EasyChair Preprint

№ 1975

Epileptic Seizure Prediction Using Data Mining Algorithms.

Ilyas Zidane, Jamal Mhamdi and Nordine Zidane

November 18, 2019

Epileptic Seizure Prediction Using Data Mining Algorithms.

Zidane I, Mhamdi J, Zidane N
Mohammed V University
Electronic Systems Sensors and Nanobiotechnology
ENSET Rabat Morocco
Email:zidane.ilyas@yahoo.fr

Abstract— Epileptic seizures are caused by sudden electric discharges in neurons irregular transient operation; hence this generate cerebral neurological dysfunction. The seizures can occur in different regions of the brain and are reflected in various clinical expressions. As it happens, depending on the location and job of the affected neuron, seizures with different types occur: simple partial seizures (SPS), complex partial seizures (CPS), secondarily generalized seizures(SGS) with epilepsy locations classified in temporal lobe epilepsy (TLE), frontal lobe epilepsy (FLE), posterior epilepsy (PE including the parietal, occipital, occipito-temporal, temporo-occipito-parietal junction regions), operculo-insular epilepsy (OIE) and multifocal epilepsy (MFE).

In this paper, we propose a new approach of epilepsy seizure prediction by using data mining algorithms to analyze seizures of a huge number of patients EEGs. This method is most powerful and most accurate than the others prediction methods.

Keywords—Epilepsy, Seizure, EEG, Data mining, Supervised learning, Algorithms.

I. INTRODUCTION

Epilepsy is the most common neurological disease after migraine. It has been reported to affect 0.5% to 1% of the world's population, namely 50 million people in the world. It is estimated that about 430 000 people are affected in France, with 33 000 new cases each year. Nearly 5% of persons are experienced a seizure in their life time and 1% of persons suffer of epileptic seizure [1].

The statistics shows that Epilepsy increases the death risks. Although medicine generate many antiepileptic drugs, but they are neither curative and nor efficient to reduce seizures. Some patients are antiepileptic drugs resistant; in this case the solutions are surgical treatment to remove the epileptogenic area or cerebral electrical stimulation.

Epilepsy disease affects all ages, most affected are children and old persons, thus 75 % of diseases are established before 18th birthday. We analyze a hundred epileptics patients of all categories of age: Children (5-14 years), Youth (15-24 years), Adults (25-64 years) and Seniors (65 years and over). In babies, seizures might not be obvious to an observer.

Many methods have been developed for seizure prediction over several years since 1995. The most of these methods used state similarity features for seizure prediction[2],also they employed independent component analysis to separate isolated epileptiform discharges from EEG background and analyse magnetoencephalography data to localize epileptic spikes using independent components analysis and spatio-temporal clustering [3],they used EEG spatiotemporal correlation structure with delay correlation and covariance matrices [4], they applied autoregressive modeling and least-squares parameter estimator [5], they studied effects of linear univariate features [6], Classification of patterns of EEG synchronization with bivariate features [7], they used spectral power from raw and bipolar time-differential signals [8] and combination of mean phase coherence and the dynamic similarity index methods [9] [10] , they used cost-sensitive support vector machines in spike rate [11] and they employed Fourier transformation to perform signal processing of epileptic EEG [12].

II. DATA SET DESCRIPTION

The data set is issued from EEGs recordings of epileptic's patients. The information in large clinical EEGs database has been entered during two years because of the huge number of patients and their time's unavailability.

We choose only hundred patients to study the prediction of their healing according to the result of treatment using data mining algorithms so to predict life threatening caused by epileptic seizure and may be able to analyze signals in real-time and having a dashboard to generate alerts before the time of epileptic seizure.

