







# Feature Selection and Fuzzy Rule Mining for Epileptic Patients from Clinical EEG Data

Abhijit Dasgupta<sup>1</sup>, Losiana Nayak<sup>1</sup>, Ritankar Das<sup>2</sup>, Debasis Basu<sup>3</sup>, Preetam Chandra<sup>3</sup>, and Rajat K. De<sup>1</sup>

<sup>1</sup> Machine Intelligence Unit, Indian Statistical Institute, Kolkata, India  
rajat@isical.ac.in

<sup>2</sup> Department of Bioinformatics, West Bengal University of Technology,  
Kolkata, India

<sup>3</sup> Department of Neuro-Medicine, Medical College and Hospital, Kolkata, India

**Abstract.** In this paper, we create EEG data derived signatures for differentiating epileptic patients from normal individuals. Epilepsy is a neurological condition of human beings, mostly treated based on a patient's seizure symptoms. Clinicians face immense difficulty in detecting epileptic patients. Here we define brain region-connection based signatures from EEG data with help of various machine learning techniques. These signatures will help the clinicians in detecting epileptic patients in general. Moreover, we define separate signatures by taking into account a few demographic features like gender and age. Such signatures may aid the clinicians along with the generalized epileptic signature in case of complex decisions.

**Keywords:** Epileptic signature · Brain connectivity network · Fuzzy rule mining · EEGLAB · MATLAB

## 1 Introduction

Epilepsy is one of the most serious and frequently occurring neuropathological condition affecting around fifty million people globally<sup>1</sup>. Epileptic seizures can be lethal [1]. Epileptogenesis is a long term dynamic progressing process of hyperexcitability and abnormal synchronization of brain neurons until the manifestation of seizure. Available drugs treat epileptogenesis indirectly by suppressing ictogenesis (the expression of seizures) [2]. Treatment of epileptic patients is mostly symptomatic based on clinical features.

The “epileptic network”, can be defined as a distributed network of distinct and distant brain regions causing hyperexcitability and hypersynchrony in a case of epilepsy [3]. Here we discuss the EEG data based computationally derived epileptic networks. Analysis of the epileptic networks may help in localization of a brain-region based signature in epileptic patients [4]. Till date, we have a

<sup>1</sup> <http://www.who.int/mediacentre/factsheets/fs999/en/> visited on 21st March 2017 10:48 AM Indian Standard Time.

limited idea of epileptogenesis. Epilepsy does not have a cure yet. Moreover, there does not exist any signature pattern of brain region based connections that can successfully categorize epileptic patients from normal human beings. The case becomes more complex when we try to incorporate demographic features like gender or age to EEG data of epileptic patients.

In this article, we differentiate epileptic patients from normal individuals based on brain-region connection based signatures derived from EEG data. Moreover, we try to answer a few questions like how different is male epilepsy from female epilepsy? Do they over-represent different connections among various brain regions? How child epilepsy differs from teenage and adult epilepsy? Do we find different patterns of over/under-representation of connections among various brain regions for different demographic categories? Can they be used as signature patterns clinically? In addition, we have aided our results with findings by a fuzzy rule mining based approach. Each discovered rule is a different combination of presence/absence of brain region based connections.

## 2 Data

The dataset consists of electroencephalography (EEG) data collected from 60 healthy individuals and 80 patients suffering from epilepsy. The data has also been grouped and studied according to gender and age. We have data of 31 normal males and 29 normal females in our dataset. Likewise, we have data of 43 epileptic males and 37 epileptic females in our dataset. Patients with less than 13 years of age have been defined as children. We have data of 23 normal and 41 epileptic children in our dataset. Teenagers have an age range of 13–19. We have found 14 such normal and 22 epileptic cases in our data. Patients with more than 19 years of age have been defined as adult epileptic patients. We have 23 normal and 17 epileptic adult patients.

