UCONN
UNIVERSITY OF CONNECTICUT

Towards Automatic Seizure Onset Detection in Epilepsy Patients: A Systems Approach

Edward Novikov
University of Connecticut
Department of Electrical & Computer Engineering

September 6th, 2016

Supervisors:
Shalabh Gupta, PhD, Electrical and Computer Engineering
Sabato Santaniello, PhD, Biomedical Engineering
Maria Gordina, PhD, Mathematics

This Page is Intentionally Left Blank

Abstract

This research has two main objectives: one, is to design, implement, and validate a support-vector-machine (SVM)-based software tool for the automatic detection of onset time of seizure events in subjects with drug-resistant epilepsy (DRE) and second, is to present a systematic review of the scientific literature regarding the application of machine learning principles and algorithms in epilepsy.

Regarding the first objective, the tool operates on time series of feature vectors computed from multichannel intracranial EEG (iEEG) recordings collected continuously across one week from DRE epilepsy patients. This work builds upon mathematical and computational tools for interpreting iEEG signals to answer the question: given promising features, can seizures be detected at the initiation stage in real-time?

Regarding the second objective, outlined is a comprehensive assessment of the algorithms proposed in the last 10 years to process intracranial electroencephalographic recordings (iEEGs) aiming to (i) detect an incumbent seizure episode or (ii) identify the brain region where seizures naturally initiate (i.e., the "epileptogenic zone", EZ).

Machine learning principles have been largely applied to epilepsy in recent years but, despite encouraging "proof-of- concept" preliminary results on small samples of data, very few methods have been translated into tools of practical use for clinicians. This study is instrumental toward the identification of the limitations of the current approaches with respect to the specific nature of epilepsy, which is often associated non-stationary, heterogeneous, and abrupt temporal dynamics in the iEEG time series.

Table of Contents

1 Introduction

The first human electroencephalogram (EEG) recording is documented to have been administered by Hans Berger in 1929 [6]. EEG captures time-varying electrical potentials in the brain produced by groups of neurons [13]. Since this monumental historical event, EEG has become widespread in the recording of brain activity and analysis of neurological disorders, particularly epilepsy. Epilepsy affects over 60 million people worldwide with chronically recurring, abrupt, and severe seizures. Seizures are finite-time episodes of disturbed cerebral function induced by abnormal, synchronous, and excessive electrical discharges of large groups of cortical neurons. Seizures lead to debilitating phenomena (convulsions) or remain clinically unapparent, last seconds to minutes, and can be followed by hours of confusion, psychosis, or sensory impairment. Currently, about one in a hundred people are diagnosed with epilepsy, making it the second most common neurological disorder second only to stroke. However, the origins of the disorder remain largely mysterious [22]. About 66% of diagnosed patients achieve acceptable seizure control with medication and approximately 10% are cured through surgical resection, but there are no sufficient treatments for the remaining 25% of patients [22]. Despite the demand for answers, epilepsy is receiving minimal attention and funding compared to other neurological disorders.

The unpredicted nature of epileptic seizures has a debilitating effect on patients' everyday lives and carries an unimaginable psycho-social effect [8]. One of the most unbearable aspects of the disease is the unforeseen and unpredicted way seizures erupt. Patients usually feel a sense of helplessness and failure, are forced to experience drastic lifestyle changes, and are in constant fear of serious injury [22]. Thus, there exists a critical demand to improve the quality of life for patients suffering from intractable drug-resistant epilepsy.