The sampling frequency is 256 Hz with 128 channels of epileptic EEG signals which are recorded by focal electrodes form brain locations: temporal lobe epilepsy (TLE), frontal lobe epilepsy (FLE), posterior epilepsy (PE including the parietal, occipital, occipito-temporal, temporo-occipito-parietal junction regions), operculo-insular epilepsy (OIE) and multifocal epilepsy (MFE).The records may be taken before the seizure (Preictal), during the seizure(ictal), between seizures (interictal) and after the seizure(postictal). Preictal or prodromeis the time before the seizure. It can last from minutes to days and make patient act and feel differently, for example, duration of ictal varies from a few seconds to 5 minutes. Therefore, ictal records during epileptic seizures can have at least 60 minutes of preictal signal preceding each seizure.

TABLE 1. GLOBAL DATASET

N	Age	Sex	type	Brain Location	Seiz.	H
1	Adults	Female	SPS	Temporal	5	yes
2	Adults	Female	SPS	Temporal	3	yes
3	Adults	Male	SPS	Temporal/Occipital	5	yes
4	Adults	Female	CPS	Temporal	4	yes
5	Adults	Female	SPS	Frontal	5	no
6	Seniors	Male	SPS	Temporal/Parietal	4	yes
7	Adults	Female	SPS	Temporal	5	yes
8	Adults	Female	CPS	Temporal	4	yes
9	Adults	Female	SPS	Frontal/Temporal	4	yes
10	Children	Female	SPS	Frontal	4	yes
95	Adults	Male	SPS	Frontal	5	no
96	Adults	Male	SPS	Temporal/Occipital	3	yes
97	Adults	Female	SGS	Temporal	5	yes
98	Adults	Female	SPS	Parietal	4	no
99	Adults	Female	SPS	Temporal/Occipital	5	no
100	Youth	Female	SPS	Temporal	2	yes

We have use the flowing fields of dataset table 1: Patient number(N), Age , Sex, type, Brain location , Total seizures(seiz) and Healing(H) .

III. DECISION TREE WITH FEW DATASET

In this paper our method is to use an ameliorated supervised learning algorithm in order to build a decision tree which shows clearly the prediction patient's response to treatment with a great precision better than other methods.

The problem is to find a function which would allow to represent the prediction?

The function would be minimum when the node is pure (all data are in the same class) and maximum when the data is equally distributed.

Two functions are often used in binary decision trees:

$$Entropy(p) = -\sum_{k=1}^{c} P(k/p) \times \log(P(k/p))$$

$$Gini(p) = 1 - \sum_{k=1}^{c} P(k/p)^2$$

$$= 2 \times \sum_{k<k'}^{c} P(k/p) \times P(k'/p)$$

Also, we need a function to choose a test:

$$Gain(p, test) = i(p) - \sum_{j=1}^{n} P_j \times i(p_j)$$

p : position test : test with n parameter

P_j : part of sample items in p position moving to p_j position.

$i(p)$: Entropy (p) or Gini (p)

It's obvious to remind some definition: Entropy is information's measurement (or rather lack thereof). The information gain is calculated by making a split. Gini measurement is the probability of a random sample being classified incorrectly if we randomly pick a label according to the distribution in a branch.

A decision tree contains a root node and the other nodes are growing upside down as new nodes by splitting the data into levels. A node of any level could have a node as parent and others as children nodes.

Generic algorithm for building a tree:
input: sample
begin
  Initialize to the empty tree
the root is the current node
repeat
  Decide if the current node is terminal
  if the node is terminal then
    Assign a class
  else
    Select a test and create the subtree
  end if
  Skip to the undrawn tracking node if there is one
until you get a decision tree
end

Table 2 below is a random extraction of 8 records sample from databases to apply Iterative Dichotomizer algorithm in order to generate decision tree.

TABLE 2. PARTIAL DATASET

N	A	S	T	L	H
40	Youth	Male	SPS	Temporal	yes
41	Youth	Female	SGS	Frontal	no
57	Adults	Female	CPS	Frontal	no
70	Youth	Male	CPS	Frontal	yes
75	Children	Male	SPS	Parietal	yes
80	Adults	Male	SGS	Parietal	no
82	Adults	Male	SPS	Parietal	no
85	Youth	Female	CPS	Temporal	no

Seizure Type : T = {SPS, CPS, CGS}

Age Range : A = {Children, Youth, Adults}

Brain Location : L = {Temporal, Frontal, Parietal}

SEX : S = {Female, Male}

Healing : H = {yes, no}

We want to build a decision tree that is able to determine if a patient is healing and therefore treatment responsive according to the attributes: A, S, T and L.