## 3 Data Acquisition and Filtering

Data collection has been done with a computerized EEG machine (16 channels Recorders & Medicare Systems Pvt. Ltd. (RMS)). It has been used to record the EEG for an interval of 20–30 min. The data of each epileptic patient has been split into epochs of 10 seconds interval. The internationally accepted Modified Combinatorial Nomenclature (MCN) system accepted scheme for the location of electrodes has been followed. According to this scheme, each location is denoted by a combination of letter(s) and number. The letter(s) are used to identify the position of the electrodes on the brain lobes, whereas the numbers denote the hemispherical regions on the brain. The frontal polar, frontal, temporal, parietal and occipital lobes are represented by the letters ‘FP’, ‘F’, ‘T’, ‘P’, and ‘O’ respectively, whereas odd and even numbers stand for electrode position on the left and right hemisphere respectively. The letter ‘C’ is used for identification only. During the collection of data, it has been asked to all participants to stay awake and motionless with wide open eyes. Subsequently, they have been

requested to attain a no-thinking state as far as possible. Each data has been recorded using a series of activation procedures, *i.e.*, eye blinking, photic stimulation, and hyperventilation among others.

After the recording of EEG data, firstly, the noise has been removed manually by experienced neurotechnologists. Then we have used EEGLAB toolbox version 13 [5], implemented in MATLAB R2015a for further data filtering purpose. The EEGLAB plugin, called CleanLine, has been used to remove sinusoidal noise from raw EEG data. The resultant data have been filtered again using Finite Impulse Response (FIR) filter, within the range of 4–60 Hz to remove sleep waves and noise due to electrical circuits. Here, we have used Independent Component Analysis (ICA) by applying the Runica algorithm [6]. Thus, a multivariate signal is decomposed into its additive independent non-Gaussian components. We have separated the maximum likely components from a number of noisy components using neural networks. Lastly, a final manual check has been done to assure that artifacts from the data have been removed.

### 4 Methodology

We have divided the methodology into four different steps. The steps have helped in selecting significant features and in finding important rules for epileptic patients in general and for different demographic categories. Figure 1 depicts the flowchart of our methodology pipeline.

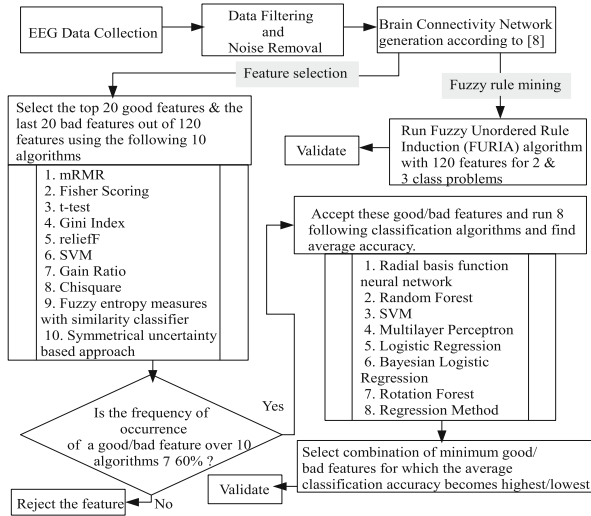


Fig. 1. Flow chart of the methodology.