The goal of detection of epileptic events is to extract information, i.e. features, on the onset time of the seizure event (i.e., the time after which the brain unequivocally transitions into a seizure

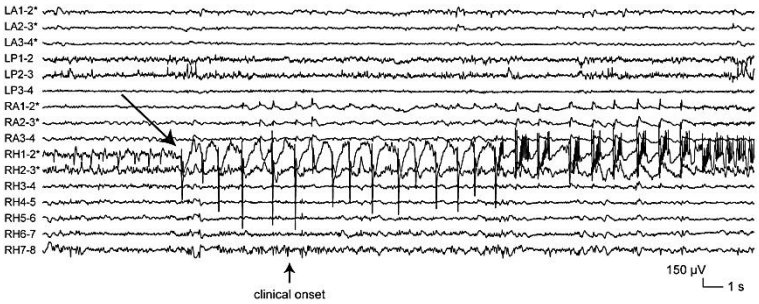


Figure 1 – Displays starting point of clinical seizure onset with low frequency high amplitude spikes. Image provided from the *Brain* journal of neurology [43].

condition, see Figure 1) which is specific to a particular patient. Epilepsy is divided into two subcategories based on the initiation of seizures in the brain and how propagation occurs. In one type, called primary generalized epilepsy, seizures begin with a pervasive electrical discharge that involves the whole brain. The second type, known as partial seizures, begins as an electrical discharge in a segregated area of the brain. Partial seizures can be further classified based on the brain region involved – frontal, temporal, parietal, or occipital lobe epilepsy [23]. However, such generalized classifications fail to accurately and completely define any particular case of epilepsy. For example, research has shown that epilepsy classified as generalized does have focal onset points from which propagation begins spontaneously [24]. Thus, a central aim is detecting and localizing epileptogenic foci, which will aid in more accurately diagnosing particular cases of epilepsy and in better understanding the disorder as a whole.

There are two analytic descriptions of brain architecture: functional segregation views specialized anatomical brain regions in a segregated fashion; functional integration includes the functional interaction between the various areas of the brain [25]. The functional integration perspective is essential to the study of epilepsy and can be characterized by means of functional connectivity, which quantifies the statistical dependencies between spatially distinct neurophysiological events [23]. Formal mathematical definitions have been developed to define directed versus undirected and linear versus nonlinear correlations, giving valuable information about the degree of functional connectivity in the time or frequency domain. The calculation of these measures is based on either amplitude or phase and can be bivariate or multivariate based on the number of signals, or variables, which are analyzed at a time. Four main categories of functional

connectivity measures are as follows – correlation and coherence, phase synchronization, information-based, and Granger causality measures [23].

Applying these functional connectivity measures to a patient's EEG signals will give information about brain connectivity before, during, and after a seizure. This approach treats the brain of an epileptic patient as a complex system, or mathematical abstraction for a physical structure composed of various parts or subsystems; therefore, identifying transitions in the complexity of the internal states of the system over time becomes a tractable computational problem [26]. Changes in certain functional connectivity measures and behaviors, such as the synchronization and nonlinear dynamics of the system, may allow prediction of an upcoming seizure [23].

To assist the subset of patients for whom antiepileptic drugs do not have a positive effect, many scientists and engineers are investigating techniques in the prediction of epileptic seizures [8]. The goal of this current research is to implement diagnostic alerting systems capable of providing therapeutic measures [7]. At the core of such systems is a seizure detection algorithm that can detect imminent seizures early in an accurate way. Identifying key events that occur close to seizure onset is crucial in accurately determining seizure behavior. Understanding the role of seizure onset times can assist in assessing early changes with electrographic and neurophysiological data that may predict seizure and/or discern transitions from seizure to non-seizure states. Furthermore, in therapeutic systems, a central question is choosing a time to administer therapy that will result in most effective treatment relative to clinically marked seizure onset [12].

1.1 Clinical Need and Challenges

The current gold standard for analyzing EEG signals and labeling seizure onset times is by a consensus of trained encephalographic clinicians. Manually labeling onset times for multiple seizures from a large dataset can be a monotonous and highly inefficient task [12]. While it is essential to develop a diagnostic tool to aid clinicians in monitoring and localization of abnormal electrical brain activity, the utilization of such a tool offers several challenges. First, in a clinical

setting, EEG recordings generate an extensive dataset that can only be analyzed by a human expert neurophysiologist. With such a tremendous amount of data, the task is close to impossible. Second, there is no standard reference for EEG recordings analyzed manually by clinicians [6]. Therefore, a consensus of three trained physicians is required to correctly annotate an EEG reading. Qualified physicians often disagree on the analysis, culminating in an estimate in marking seizure onset. In such cases, a trustworthy seizure detection algorithm is paramount to warn epilepsy patients and assist clinicians in diagnosis and earlier intervention [6].

