International Journal of Computer Applications (0975 – 8887)
Volume 57– No.16, November 2012

Informed Under-Sampling for Enhancing Patient Specific Epileptic Seizure Detection

Hussein El-Saadi
Masters student, college of computing, AAST. Cairo, Egypt

Ahmed Farouk Al-Sadek
Central Lab for Agricultural Expert Systems, Giza, Egypt & Faculty of computer science, MSA, Giza, Egypt

Mohamed Waleed Fakhr
College of computing, AAST Cairo, Egypt & Electrical Engineering Department, University of Bahrain

ABSTRACT
Thirty percent of epileptic patients encounter intractable seizures, (seizures that do not respond to medication), thus, an accurate seizure detector would help improve their quality of life. Unfortunately, seizure detection is one of the many fields that suffer from imbalanced dataset i.e. the ratio between ictal and inter-ictal records is huge which makes it difficult to build an accurate classifier. This paper attempts to build a classifier that is able to overcome the previously mentioned challenge by dividing the dataset in ensembles and utilizing multiple SVM classifiers. As a result, the detector was able to reach an overall accuracy of 97.3%; thus, opening the field for building strong classifiers from highly imbalanced datasets in the biomedical domain.

General Terms
Pattern Recognition, Neuroinformatics, Brain Machine Interface (BCI) And Algorithms.

Keywords
Seizure detection, imbalanced dataset, SVM ensemble, approximate entropy.

1. INTRODUCTION
Epilepsy is one of the oldest and well known chronic neurological disorders, effecting around 50 million people worldwide (1% of the words population). Epilepsy is characterized by frequent and recurrent seizures. The problem with seizure detection is finding appropriate features that distinguish between ictal and inter-ictal brain states as well as building a classifier that is capable of learning from the highly imbalance dataset due to the inferior ratio of ictal to inter-ictal records in the dataset. In recent years, various methods were developed to detect seizures either through scalp EEG or using invasive signal acquisition methods such as intracranial EEG.

2. RELATED WORK
Seizure detection algorithms are divided into multiple categories which includes the classification between epileptic and non-epileptic patients, seizure episode counting and onset seizure detection. In case of the latter, the algorithm focuses on detecting the seizure with the least possible delay unlike the seizure episode counting algorithms which are more focused on getting the number of seizure episodes encountered by the patient rather than its early detection.

Seizure detection approaches described in other papers included the use of frequency based analysis [1] for feature extraction; using methods such as Fourier transformation, wavelet transformation or filter banks. Other approaches includes the use of nonlinear analysis [2] such as largest Lyapunov exponent, Kolmogorov entropy or approximate entropy in the formation of the feature vector additionally other papers suggested a combination of both (spectral based and non-linear) [3] features for the feature vector construction. Furthermore, one paper introduced the use of numerical differentiation technique [4] by calculating the time derivative and zero-crossings and using them as features but the objective of that paper was to differentiate between epileptic and non-epileptic patients, finally another paper tested the use of genetic algorithms for optimum feature selection [5].

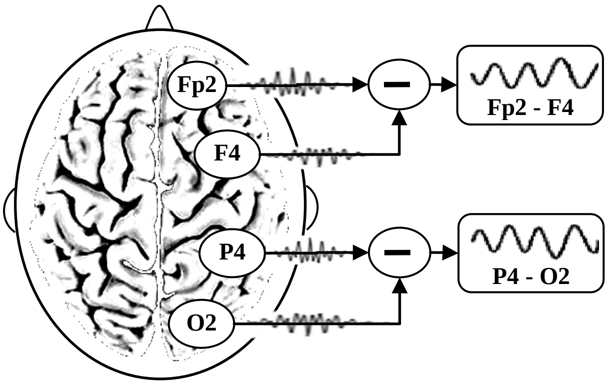
This paper focuses on detecting seizures from scalp EEG with the least possible delay and without sacrificing the overall accuracy of the seizure detector by addressing the prior mentioned challenges.