1. **Brain connectivity network generation:** We have calculated the correlation between two certain electrode positions using Pearson’s correlation coefficient [7]. We have considered the 16 electrode positions as vertices. If two vertices are found to be correlated (positively or negatively), we have created an edge between them. Thus, we have developed the brain connectivity network of normal individuals and epileptic patients. The detailed explanation of this step can be found in one of our previous research work [8].
2. **Determination of maximum features:** We have developed a  $16 \times 16$  adjacency matrix for each brain connectivity network representing either a healthy volunteer or an epileptic patient. This adjacency matrix is symmetric with ‘0’s in its diagonal elements. Thus, maximum  $(\frac{16 \times 15}{2}) = 120$  unique undirected connections are possible from such a matrix. These 120 connections have been considered as features.
3. **Feature selection and classification:** We have identified the key features discriminating normal individuals from epileptic patients (a two class problem), male epileptic patients from female epileptic patients (another two class problem) and child epileptic patients from teenage as well as adult epileptic patients (a three class problem). We have used ten well-established feature selection algorithms, *i.e.*, mRMR [9], Fisher scoring [10], t-test, Gini index, reliefF, Support Vector Machine (SVM) [11], gain Ratio, Chisquare, fuzzy entropy measures with similarity classifier [12], and symmetrical uncertainty based approaches [13–15] among others to identify top twenty most significant features and last twenty least significant features for the above mentioned classification problems. We have selected those most/least significant features whose frequency of occurrence over the ten algorithms is more than 60%. We have used eight well-known classification algorithms, *i.e.*, radial basis function neural network [16], random forest [17], SVM [11], multilayer perceptron, logistic regression, Bayesian logistic regression, rotation forest [18], and regression method among others to classify the combination of the minimum number of most and least significant features. It has been done to maximize the average classification accuracy of the most significant features and minimize the same for the least significant features. Thus, a combination of minimum features has been selected for each of the classification problems mentioned earlier.
4. **Fuzzy rule mining:** We have used Fuzzy Unordered Rule Induction Algorithm (FURIA) [19] to identify a few rules for differentiation of the aforementioned classes. Here, we have considered all 120 features as input to FURIA. Besides, we have calculated the Certainty Factor (CF) for each rule. CF lies in a range of [-1, 1]. If the antecedent and consequent are related, the value of CF becomes positive. A higher value of CF represents more significant rule.

## 5 Results and Discussion

We have found certain brain-region based connections (features) over represented and a few under represented while comparing epileptic patients with normal volunteers. C3-F3, F7-O1, F7-O2, T7-F7, F8-FP1, P8-P4 and P8-P7 features have

been found to be over represented in epilepsy. On the other hand, P7-P4 and T8-F3 features have been found to be under represented. The combination of these nine significant features (Table 1) has shown 82.59% average classification accuracy with 20 fold cross validation to separate normal individuals from epileptic patients in general. The representation pattern of these features constitutes a generalized epileptic signature.

In a similar way, we have found a few over represented and under represented features while comparing male and female epileptic patients. P4-F4 and P8-C4 features have been found to be over represented along with under representation of O2-C3, O2-P3, P7-O2 and F8-F7 features in case of female epileptic patients. Here, we have found 70.94% average classification accuracy with 20 fold cross validation for a combination of these six significant features (Table 1) for identifying male and female epileptic patients separately. Over or under representation pattern of these features constitute the gender-specific epileptic signatures.

On the other hand, the combination of eight promising good features (Table 1) has shown 62.14% average classification accuracy to discriminate child epileptic patients from teenage along with adult epileptic patients. The F8-F3 feature has been found to be over represented along with two under represented P8-FP1 and P8-T8 features in case of child epilepsy. In the cases of teenage epilepsy we have found an over represented O2-F3 feature and under represented F7-P4 feature. P8-T8 and O2-F4 features have been found to be over represented in the case of adult epilepsy. The representation pattern of these features constitutes the age-specific epileptic signatures. The rest of the features seem inconclusive at this point according to their frequency of occurrence. However, the average classification accuracy value may increase with a larger sample size.

In support of the aforementioned results, a fuzzy rule based association mining study has been done with the all possible 120 features (connections among different nodes) as given in Table 2 to find unique rules. We have performed this study to generate support for our earlier findings. Moreover, this study has provided additional associated features. The additional features helped in defining a proper rule for the general, age-wise and gender-wise epileptic signatures. Some features, i.e., T7-F7 and P8-P7 has been found to be associated with epilepsy as described earlier in Table 1. The presence of the feature T7-F7 has been found coupled with the presence of two additional features P7-F3 and P3-FP1. The rule indicates epilepsy with a certainty factor of 0.95. The presence of P8-P7 feature has been found with the presence of a new feature C4-FP1 with a certainty factor of 0.93.