A topic of increasing interest is neurostimulation, an alternative to surgical resection. Still under clinical trials, neurostimulation has provided encouraging results [23], but its effectiveness critically depends on the electrode placement, the seizure's morphology, and, most importantly, the seizure's onset time, i.e., stimulation is mostly effective when administered immediately prior to or at the onset of the seizure. It is critical to develop a tool that accurately detects the seizure onset from sequential electroencephalographic (EEG) measurements. Although several algorithms have been introduced, automatic online seizure detection is still an open problem. The algorithms proposed thus far have high sensitivity (large number of true positives) but low specificity (large number of false positives), which ultimately limits their clinical use.

Current research methods aim to develop novel computational tools to detect the onset of epileptic seizures, which will assist clinicians in earlier intervention and, eventually, in real-time treatment via neurostimulation. The detection of seizure events can also prove useful in developing closed-loop intervention strategies, review the electrographic activity in a patient for sake of diagnosis, and ultimately, improve the patient's quality of life.

1.1.1 Performance Metrics

There are several performance metrics used to evaluate and compare seizure detection algorithms in the literature. They are defined below:

(1) Sensitivity: The ratio of true positives to the total number of positives. $\frac{TP}{TP+FN}$

(2) Specificity: The ratio of true negatives to the total number of negatives. $\frac{TN}{TN+FP}$

(3) Total Accuracy: $\frac{TP+TN}{TP+FN+TN+FP}$, where TP = true positives, FN = false negatives, TN = true negatives, and FP = false positives.

(4) False Alarm Rate: $\frac{FP}{hr}$

(5) Latency: The delay between expert identified electrographic onset and the time a detection algorithm indicates a seizure.

While there have been many published seizure detection algorithms with high performance indicators in the research realm, implementations of seizure detections algorithms in the clinical setting has been less encouraging. Reasons for this failure can be explained by the following factors that inhibit the detector's performance. First, epileptic seizure morphology has highly varied fluctuations between individual patients. Thus, a generalized patient independent algorithm with clinically acceptable performance across a large patient population can perform mediocrely with respect to specific patients who are not adept to the design criteria. Second, epileptic seizure morphology of a particular patient can exhibit variations over time due to a change in physiological state or recording electrodes. Thus, a patient's individual seizure characteristics can vary with time due to internal and external changes. Third, epileptic seizure activity can be imitated by EEG recordings polluted with physiological and non-physiological artifacts [6]. With respect to the first challenge, about 70% of studies analyzed in this paper [1-9, 10-14] have demonstrated a wide improvement in seizure detection results using patient-specific algorithms as opposed to patient-independent algorithms. However, these methods exhibit little to no improvement in challenges two and three; they require a classifier to be trained specifically on a per patient basis, which is a burden in the clinical setting. There exists a need to develop a patient-independent detector that adapts well to time varying seizure morphologies across patients and resists contamination of EEG artifacts during recordings.

2 Review of Literature

This study includes peer-reviewed research articles that satisfy the following criteria: first, they propose machine learning methods for seizure detection or EZ localization; second, the proposed methods are tested on iEEG time series collected in vivo from subjects with drug-resistant epilepsy; and third, the articles have been published in the last 10 years.