Tree root (no test)

Label (3,5) for : (L(Є,yes), L(Є,no)).

What is the first test to perform?

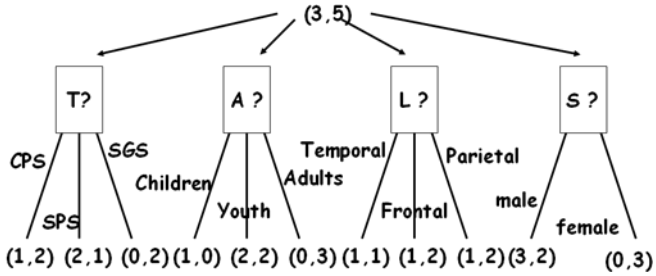

FIGURE 1. TREE OF TESTING ITEMS.

The test on T is not discriminatory as it is shown in figure 1.

The test on A is interesting on the branches Children and Adults.

The chosen test is the one with the biggest gain

Data processing using entropy:

A node p is terminal if: all the items associated with this node are in the same class or if no test could be selected.

We have chosen the test that maximizes:

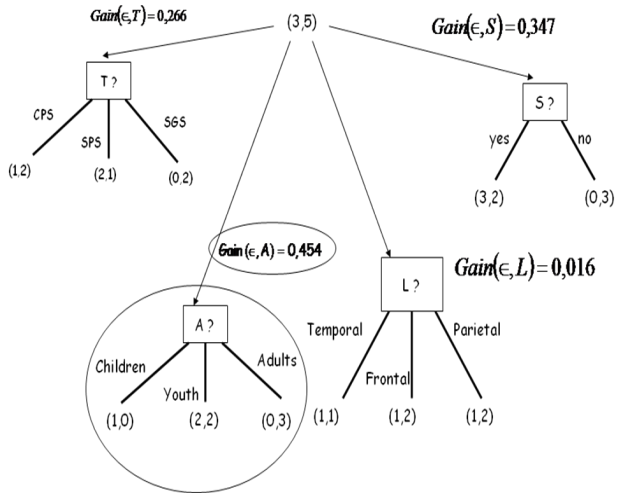

FIGURE 2. TEST CHOSEN.

So the test A(Age) is chosen because it has the maximal gain .

We use Iterative Dichotomiser supervised learning algorithms with the function below :

function Iterative_Dichotomiser (I, O, T)
I set of input attributes
O output attribute
T set of learning items
If (T is empty)
    Resend Error
If (all items of T belong to the same class)
    Return a node with the class label
If (I is empty)
    Returning a node with the most frequent label on the output attribute of T
Calculate the information gain for all attributes of I relative to T
X is the attribute with the largest gain
{x_j / j = 1,2, ..., m} are the values of X
{T_j / j = 1,2, ..., m} are the subsets of T decomposed with respect to x_j
Returning a tree with X as the root node's label and x_1, x_2, ..., x_m as labels of arcs going to trees Iterative_Dichotomiser(I-{X},O,T_1), Iterative_Dichotomiser(I-{X},O,T_2),….., Iterative_Dichotomiser (I- {X}, O, T_m).

After executing the iterative dichotomiser algorithm with all subtrees the final decision tree for epileptic predictive healing is:

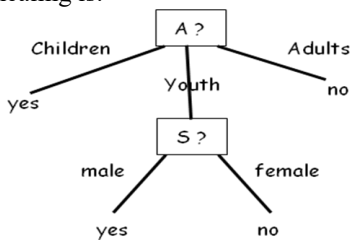

FIGURE 3. FINAL TREE.