3. MOTIVATION
An accurate seizure detection algorithm could provide a permanent cure for epilepsy or at least provide a permanent method for seizure control. Although plenty of medications exist on the market that claim to provide seizure control, yet 30% of medicated epileptic patients still encounter seizure episodes. With the appearance of non convulsive drugs and vagus nerve stimulation, a patient could have permanent control over seizures [6], only if such drug or stimulation can be activated at the start of the seizure episode and this is where a seizure detection algorithm can be used to permanently control seizure episodes. Seizure detection is a very challenging topic due to the similarities between ictal and inter-ictal records plus the existence of artifacts and noises in the acquired data in addition to the imbalanced dataset nature of the problem which is due to the huge ratio between ictal and inter-ictal records.

Each of above challenges is considered a research topic on its own, however this paper focuses on finding suitable features between ictal and inter-ictal records as well as finding an appropriate classifier which can accurately classify between the prior features. This paper describes the use of an informed under sampling algorithm to overcome the imbalanced dataset challenges.

4. SIGNAL ACQUISITION
The EEG data mentioned in this paper was collected at the Children's Hospital Boston for pediatric subjects suffering from intractable seizures [7] (seizure that do not respond to medication). All signals were sampled at 256 samples per second with 16-bit resolution. The recorded data used the International 10-20 system of EEG electrode positions with bipolar electrode measurements see figure1, as described by [8].

Bipolar EEG montage is calculated by getting the potential difference between a pair of electrodes. One of the advantages of using bipolar over uni-polar EEG montage is that bipolar montage usually displays well the local abnormalities, since a phase reversal is often present. The exception occurs when the discharge is maximal at either the beginning or the end of the sequential chain. The second advantage is that bipolar montages can help resolve ambiguous findings on referential montages due to an active reference [9].



Figuer1.Bipolar electrode measurement.

An experienced technician marked the seizure onset times; however, seizure classification (partial or generalized) was not provided, that's why a patient specific approach was a must.

The channels used by the proposed seizure detection algorithm are FP1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8, F8-T8, T8-P8, P8-O2, FZ-CZ and CZ-PZ; these were the only channels available for all patients provided as per the dataset used.

5. PROPOSED SOLUTION

Seizure detection is no different than any other pattern recognition problem (a little bit more complex maybe). Pattern recognition can be summarized in training (which is composed of Pre-processing – Processes the data so it is in a suitable form, Feature extraction – Reduces the amount of data by extracting relevant information—usually results in a vector of scalar values, and Model Estimation – from the finite set of feature vectors, needs to estimate a model) and testing (which is composed of Pre-processing, Feature extraction – both same as above and Classification – compares feature vectors to the various models and finds the closest match).

The proposed algorithm focuses on the feature extraction and handling the imbalance of the dataset.

5.1 Feature extraction

In order to be able to extract useful features from the EEG signals, firstly a seizure definition must be established, by definition, a seizure is “when there is a synchronized activity of brain cells and they are firing at the same time; also known as ‘Burst of activity’ where normally the brain functions in a desynchronized manner” [10].

From the above definition, the features of a seizure can be characterized into two main categories. Firstly, the amount of energy in certain frequency bands depending on the type of seizure and secondly, a measure of how the signals of the brain are synchronous (complexity measure).

The feature extraction algorithm extracts both features mentioned before, from an EEG signal using the following methodology.

5.1.1 Pre-processing

Each channel of the EEG channel set is sliced into 2 second epoch (window) with an overlap of 1 second between each window. This two second window was chosen to provide high amount of data; however, the overlap was set to one second in order to minimize detection delay. Figure 2 illustrates the slicing process.

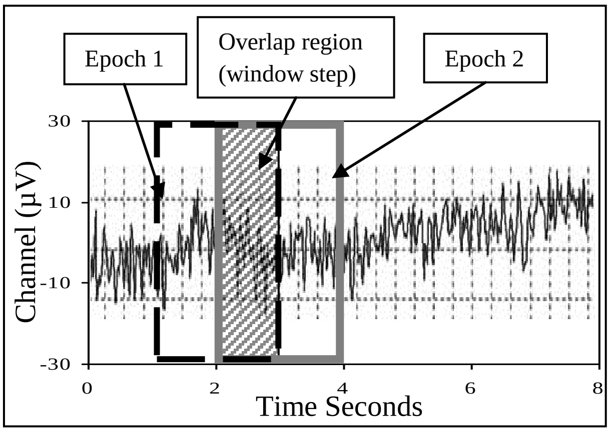


Figure2.Slicing the Signal into epochs.