2.1 SVM Technology

The Support Vector Machine (SVM) is a supervised machine learning algorithm that can be trained to solve binary decision classification problems, i.e. discriminate between 2 separable high-order data classes via non-linear decision boundaries [27]. An SVM takes as input a feature set representing the highly non-linear non-stationary EEG signals; it is trained on the feature vectors (in many cases a combination of temporal and spectral features in each feature window) labeled a priori as belonging to either seizure or non-seizure classes. Based on the training data, the SVM formulates a decision boundary hypersurface by maximizing a distance metric between the two brain-state classes in higher-dimensional space. The maximization of distance between both classes of data results in extremely high classification accuracy for novel, unseen, data.

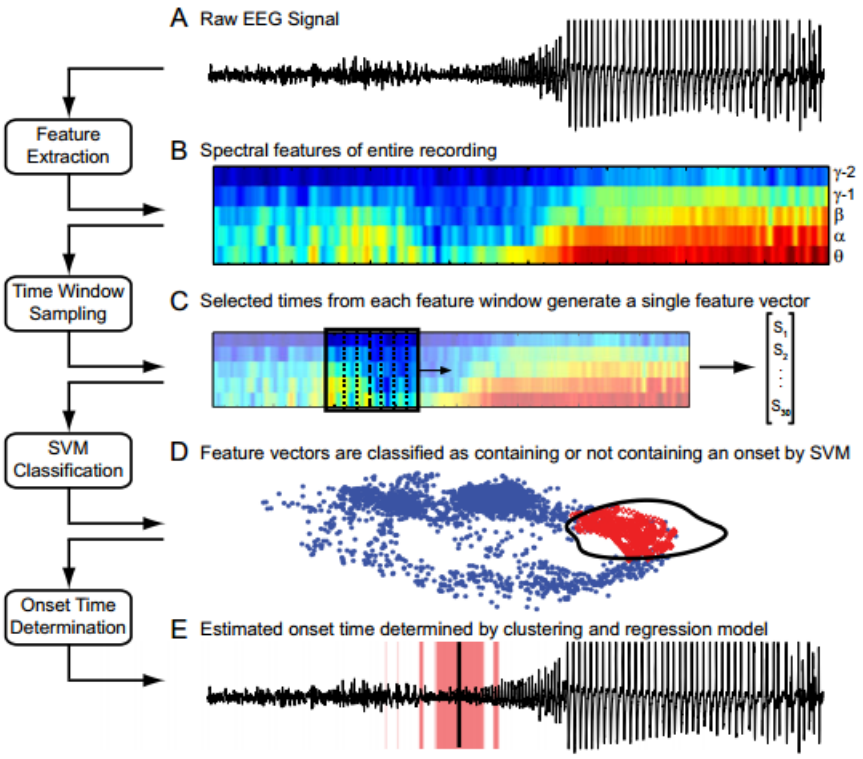


Figure 2 – A four-component seizure onset detection architecture consisting of feature extraction, time window sampling, support vector machine classification, and onset time determination [12].

Essentially, SVM compares class 1 (onset-containing) feature vectors and class 2 (non-onset containing) feature vectors to derive an optimal decision boundary that characterizes the data via a quadratic optimization problem. Unlike its counterpart, neural networks, the SVM does not converge on local minima because of its robustness in handling different sized training sets [12]. Additionally, SVM has been shown to exhibit excellent performance in classification of neurological data, especially pertaining to EEG recordings epileptic seizure onset detection [12].

2.1.1 Technical Background

Based on Figure 3 below, the classification phase of the general seizure detection algorithm maps a feature vector to one of two possible classes, either seizure or non-seizure. The main motivations for utilizing SVM classification are described next. First, SVM classification is ideal for real-time applications because the training phase, the majority of the computational cost, is completed offline. Second, through thorough study and analysis, SVMs predict well in theoretical and practical settings. Third, the SVM technique can solve big-data, high-dimensional, problems due to the various optimization techniques developed [11].

Figure 3 – A four-component seizure onset detection architecture consisting of feature extraction, time window sampling, support vector machine classification, and onset time determination [11].