Figure 3 shows clearly that Children (5-14 years) are responsive to treatment; Adults (25-64 years) are not. the Youth (15-24 years) male are responsive but not female.
It is shown in figures 3 and 4 with percentages of responses to treatments.

Tree size: 5 nodes, 3 leaves
Edge widths: Relative to parent
Target class: H=yes

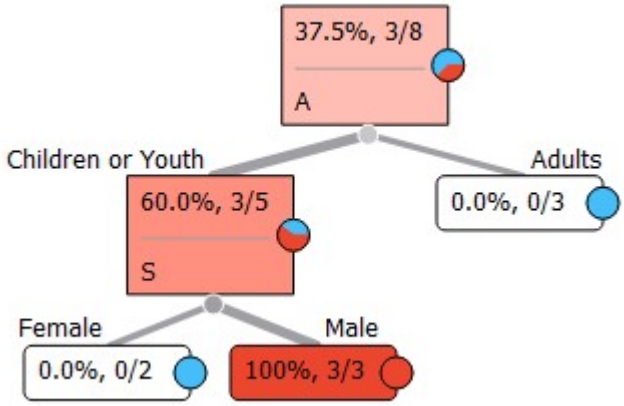

FIGURE 4. FINAL TREE. FOR H=YES

Tree size: 5 nodes, 3 leaves
Edge widths: Relative to parent
Target class: H=no

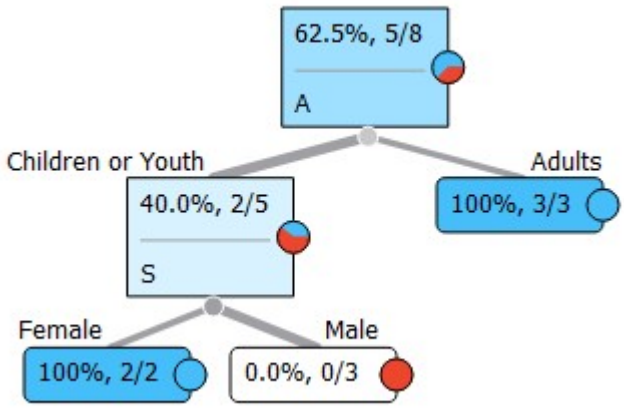

FIGURE 5. FINAL TREE. FOR H=NO

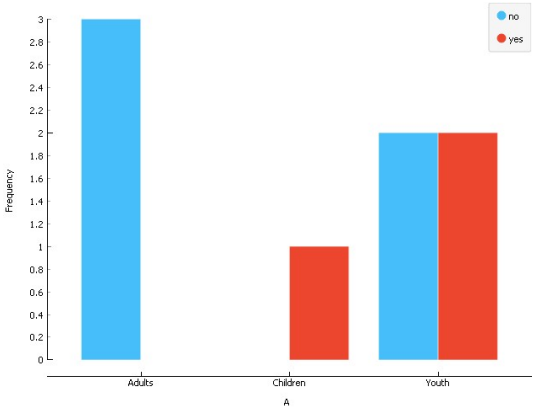

FIGURE 6. HEALING DISTRIBUTION DIAGRAMME

Figure 6 specifies that adults are not responsive to treatment because healing field contains no as value.

IV. DECISION TREE WITH GLOBAL DATASET

Our method predicts epileptic healing in this sample for only 8 patients dataset. We complete the experiment within 100 patients dataset and we find the same conclusion of decision tree.
We use Iterative Dichotomizer supervised learning algorithms with all patients dataset.
The settings of model parameters configuration are as follow:
Pruning: at least six instances in leaves, at least two instances in internal nodes, maximum depth 10
Splitting: Stop splitting when majority reaches 100% (classification only)
Binary trees: Yes
Data:
Data instances: 100
Features: Age, Sex, type, Brain Location
Target: Healing
We obtain 19 nodes with 10 leaves the tree is illustrated in figure 7 below:

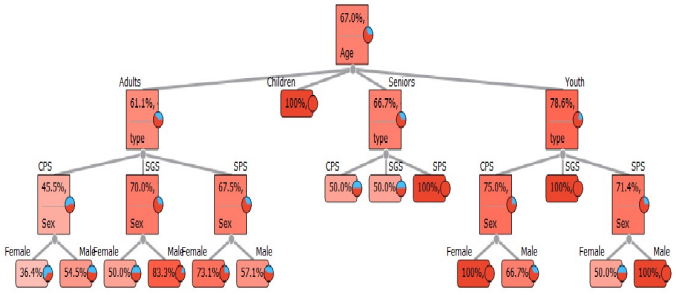

FIGURE 7. DECISION TREE OF ALL DATASETS

Figure 7 indicates that the children (5-14 years) are 100% responsive to epileptic treatment, 61.1% of Adults (25-64 years) , 78.6% of Youth (15-24 years) and 66.7% of Seniors (65 years and over). Youth males with SPS type are 100% responsive to epileptic treatment and females only 50% for the same disease.

Figure 8 illustrate filtered distribution with kernel density which confirms the result of healing prediction of our method data mining algorithms. The frequencies of red bars (treatment successful) are greater than blue ones (treatment failed) for most of patients.

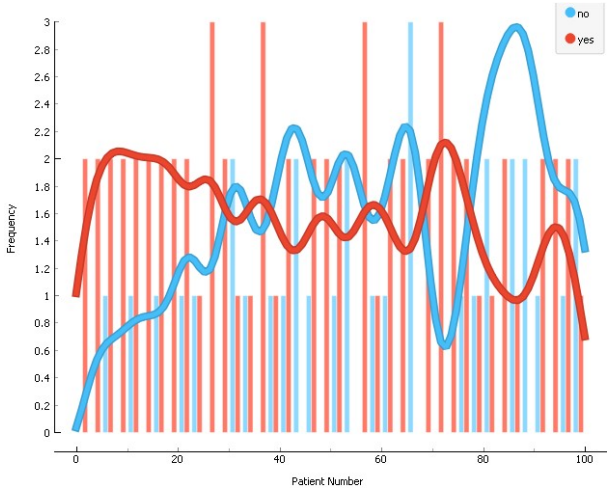

FIGURE 8. Filtered distribution with kernel density.

The figure 9 illustrate in 3 dimensions(patient number,sex,total seizures) the result of treatment according to age and epileptic type of seizures, and here again we have the same result as in our method of decision tree algorithm.

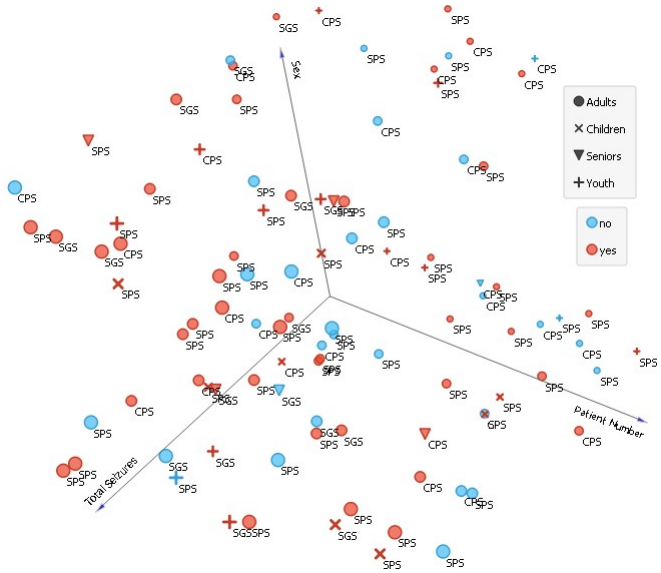

FIGURE 9.PREDICTIVE DATA MINING MULTIVARIATE VISUALIZATION

This graph (figure 9) provides a dataset visualization of the epileptic patients database and allow classification of epileptic types, expose the pertinent relations between classes and features and show feature interactions.