5.1.2 Spectral features

Each epoch is then transformed from the time domain to the frequency domain using FFT.

Only a subset of the frequencies generated by the FFT is used, frequencies ranging from 0.5 to 24 [11] (which seizure activity mostly in) are used for further processing and the rest is discarded. The subset of frequency bands ranging from 0.5 to 24 are divided into eight overlapping sub-bands. The sub-band size (number of frequency points within each sub-band) is calculated as follows:

$$\boldsymbol{SubBand\ size} = (o.5 < number\ of\ frequency\ points < 24)\ )\ DIV\ numberOfSubBands + overLap\ region\ size$$

And the overlap region between sub-bands is calculated as follows:

$$\boldsymbol{Over\ lap\ Regoin} = (o.5 < number\ of\ frequency\ points < 24)\ )\ MOD\ numberOfSubBands$$

See figure 3, for a description to the process of extracting the spectral features.

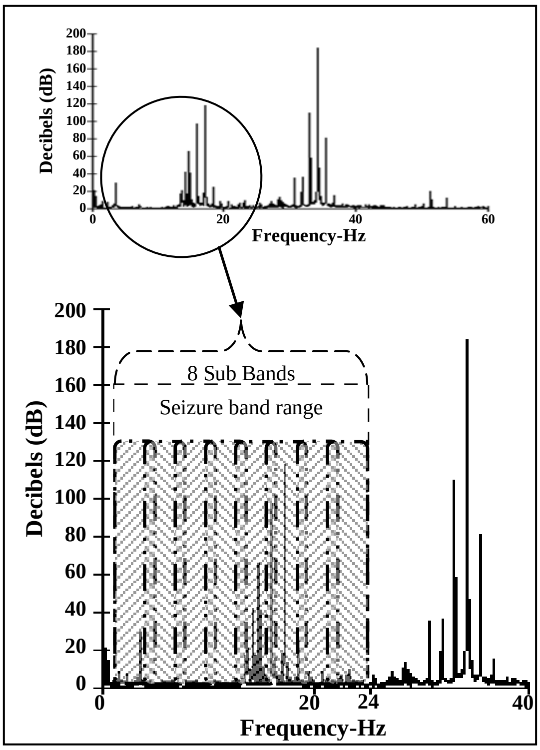


Figure3.Illustrates the process for extracting spectral features

The total corresponding energy for each sub-band is calculated. Thus, each epoch provides eight features per EEG channel.

5.1.3 Complexity features

Nonlinear analysis techniques such as approximate entropy (ApEn) measure the regularity, predictability or chaos of a time series and so higher values of ApEn indicates a more chaotic signal and low values indicate that there is a high likelihood that similar patterns will follow[12]. ApEn proved to be a good discriminate between ictal and inter-ictal data, see figure 4 as describe in [3].

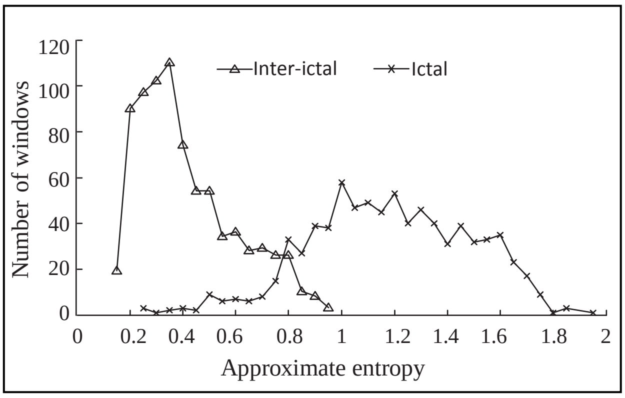


Figure4.Approximate entropy analysis to inter-ictal and ictal EEG data [3].