Moreover, multiple rules have been found for the P7-P4 feature. The absence of the feature along with the absence of two additional features P8-F7 and O2-C4 have been found to be associated with epilepsy with a certainty factor of 0.97. On contrary, the presence of the P7-P4 feature along with the absence of three additional P4-FP2, T7-F7, and F8-P3 features has been found in normal individuals with a certainty factor of 0.96. Also, the presence of the same factor along with the absence of three additional O2-FP2, T7-F7, and P3-C3 features

**Table 1.** List of significant features differentiating epilepsy from normal individuals; male from female epilepsy; and child from teenage and adult epilepsy. Odd numbers succeeding node names indicate left hemisphere of the brain and similarly, even numbers indicate right hemisphere of the brain. The frequency of occurrence is given in percentage.

<b>Normal Individuals vs Epileptic Patients</b>				
Connection between lobes (Features)	Nodes	Epileptic	Normal	In epilepsy
Left identification point & frontal lobe	C3-F3	40	16.66	More
Left frontal lobe & occipital lobe	F7-O1	57.50	23.33	More
Left frontal lobe & right occipital lobe	F7-O2	52.50	0.10	More
Left temporal lobe & frontal lobe	T7-F7	35	0.12	More
Left parietal lobe & right parietal lobe	P7-P4	45	88.33	Less
Right frontal lobe & left frontal polar lobe	F8-FP1	42.50	0.13	More
Right temporal lobe & left frontal lobe	T8-F3	0.15	38.33	Less
Right parietal lobe & parietal lobe	P8-P4	33.75	0.10	More
Right parietal lobe & left parietal lobe	P8-P7	35	0.07	More
<b>Male vs Female Epileptic Patients</b>				
Connection between lobes (Features)	Nodes	Male	Female	In female
Right parietal lobe & frontal lobe	P4-F4	0.06	35.13	More
Right occipital lobe & left identification point	O2-C3	44.18	27.02	Less
Right parietal lobe & identification point	P8-C4	60.46	83.78	More
Right occipital lobe & left parietal lobe	O2-P3	32.55	18.91	Less
Left parietal lobe & right occipital lobe	P7-O2	48.83	29.72	Less
Right frontal lobe & left frontal lobe	F8-F7	48.83	29.72	Less
<b>Child vs Teenage &amp; Adult Epileptic Patients</b>				
Connection between lobes (Features)	Nodes	Child	Teenage	Adult
Right frontal lobe & left frontal lobe	F8-F3	49	13.63	11.76
Right parietal lobe & left frontal polar lobe	P8-FP1	0.07	40.90	35.29
Right parietal lobe & temporal lobe	P8-T8	29.26	40.90	64.70
Right occipital lobe & frontal lobe	O2-F4	12	9.09	35.29
Right occipital lobe & left occipital lobe	O2-O1	68	86.36	47.05
Left occipital lobe & frontal polar lobe	O1-FP1	34.14	18.18	11.76
Right occipital lobe & left frontal lobe	O2-F3	48.78	54.54	23.52
Left frontal lobe & right parietal lobe	F7-P4	19.51	9.09	17.64

has been found in normal individuals with a certainty factor of 0.96. In both the cases, the feature T7-F7 and its absence is a common factor. Such findings reflect the variable states of brain regions even for a very specific disease state.

**Table 2.** Fuzzy rules with its certainty factor (CF) for differentiating epilepsy from normal individuals; male from female epilepsy; and child from teenage and adult epilepsy. Range of CF is [-1,1].

