Confidence in Medical Decision Making: Application in Temporal Lobe Epilepsy Data Mining

Shobeir Fakhraei[1,2], Hamid Soltanian-Zadeh[2,3], Farshad Fotouhi[1], Kost Elisevich[4]

1. Dept. of Computer Science Wayne State University Detroit, MI, USA
2. Image Analysis Lab. Dept. of Radiology Henry Ford Health System Detroit, MI, USA
3. CIPCE, School of Elec. and Comp. Eng. University of Tehran, Tehran, Iran
4. Dept. of Neurosurgery Henry Ford Health System Detroit, MI, USA

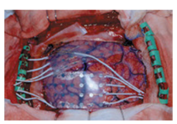
Abstract

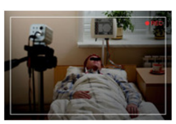
Prior to neurosurgical resection of abnormal brain tissues in mTLE patients, focal points of the seizure should be identified via a set of examinations. Once decisive evidence is not present in noninvasive clinical profile of mTLE patients, extraoperative Electrocorticography (ECoG) is required which is the practice of using electrodes placed directly on the exposed surface of the brain. Through classification techniques on a dataset of mTLE patients we have studied the possibility of reduction of such requirement and shown significant results. Furthermore, we have shown that in critical domains such as medicine use of AUC does not provide sufficient information about the confidence of the classification and further measures are needed.

Introduction

- Epilepsy is a disorder of the brain characterized by an enduring predisposition to generate epileptic seizures and by the neurobiological, cognitive, psychological and social consequences of this condition.
- Neurosurgical resection of the abnormal brain tissues in patients suffering from Mesial temporal lobe epilepsy (mTLE) is a way of eliminating and reducing the occurrence of epileptic seizure onsets.
- Prior to such operation, focal points of the seizures should be identified via a set of examinations.

- Once decisive evidence is not present in noninvasive clinical profiles of mTLE patients, extraoperative Electrocorticography (ECoG) is required. ECoG is the practice of using electrodes placed directly on the exposed surface of the brain to record electrical activity from the cerebral cortex.

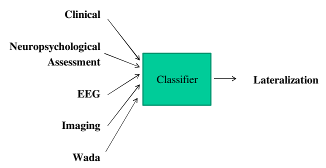




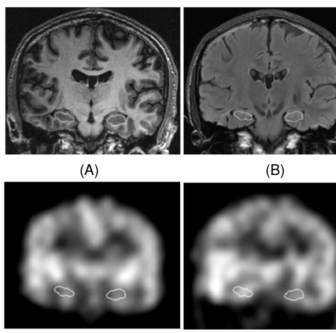
- Data mining techniques have is applied in this study to provide decision assistance in lateralizing focal epileptogenicity.
- The goal of this paper is to reduce the need for ECoG via data mining techniques and finding the best classifier for this purpose.
- Since decision making is highly critical in medical domains, classifiers that result in higher decision confidence are preferred.
- To be able to evaluate such confidence in different classifiers, we propose a new measure and compare it with well known area under receiver operating characteristic (ROC) curve (AUC) measure.

HBIDS

To integrate several clinical attributes of TLE patients from various sources and subsystems, human brain image database system (HBIDS), which is a clinical and imaging database of TLE patients, is developed at the radiology research department of Henry Ford Health System in Detroit Michigan:



1. Imaging features were generated using the hippocampi outlines. A domain expert outlined all hippocampal contours on coronal slices of T1-weighted images using in-house software and a previously established protocol. These were then verified by another expert.
The T1-weighted and fluid-attenuated inversion recovery (FLAIR) magnetic resonance (MR) image sets were co-registered using a rigid registration technique based on mutual information. Similarly, ictal and interictal single photon emission computed tomography (SPECT) image sets were co-registered to the T1-weighted image set.
Four sets of features were extracted from each hippocampal ROI: mean and standard deviation of the FLAIR MR signal intensity, wavelet transform-derived energy, volumetry, and SPECT ictal-interictal mean difference.