In this algorithm, ApEn was applied to each epoch of 512 points, 3 times, each of which with different window sizes m. Table 1 shows the parameters used for each of the 3 features extracted.

Table 1. ApEn Parameters per feature

Feature#	r	m
1	$0.2 * \sigma$	5
2	$0.2 * \sigma$	25
3	$0.2 * \sigma$	75

Three ApEn Values with different window sizes give a better complexity measure as well as increase the complexity measure weight in the feature vector.

5.1.4 Feature vector structure

The energies corresponding to each epoch per channel are concatenated with the three ApEn values mentioned earlier thus forming FVi. All FV's corresponding to the same epoch regardless of the EEG channel are concatenated together forming the epoch feature vector, which will be referred to as X, see figure 5

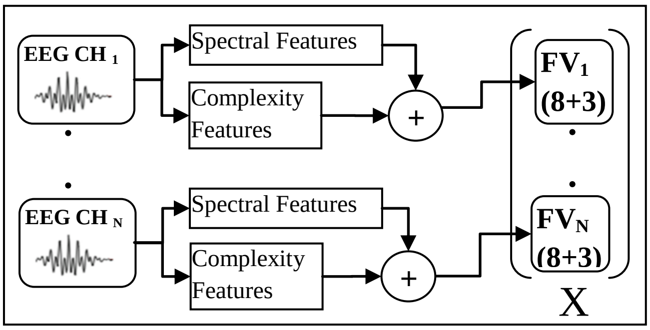


Figure5.Including EEG channels in the feature vector.

Finally, feature vector X holds seizure spectral and complexity features; however, it lacks information on how the seizure progresses which is needed to detect a seizure event with the least possible delay [6].

In order to include seizure progress information, a bigger feature vector W should be formed. W is compromised of every 3 non-overlapping epoch of feature vector X, where W is considered a seizure, if at least one of the feature vectors X (contained in W) was labeled as a seizure, see figure 6 as described in [6].

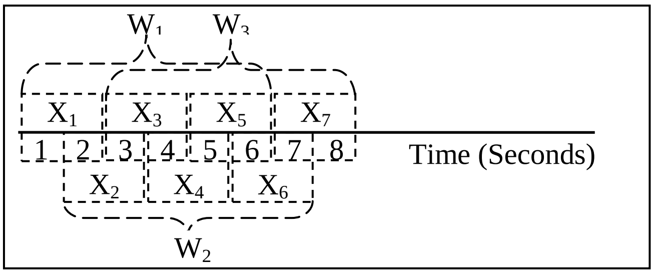


Figure6.Feature vector formations to include seizure progress information.

The feature vector holds spectral, chaos and spatial information which represents all the properties of an epileptic seizure. The Inclusion of discriminating features between ictal and inter-ictal signals is not enough to build a strong seizure detector, this is due to the fact that it is still facing two major challenges namely, the presence of artifacts in the EEG signal and the huge ratio between ictal and inter-ictal data which leads to an imbalanced data set. In the next section, the proposed ensemble based under sampling algorithm which attempts to overcome both priory mentioned challenges will be described.

5.2 Imbalanced data Classification

Highly imbalanced datasets is one of the great challenges in machine learning, especially in a cost-sensitive environment. Class imbalance can be characterized by having a huge ratio between classes in a dataset i.e. the negative class (majority class) has much more samples than the positive class (minority class) which is the class of interest [13].

The class imbalance problem is presented in this paper by the relation between the ictal intervals (in seconds) and the inter-ictal intervals (in seconds). In some patients the ratio between ictal and inter-ictal intervals is up to 1:1570 which is a challenge for building an accurate classifier; see table2 for the ictal to inter-ictal sample distribution within the data set used.