<b>Normal Individuals vs Epileptic Patients</b>		
IF condition	Then	CF
(P7-P4) absent & (P8-F7) absent & (O2-C4) absent	Epilepsy	0.97
(P4-FP2) present & (O2-FP2) present & (C3-FP1) absent	Epilepsy	0.96
(F7-O2) present & (O2-FP2) absent & (O2-F4) absent	Epilepsy	0.96
(T7-F7) present & (P7-F3) present & (P3-FP1) absent	Epilepsy	0.95
(P8-P7) present & (C4-FP1) present	Epilepsy	0.93
(P3-C3) present & (F8-FP2) absent & (O2-FP1) absent	Epilepsy	0.92
(P7-P4) present & (P4-FP2) absent & (T7-F7) absent & (F8-P3) absent	Normal	0.96
(F7-O2) absent & (T7-P3) absent & (P7-F7) absent	Normal	0.92
(P7-C3) absent & (P4-FP2) absent & (F7-C4) present & (P3-F3) absent	Normal	0.94
(P8-O1) present & (F8-O1) present & (O1-C3) absent	Normal	0.84
(P7-P4) present & (O2-FP2) absent & (T7-F7) absent & (P3-C3) absent	Normal	0.96
<b>Male vs Female Epileptic Patients</b>		
IF condition	Then	CF
(F8-T7) present & (O1-FP1) present	Male epilepsy	0.92
(O2-P3) present & (O1-FP1) absent	Male epilepsy	0.91
(P7-O2) present & (F7-F3) absent & (T8-F8) absent	Male epilepsy	0.93
(P4-F4) present & (P8-P3) present	Female epilepsy	0.89
(C3-FP1) present & (T7-C4) absent & (F8-T7) absent	Female epilepsy	0.91
(P8-F7) present & (F8-F7) absent & (P3-C3) present	Female epilepsy	0.88
(F8-O1) present & (F4-P3) absent & (C3-FP1) absent	Female epilepsy	0.86
<b>Child vs Teenage &amp; Adult Epileptic Patients</b>		
IF condition	Then	CF
(F8-F3) present & (T7-FP2) absent	Child epilepsy	0.94
(O1-C3) present & (C4-C3) present	Child epilepsy	0.89
(F8-P4) present & (P8-FP1) absent & (C4-F3) present	Child epilepsy	0.92
(P8-F4) present & (T7-C4) absent & (FP2-O1) absent	Child epilepsy	0.89
(P7-P4) absent & (P8-FP1) present & (F7-P4) absent	Teenage epilepsy	0.86
(P7-P3) present & (C3-FP1) absent & (O1-C3) absent	Teenage epilepsy	0.82
(P8-T8) present & (F8-P4) absent & (T8-FP1) absent	Adult epilepsy	0.8
(F4-C3) absent & (T7-F7) present & (F7-O1) present	Adult epilepsy	0.74
(T8-FP2) present & (P8-P4) present & (F3-FP1) absent	Adult epilepsy	0.74

The feature F7-O2 has been found to be associated with both epileptic patients and normal individuals. The presence of the feature along with the absence of two additional O2-FP2 and O2-F4 features has been associated with epilepsy with a certainty factor of 0.96. On contrary, the absence of the feature along with the absence of two additional T7-P3 and P7-F7 features has been seen in normal individuals with a certainty factor of 0.92. These rules altogether contribute to the generalized epileptic signature.

In the case of male versus female epileptic patients, the presence of O2-P3 feature along with the absence of an additional feature O1-FP1 has been found to be associated with male epilepsy with a certainty factor of 0.91. Similarly, the presence of the feature F8-T7 along with the presence of another additional feature O1-FP1 has been found to be associated with male epilepsy with a certainty factor of 0.92. In contrast, the presence of P4-F4 feature along with the presence of an additional feature P8-P3 has been found to be associated with female epilepsy with a certainty factor of 0.89. The absence of the feature F8-F7 along with the presence of two additional features P8-F7 and P3-C3 have been found to be associated with female epilepsy. Such results support our earlier finding of features in male and female epilepsy as given in Table 1. These rules contribute to the gender-specific epileptic signatures.

We have also done an age-specific fuzzy rule mining study as given in Table 2. Here the presence of F8-F3 and the absence of P8-FP1 features have been found in two separate rules. They have been found to be associated with child epilepsy with certainty factors 0.94 and 0.92 respectively. The absence of P8-FP1 feature along with the presence of two additional features P7-P4 and F7-P4 with a certainty factor of 0.86 has been found to be associated with teenage epilepsy. In the case of adult epilepsy, the presence of the feature P8-T8 along with the absence of two additional features F8-P4 and T8-FP1 has been found to be associated with a certainty factor of 0.8. These results provide support to our earlier findings as given in Table 1 as well as contribute to the age-specific epileptic signatures.