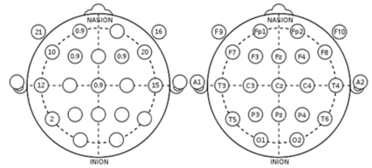
(A) (B)

(C) (D)

Manual segmentation of the hippocampi in a representative coronal T1-weighted MR image (A) and its map on the FLAIR MRI (B), interictal SPECT (C), and ictal SPECT (D).

2. Neuropsychological profile of the patients include pre- and post-operative measurement through Boston naming test (BTN), Wechsler memory scale (WMS), Rey-Osterrieth complex figure test (ROCF), California verbal learning test, and intelligence quotient (IQ) test. Quantitative measures of patients' verbal and non-verbal memory are recorded and stored in the database.

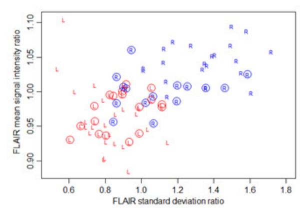
3. Dataset includes descriptive noninvasive electrographic features. Ictal onset locations and three most predominant interictal localities of sharp and slow waveforms are provided. Ictal onset locations were integrated into one feature indicating the probability of focal epileptogenicity in the right temporal lobe. Figure 5B demonstrates the surface electrode configuration of the EEG recordings. The abundance percentage of the two most dominant interictal sharp wave activities in each location for all patients in the dataset are shown in the figure below.



4. Wada test which is also known as the "intracarotid sodium amobarbital procedure" (ISAP), is used to establish cerebral language and memory representation of each hemisphere. Laterality of language dominance in cerebral hemispheres, and the number of correctly recalled items after left and right carotid injections, is also stored in the database.

Patient Cohort

- In this study, 79 patients are selected (31 males, 48 females) with 197 medical features. The patients have an average age of 38y (S.D. 12.2). Temporal lobe epileptogenicity is found to be on the left side in 43 patients and the right side in 36 patients.
- In 46 patients, standard noninvasive evaluations lateralize the TLE sufficiently well to proceed with resection of the site of epileptogenicity directly, whereas, 33 patients require ECoG (Phase II)(41.7%).
- The dataset contains missing values in different features due to various reasons such as inability to perform all medical tests for each patient. Missing values are identified for EEG features in 21% of cases, for Wada studies in 31%, for SPECT imaging features in 35%, and for FLAIR and volumetric imaging in less than 10% of cases. The missing values of the remaining features are found in about 20% of cases on average.

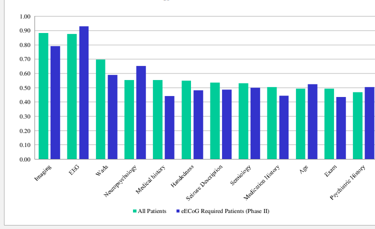


Side of abnormality in patients is shown with "R" and "L" letters, respectively. Phase II patients are outlined.

Feature Selection

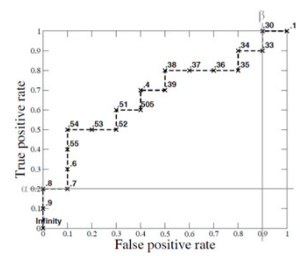
With 79 patients and 197 attributes, the need for feature selection is apparent. we have applied a heterogeneous ensemble of single variable classifiers to rank the medical features based on their individual predictive performance.

$$P(f_i) = \sum_{c_k \in C} AUC(c_k, f'_i)/|C|$$



Confidence Evaluation

Our domain has zero tolerance for invalid decision, although the classification is binary, two thresholds should be used, and the final classification response should be "left", "right", or "undecided".
To achieve such classifier, any chosen thresholds have to be set on points where no mistakes are made when the side of abnormality is predicted. The limits of the thresholds are invalid predictions with the highest predicted probabilities. (α and β)



The α and β limit are the upper bounds for the classifier performance in this fashion and could indicate the classifier's potential in such classification. We refer to this measure as confident prediction rate" (CPR):

$$CPR = \frac{\text{\# of possible confident predictions}}{\text{Total \# of samples}} * 100$$

where samples with predicted probabilities more extreme than α and β are possible confident predictions.

Experimental Results

Performance evaluation of different classifiers on different feature subsets.