We now present the generic binary classification problem that constructs a training classifier to separate two disconnected sets of points in the standard Euclidian space. First, let the training set be defined by $TS = \{(u_i, l_i), u_i \in \boldsymbol{R^p}, l_i \in \{-1, 1\}, i = 1, \ldots, P\}$. We assume that there exists a hyperplane $H(w, b) = \{u \in \boldsymbol{R^n} : w'u + b = 0\}$, i.e. the data points are linearly separable, with

$$w'u_i + b \geq 1 \ \forall i : l_i = 1$$

$$w'u_i + b \leq -1 \ \forall i : l_i = -1.$$

We have that the standard Euclidian distance metric $\rho(w, b)$ of the hyperplane $H(w, b)$ is defined as the distance from the hyperplane to the closest training point(s), with

$$\rho(w, b) = \min_{i=1,2,\ldots,P} \frac{|w'u_i + b|}{\|w\|}.$$

As shown in Figure 4, the linear SVM method attempts to choose the optimal hyperplane, i.e. the hyperplane that exhibits maximum distance between disjoint sets. In SVM, the basic idea behind the training phase is based on concepts from statistical learning theory [27]. Due to the optimization techniques underlying SVM, it is assumed that the classification error on novel datasets is minimized [11]. Thus, SVM is an excellent in developing predictive models, as needed in epileptic seizure detection. SVM solves the following quadratic programming problem to obtain the optimal hyperplane for discriminability of data points,

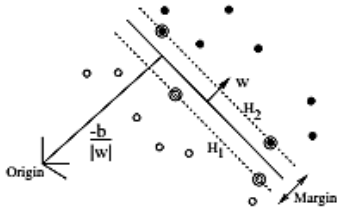


Figure 4 – Linear separating hyperplanes with linear SVM. The support vectors are circled. Image provided by A Tutorial on SVM for Pattern Recognition [28].

$$\min_{w \in \boldsymbol{R}^p, b \in \boldsymbol{R}} \frac{1}{2}\|w\|^2$$

$$\text{subject to } l_i(w'u_i + b) \geq 1, \ i = 1, \ldots, P.$$

In practice, linear SVM classifiers may produce poor results if the data is not linearly separable [11]. Thus, it most often is the case that nonlinear classifiers are needed. The principle behind nonlinear SVMs is different when compared to linear SVMs. A nonlinear SVM maps input feature vectors to another space known as feature space H. Mathematically the transformation is given by, $\varphi : \boldsymbol{R}^n \rightarrow H$. The optimal hyperplane is determined in the space H and the new quadratic programming problem becomes,

$$\min_{w \in \boldsymbol{R}^p, b \in \boldsymbol{R}, \xi \in \boldsymbol{R}^p} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{P} \xi_i$$

$$\text{subject to } l_i(w'\varphi(u_i) + b) \geq 1 - \xi_i, \ i = 1, \ldots, P,$$

with $\xi_i \geq 0, \ i = 1, \ldots, P$. The following terms, $\xi = [\xi_1, \xi_2, \ldots, \xi_P]'$ and $\xi i, \ i = 1, \ldots, P$, are slack variables, $\sum_{i=1}^{P} \xi_i$ acts as an upper bound on the training error, and C >0 varies the inverse relationship between $\rho(w, b)$ and the training error [11].

Without considering the map $\varphi : \boldsymbol{R}^n \rightarrow H$ in explicit form, it is possible to create a nonlinear SVM classifier by only observing the inner product kernel on *H*. Mathematically, let *X* be a subset of $\boldsymbol{R}^p$. Then the function $K : X \, x \, X \rightarrow \boldsymbol{R}$ is a kernel if

$$K(x, y) = \langle \varphi(x), \varphi(y) \rangle \ \forall \, x, y \in X,$$

where $\langle \varphi(x), \varphi(y) \rangle$ denotes the inner product, $\varphi : X \rightarrow H$ is a function, and *H* is a Euclidian inner product space.