V. Conclusion

This paper analyze a large epileptic seizures database containing a lot of records EEGs signal processing despite of its complexity. Most of predictive methods use total seizures or cost-sensitive support vector machines without analyzing the other classes and features, thus will involve inaccurate predictions.
Our method use all classes, features and parameters thanks to effective search easily and efficiently of pertinent and significant data using data mining algorithms.
The prediction using supervised learning algorithm to generate decision tree, applied to 8 records in the epileptic dataset is nearly the same applied to 100 records.
The challenge for the future is to predict the epileptic seizure time before at least several hours.

REFERENCES

[1] Scheffer IE, Berkovic S, Capovilla G, Connolly MB, French J, et al. (2017) ILAE classification of theepilepsies: Position paper of the ILAE Commission for Classification and Terminology. Epilepsia 58: 512-521.

[2] K. Kobayashi, C. J. James, T. Nakahori, T. Akiyama, and J. Gotman, "Isolation of epileptiform discharges from unaveraged EEG by independent component analysis," Clinical Neurophysiology, vol. 110, no. 10, pp. 1755-1763, 1999.

[3] A. Ossadtchi, S. Baillet, J. C. Mosher, D. Thyerlei, W. Sutherling, and R. M. Leahy, "Automated interictal spike detection and source localization in magnetoencephalography using independent components analysis and spatio-temporal clustering," Clinical Neurophysiology, vol. 115, no. 3, pp. 508-522, 2004.

[4] J.R. Williamson, D.W. Bliss, D.W. Browne, and J.T. Narayanan, "Seizure prediction using EEG spatiotemporal correlation structure," Epilepsy & Behavior, vol,. 25, no. 2, pp. 230–238, 2012.

[5] L. Chisci, A. Mavino, G. Perferi, M. Sciandrone, C. Anile, G. Colicchio, and F. Fuggetta, "Real-Time Epileptic Seizure Prediction Using AR Models and Support Vector Machines," IEEE Transactions on Biomedical Engineering, vol. 57, no. 5, pp. 1124-1132, 2010.

[6] J. Rasekhi, M.R.K. Mollaei, M. Bandarabadi, C.A. Teixeira, and A. Dourado, "Preprocessing effects of 22 linear univariate features on the performance of seizure prediction methods," Journal of Neuroscience Methods, vol. 217, pp. 9–16, 2013.

[7] P. Mirowski, D. Madhavan, Y. LeCun, and R. Kuzniecky, " Classification of patterns of EEG synchronization for seizure prediction," Clinical Neurophysiology, vol. 120, no. 11, pp. 1927-1940, 2009.

[8] Y. Park, L. Luo, K. K. Parhi, and T. Netoff, "Seizure prediction with spectral power of EEG using cost-sensitive support vector machines," Epilepsia, vol. 52, no. 10, pp. 1761-1770, 2011.

[9] H. Feldwisch-Drentrup, B. Schelter, M. Jachan, J. Nawrath, J. Timmer, and A. Schulze-Bonhage, "Joining the benefits: Combining epileptic seizure prediction methods," Epilepsia, vol. 51, pp. 1598-1606, 2010.

[10] K. Gadhoumi, J.-M. Lina, and J. Gotman, "Seizure prediction in patients with mesial temporal lobe epilepsy using EEG measures of state similarity," Clinical Neurophysiology, vol. 124, no. 9, 1745-1754, 2013.

[11] M. Paul, W. Lin, C. T. Lau, B. Lee,"Direct Intermode Selection forH.264 Video Coding Using Phase Correlation," IEEE Transactions on Image Processing, vol. 20, no. 2, pp. 461-473, 2011.

[12] S. Li, W. Zhou, Q. Yuan, and Y. Liu, " Seizure Prediction Using Spike Rate of Intracranial EEG," IEEE Transactions on Neural Systems and Rehabilitation, vol. 21, no. 6, pp. 880-886, 2013.