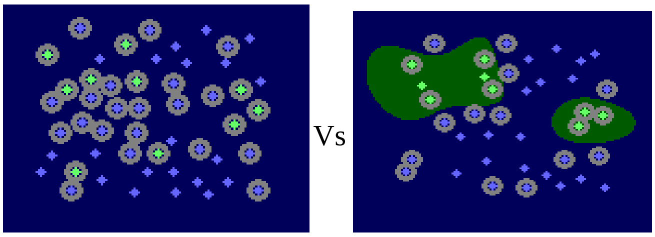
Table 2. Ictal/Inter-Ictal sample distribution

Patient #	Ictal seconds	Inter-ictal seconds	Ratio
Patient#1	442	145988	1:330
Patient#2	172	126959	1:738
Patient#3	402	136806	1:340
Patient#4	378	561834	1:1486
Patient#5	153	240246	1:1570
Patient#6	325	241388	1:742
Patient#7	276	244338	1:885
Patient#8	447	180084	1:402

In this paper, an attempt to overcome the class imbalance problem associated with seizure detection is established by utilizing multiple informed under sampling techniques in addition to the use of ensembles of support vector machines (SVM) which will be explained at a later stage.

SVM minimizes the upper bound on the expected risk by utilizing the structural risk minimization principle; thus providing a reasonable trade-off between the training error and the modeling complication hence providing a superior generalization capability [13].

The modeling algorithm for the SVM constructs a hyper-plane separating positive from negative examples with the maximal margin. On reasonably imbalanced data, SVM is considered a better classifier than other standard classifiers[13] because classification is completed using support vectors only; thus discarding numerous majority samples that are far from the decision boundary without disturbing classification results[13]; However, high-class imbalance can cause an SVM classifier to build a strong estimation bias towards the negative (majority) class leading to a decline in performance by resulting in a large number of false negatives, see figure 7.



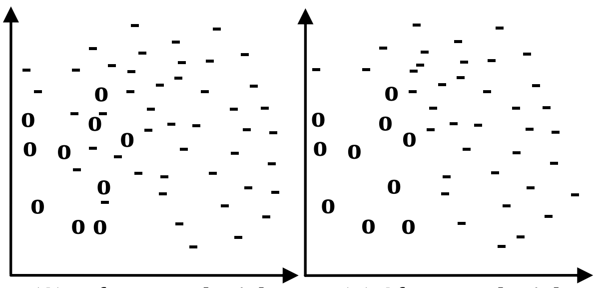
Un-Balanced **Balanced Data**

Figure7.Represents the decision boundary for unbalanced vs. balanced SVM.

This paper proposes the handling of the imbalanced dataset using two approaches. The first approach is data cleansing where the majority side for each tomek link in the training set is found and removed. Then, apply the neighborhood cleaning rule which in turn removes some of the majority class's examples that are either considered noisy or borderline. In the second approach, an ensemble of SVMs is used to facilitate the classification process. The two previously mentioned approaches will be further discussed in the following sections.

5.2.1 Data Cleansing

Data cleansing is the process of removing borderline and noisy data samples from the training set. However; due to the high imbalance ratio between ictal and inter-ictal samples, it is assumed that noisy and border line samples exist only in the inter-ictal data samples; thus, under-sampling the majority class. In order to do so, firstly the tomek links in the training set are identified. A tomek link can be defined as follows: "If Ei, Ej belong to different classes, d (Ei, Ej) is the distance between them. The pair (Ei, Ej) is called a Tomek link if there is no example El, such that d (Ei, El) < d (Ei, Ej) or d (Ej, El) < d (Ei, Ej)."[14]. in this paper, tomek links are used as an under-sampling method accordingly; only samples belonging to the majority class are removed from the training set. This method removes both noise and borderline examples thus reducing the degree unbalance, see figure 8.



(A) Before Tomek Link removal **(B) After Tomek Link removal**

Figure8.Shows the data dispersion before and after the removal of tomek links.

Secondly, the neighborhood cleaning rule (NCL) algorithm [15] is applied to the remaining training samples. This algorithm identifies the three closest neighbors for each sample. If the classification identified by the three closest neighbors differs from this sample's class and that sample belongs to the majority class then this sample will be discarded from the training set; however, if the sample belongs to the minority class, then its nearest neighbors that belong to the majority class are discarded from the training set.

