific precipitants that may lead to seizures. Individualization of these precipitating factors can give an individual with epilepsy more confidence about high risk and lower risk days for a seizure. This knowledge could affect decisions on activities, decisions on medication dosing or timing, and in some cases decisions on closer monitoring by caregivers during times of

highest seizure risk. Overall, this study created innovative multilevel approaches to building personalized action plans for people living with epilepsy.

Acknowledgements

This is work is based upon electronic seizure diary data that was provided by Michael Privitera, MD, and Sheryl Haut, MD who served as the primary investigators for the Stress Management Intervention for Living with Epilepsy (SMILE) Study. Financial support to carry out this study was offered 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] 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.

[8] 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.

[9] 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.

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

[11] 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.

[12] Patel N, Upadhyay S., "Study of various decision tree pruning methods with their empirical comparison". WEKA. Int J Comp Appl; vol 60, no 12, pp 20-25, 2012

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

[14] 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.

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

[16] D'Ambrosio A., Tutore V.A. "Conditional Classification Trees by Weighting the Gini Impurity Measure"., Ingrassia S., Rocci R., Vichi M. (eds) New Perspectives in Statistical Modeling and Data Analysis. Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Berlin, Heidelberg, 2011.