Moreover, the fuzzy rule mining study has discovered some new rules which we have not spotted in our earlier findings. We have found 2 new rules each for epileptic patients and normal individuals. One male and two female epilepsy related new rules have also been found by this study as given in Table 2. In addition, two child, one teenage and two adult epilepsy associated new rules have been found. These new rules along with our earlier findings (Table 1) helped in creating well-defined signatures for future machine learning based detection of epilepsy in general and complicated cases.

## 6 Conclusions

In this paper, we have found a generalized epileptic signature from EEG data of patients along with gender and age specific signatures. It is not always easy for medical practitioners to rightly identify an epileptic patient, the reason being similar kind of EEG spikes found in other neurobiological disorders. The created epileptic signatures may help them in overcoming these hurdles in epilepsy

patient detection. Moreover, these kinds of predictions can come in handy with minimal human intervention in peripheral areas where the proper medical facility is not available yet. They can quickly be employed to detect probable epilepsy. Then the patient can be referred for proper medical care. In this paper, we were also able to distinguish between patients from different age groups and gender but it has not yet been clinically tested. According to clinicians, identifying the gender-specific epileptic signatures is an interesting concept and may be helpful in the future.

**Acknowledgment.** Abhijit Dasgupta acknowledges Visvesvaraya Ph.D. scheme for Electronics and IT empowered by Ministry of Electronics and Information Technology, Government of India, for a research fellowship. Losiana Nayak acknowledges University Grants Commission, India for a UGC Post-Doctoral Fellowship (No. F.15-1/2013-14/PDFWM-2013-14-GE-ORI-19068(SA-II)).

## References

1. Devinsky, O.: *N. Engl. J. Med.* **365**(19), 1801–11 (2011)
2. Giourou, E., et al.: *Cyberphysical Systems for Epilepsy and Related Brain Disorders*, pp. 11–38. Springer, Cham (2015)
3. Wendling, F., et al.: *Front Syst. Neurosci.* **4**, 154 (2010)
4. Dixit, A.B., et al.: *Neurol. India* **63**(5), 743 (2015)
5. Delorme, A., et al.: *J. Neurosci. Methods* **134**(1), 9–21 (2004)
6. Naik, G.R., et al.: *8th IEEE International Conference on Computer and Information Technology* (2008)
7. Sedgwick, P.: *BMJ* **345**, 7 (2012)
8. Dasgupta, A., et al.: *IEEE International Conference Bioinformatics Biomedical*, pp. 815–821. IEEE (2015)
9. Peng, H., et al.: *IEEE Trans. Pattern Anal. Mach.* **27**(8), 1226–38 (2005)
10. Mika, S., et al.: *9th IEEE Workshop on Neural Networks for Signal Processing*. IEEE (1999)
11. Wang, L. (ed.): *Support Vector Machines: Theory and Applications*, vol. 177. Springer, Heidelberg (2005)
12. Luukka, P.: *Expert Syst. Appl.* **38**(4), 4600–7 (2011)
13. Saeys, Y., et al.: *Bioinformatics* **23**(19), 2507–17 (2007)
14. Shang, W., et al.: *Expert Syst. Appl.* **33**(1), 1–5 (2007)
15. Liu, H., et al. (eds.): *Computational methods of feature selection*. CRC Press, Boca Raton (2007)
16. Yingwei, L., et al.: *IEEE Trans. Neural Netw.* **9**(2), 308–18 (1998)
17. Liaw, A., et al.: *R News* **2**(3), 18–22 (2002)
18. Rodriguez, J., et al.: *IEEE Trans. Pattern Anal. Mach.* **28**(10), 1619–30 (2006)
19. Huhn, J., et al.: *Data Min. Knowl. Discov.* **19**(3), 293–319 (2009)