5.2.2 SVM Ensemble

After the completion of the data cleansing process, the training dataset should be free of borderline and noisy samples; however, the ratio of ictal to inter-ictal samples is still very high. Therefore, an ensemble of SVMs is proposed, where the inter-ictal (majority class) data samples are split into fifty random non-overlapping subsets where the number of samples in each subset is equal to five times the number of ictal (minority class) samples. Each of the majority class subsets is then grouped with all the minority class data samples and fed to an RBF kernel SVM. As a result, there should be fifty SVM classifiers for each patient and each of those SVM classifiers was trained with the following parameters epsilon and Gamma equal 0.01 and 0.1 respectively. Finally, votes from each classifier are counted to decide on the class of the unknown sample, in case there is a tie between the classifiers e.g. twenty five of them declared a sample as a seizure and the other 25 declared it as a non-seizure, the classifier should be more biased to towards the seizure class and thus declaring the unknown sample a seizure.

Algorithm1 describes the proposed ensemble method.

Algorithm 1 The SVM Ensemble algorithm

```
procedure SVMEnsembleTraining(S,St ,C )
    ► S ← Seizure samples
    ► St ← Non-seizure samples
    ► C ← Training subsets
    ► |S| + |St | = Number of samples in the training set
    SVM ensembles  E ← 0
►such that  |E|  =|C | = 50
    for i ← 0, |E| do
    Ci   ← non-overlapping random samples from St , |Ci | = 5 * |S|
    Ei  ← TrainSVM (Ci , S) ►  Train an RBF-SVM with C i and S
    end for
    return  E
end procedure

procedure SVMEnsemblePredict(U ,E)
    ► U ← Unknown sample
    ► S ← SVM Ensembles
    return  sgn(∑Ei.predict(U ))
end procedure
```

6. EVALUATION

The proposed seizure detector was tested on eight patients, ages ranging from 1.5 to 22 thus covering all epileptic age groups. The classifiers were trained on 10 hours of continious non-seizure records and less than 70% of the seizures (complete seizures regardless of the seizure length).

Table 3. Seizure samples training/testing distribution

Patient #	#Seizures in training	#Seizures in testing	Training seconds	Testing seconds
Patient# 1	4	3	320	122
Patient#2	2	1	85	087
Patient#3	4	3	201	201
Patient#4	2	1	324	054
Patient#5	7	3	72	81
Patient#6	2	1	234	091
Patient#7	2	1	207	069
Patient#8	4	3	268	179
Average	*3.3*	*2*	*213*	*110.5*

The detector performance was judged based on three metrics specificity, sensitivity and latency.

Sensitivity measures the ability of a detector to successfully identify ictal records i.e. the probability of successfully detecting a seizure, sensitivity is calculated as follows:

Let the number of correct ictal classifications ⇒ nc.

Let the number of incorrect ictal classifications ⇒ ni.

$$Senstivity = \frac{nc}{nc + ni}$$

Specificity measures the ability of a detector to successfully identify inter-ictal records i.e. the probability that a detector will declare a seizure when there isn't one, specificity is calculated as follows:

Let the number of correct inter-ictal classifications ⇒ inc;

Let the number of incorrect inter-ictal classifications ⇒ ini;

$$Specificity = \frac{inc}{inc + ini}$$

Latency measures the time between the start of a seizure episode and the detection of its presence by the detector.

Table 4 shows the specificity, sensitivity and latency achieved by the seizure detector for each patient.

Table 4. Specificity, sensitivity and latency per patient

Patient #	Sensitivity percent	Latency in seconds	Specificity %
Patient# 1	98.36%	0.67	99.68%
Patient#2	77.01%	4.00	99.66%
Patient#3	81.59%	3.67	99.20%
Patient#4	81.48%	1.00	99.78%
Patient#5	63.85%	5.67	99.94%
Patient#6	92.31%	3.00	93.40%
Patient#7	100.00%	0.00	90.64%
Patient#8	91.06%	1.00	99.97%
Average	85.71%	2.37	97.78%