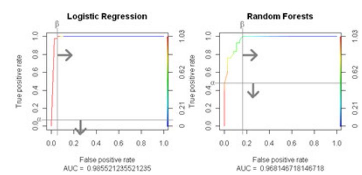
(a) AUC of the classifiers using different feature subsets

Features	NB	SVM	MLP	3NN	LR	RF
All 9	0.993	0.959	0.978	0.964	0.986	0.968
Top 4	0.973	0.970	0.974	0.952	0.966	0.957
Imaging	0.982	0.981	0.957	0.929	0.975	0.916
EEG	0.951	0.951	0.943	0.964	0.925	0.958
Wada	0.609	0.660	0.599	0.596	0.660	0.689

(b) Confident prediction rate (CPR) of the classifiers using different feature subsets

Features	NB	SVM	MLP	3NN	LR	RF
All 9	84.8%	36.7%	72.2%	43.0%	44.3%	64.6%
Top4	65.8%	72.2%	54.4%	38.0%	41.8%	36.7%
Imaging	81.0%	75.9%	59.5%	0.0%	72.2%	0.0%
EEG	53.2%	65.8%	45.6%	73.4%	30.4%	62.0%
Wada	10.1%	13.9%	12.7%	22.8%	13.9%	12.7%

Using LR with all 9 features which generated the AUC of 0.986 and CPR of 44.3% while RF results in lower AUC of 0.968 with higher CPR of 64.6%. In a medical domain such as this case, RF should be preferred over LR despite the AUCs suggesting otherwise.



Discussion and Conclusion

- **Using six classifiers, we showed the possibility of using data mining techniques to build a decision support system that could potentially lateralize 84.8% of the patients with high confidence without the need for extraoperative Electrocorticography (ECoG). Lacking such system, only 58.2% of patients were lateralized by domain experts using noninvasive methods. Using this method, it is potentially possible to lateralized 78.8% of the phase II patients, while only 8.7% of the phase I patients will be undecided.**
- **We also demonstrated that AUC does not provide sufficient information about the confidence of the classification and other measures such as our proposed "confident prediction rate" (CPR) are needed in domains such a medicine.**
- **Using the experiments, we also demonstrated that classifiers that generate high AUCs might not be sufficiently confident for domains that require reliable predictions.**

References

[1] C. P. Panayiotopoulos, A clinical guide to epileptic syndromes and their treatment: Springer Verlag, 2010.
[2] R. Bellazzi and B. Zupan, "Predictive data mining in clinical medicine: Current issues and guidelines," International Journal of Medical Informatics, vol. 77, pp. 81-97, 2008.
[3] M.-R. Siadat, H. Soltanian-Zadeh, F. Fotouhi, A. Eetemadi, and K. Elisevich, "Data modeling for content-based support environment (C-BASE): Application on epilepsy data mining," in 17th IEEE International Conference on Data Mining Workshops, October 28, 2007 - October 31, 2007, Omaha, NE, United states, 2007.
[4] Y. Saeys, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics," Bioinformatics, vol. 23, p. 2507, 2007.
[5] S. Fakhraei, H. Soltanian-Zadeh, F. Fotouhi, and K. Elisevich, "Consensus feature ranking in datasets with missing values," in 9th International Conference on Machine Learning and Applications, ICMLA 2010, December 12, 2010 - December 14, 2010, Washington, DC, United states, 2010.
[6] S. Fakhraei, H. Soltanian-Zadeh, F. Fotouhi, and K. Elisevich, "Effect of classifiers in consensus feature ranking for biomedical datasets," in 4th International Workshop on Data and Text Mining in Biomedical Informatics, DTMBIO'10, CIKM'10, October 26, 2010 - October 30, 2010, Toronto, ON, Canada, 2010.
[7] S. Fakhraei, H. Soltanian-Zadeh, K. Elisevich, and F. Fotouhi, "Attribute ranking for lateralizing focal epileptogenicity in temporal lobe epilepsy," in 17th Iranian Conference in Biomedical Engineering, ICBME 2010, November 3, 2010 - November 4, 2010, Isfahan, Iran, 2010
[9] T. Fawcett, "ROC graphs: Notes and practical considerations for researchers," Machine Learning, vol. 31, pp. 1-38, 2004.