7. CONCLUSION

The algorithm was tested on 441 hours of inter-ictal data and 14 minutes of ictal data providing a mean specificity rate of 97.78% and mean sensitivity rate of 85.71% overall the entire algorithm detected 16 of the 16 test seizures with a mean delay of 2.37 seconds. when comparing this algorithm to other algorithms with comparable results, it should be noticed that these algorithms were evaluated on intracranial EEG data, which by nature the signal sources are neither effected by noise nor other internal and external artifacts, furthermore the risk factor and complication with recording intracranial EEG is high [16], moreover these algorithms were tested on 5 patients only, where their dataset size was 4720 seconds, in contrast to the 441 hours of testing data used in the evaluation of this algorithm thus rendering it more suitable for use as it was tested on a huge dataset and trained with highly imbalanced data.

8. REFERENCES

[1] M. G. Tsipouras andD. I. Fotiadis A. T. Tzallas, "Automatic Seizure Detection Based on Time-Frequency," Hindawi Publishing Corporation Computational Intelligence and Neuroscience, vol. 2007, 2007.

[2] T. He1 L. A. Smith and L. Tarassenko P. E. McSharry, "Linear and non-linear methods for automatic seizure detection in scalp electro-encephalogram recordings," Med. Biol. Eng. Comput, vol. 40, pp. 447–461, 2002.

[3] Hsu-ChuanWang and Wan-Lin Chang Sheng-Fu Liang, "Combination of EEG Complexity and Spectral Analysis for Epilepsy Diagnosis and Seizure Detection," Hindawi Publishing Corporation EURASIP Journal on Advances in Signal Processing, vol. 2010, 2010.

[4] John Guttag Ali Shoeb, "Application of Machine Learning to Epileptic Seizure Onset Detection," in 27th International Conference on Machine Learning (ICML), Haifa, 2010.

[5] Sung-NienYu Kai-Cheng Hsu, "Detection of seizures in EEG using subb and nonlinear parameters and genetic algorithm," Computers in Biology and Medicine, vol. 40, no. 1, p. 823=830, Sep. 2010.

[6] ˙Ibrahim Öztura, Mehmet Kuntalp, Barıs¸ Baklan, and Cüneyt Güzelis Nurettin Acır, "Automatic Detection of Epileptiform Events in EEG by a Three-Stage Procedure Based on Artificial Neural Networks," IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, vol. 52, no. 1, pp. 30-40, 2005.

[7] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 101(23):e215-e220

[8] Jaakko Malmivuo and Robert Plonsey, Bioelectromagnetism - Principles and Applications of Bioelectric and Biomagnetic Fields. New York: OXFORD UNIVERSITY PRESS, 1995.

[9] Richard P. Brenner Lawrence J. Hirsch, Atlas of EEG in Critical Care.: Wiley-Blackwell, 2010.

[10] Interview with Prof.Charles Ribak,Irvine School of Medicine,The University of California

[11] Ali Shoeb, Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment. PhD Thesis, 2009.

[12] George B. Moody. Approximate Entropy (ApEn). [Online]. http://physionet.org/physiotools/ApEn/

[13] Yan-Qing Zhang, Nitesh V. Chawla and Sven Krasser Yuchun Tang, "SVMs Modeling for Highly Imbalanced Classification," JOURNAL OF LATEX CLASS FILES, vol. 1, no. 11, 2002.

[14] Miroslav Kubat Sareewan Dendamrongvit, "New Frontiers in Applied Data Mining:Undersampling approach for imbalanced training sets and induction from multi-label text-categorization domains," Lecture Notes in Computer Science, vol. 5669, pp. 40-52, 2010.

[15] Jorma Laurikkala, "IMPROVING IDENTIFICATION OF DIFFICULT SMALL CLASSES BY BALANCING CLASS DISTRIBUTION," 2001.

[16] Birkett J, Byth K, Dexter M, Somerville E, Gill D, Chaseling R, Fearnside M"Risk factors for complications during intracranial electrode recording in presurgical evaluation of drug resistant partial epilepsy," Acta Neurochir (Wien), vol. 151, no. 1, pp. 37-50, January 2009.