2.1.2 Kernel Types for Epilepsy Diagnosis

This section outlines the main characterization of supervised kernel functions that have been used in epilepsy detection. The most widely used technique is standard SVM with kernel functions as described in the previous section. Kernels are employed in order to map input feature vectors into higher-dimensional feature spaces with low computational costs [18]. The most prevalent kernels utilized in the literature are the Gaussian radial basis function (RBF) kernel and the exponential radial basis function (ERBF) kernel. About 70% of the papers [3-6, 8-13, 17-18] analyzed in this study use the radial basis function kernel in building the SVM classifier. Whereas the rest of the state of the art uses roughly 15% linear kernels, 10% polynomial kernels, and 5% other unique kernels such as extreme learning machine (ELM) [15].

The RBF is likely the preferred kernel due to its ability to linearly map input data with nonlinear attributes and class labels into higher dimensional feature spaces, a mapping the linear kernel is not capable of completing. Additionally, the RBF kernels only have one adjustable tuning parameter for the configuration of the classifier, unlike polynomial kernels, which have multiple parameters [18]. The formal mathematical expressions for the RBF and the ERBF are given below,

$$K_{RBF}(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$

$$K_{ERBF}(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|}{2\sigma^2}\right)$$

where σ is the control radius parameter that needs proper adjustment [18].

The next type of SVM is called a least-squares SVM (LS-SVM). The optimization problem is changed to return a system of linear equations. To accomplish this, the least squares cost function is solved with equality constraints. Another type of SVM is the S-SVM, which possesses a smooth unconstrained quadratic programming problem, similar to the standard SVM. In S-SVMs, two modifications to the standard SVM are adopted. First, the distance $\rho(w, b)$ between the two planes is maximized with respect to the direction *w* and location relative to the origin *b*. Second, the soft margin error bound ξ_i was minimized using the 2-norm squared. This new SVM formulation is not twice differentiable anymore; therefore smoothing techniques should be employed before use [18]. Additionally, P-SVMs employ classification a little differently compared to standard SVM. In the P-SVM approach, data points are classified by being assigned one of the two closest parallel planes. This SVM implementation can be thought of as a hybrid of S-SVM and LS-SVM, where the cost function is identical to S-SVM and the constraints are replaced with equalities [18]. Next, is the L-SVM, which achieves greater efficiency in the training process by means of an iterative approach. The cost function is the same as in the standard SVM implementation, except that the 2-norm is used instead of the 1-norm [18].

Finally, the relevance vector machine (RVM) is discussed. Conventional SVMs minimize classification error and produce a simple binary decision without modeling the data distribution. There may be scenarios where the binary classification estimate is not sufficient and a probabilistic output is preferred. This is the principle behind RVMs. An RVM has the same function as an SVM, except it is implanted within a Bayesian network. Thus, the RVM outputs a probability that the data point(s) are in a certain class. It is important to note that the error estimate in an RVM is assumed to be a normally distributed random variable with zero mean and nonzero variance [18]. RVMs have not been widely used in the area of seizure detection. However, classifying input data based on associated probabilities per class combined with smoothing and collar techniques has the potential to drastically reduce the false positive per hour (FP/hr) rate. One paper found in the literature that was successful in employing RVMs is [29]. A sensitivity of 92.94% and a specificity of 97.47% were achieved.

The tables below show results of the most widely used RBF kernels with discrete wavelet transform (DWT) and Lyapunov exponents as features. Table 1 displays the average accuracy achieved by each kernel machine type with respect to the best-configured classifiers and Table 2 conveys the percentage of the best-configured classifiers that yielded 100% total accuracy.

Kernel type	SVM	LS-SVM	RVM	L-SVM	S-SVM	P-SVM
RBF	40%	40%	40%	10%	20%	5%
ERBF	35%	35%	25%	20%	15%	5%

Table 1 – The proportion of the best-calibrated models that have achieved 100% total classification accuracy [18].

Kernel type	SVM	LS-SVM	RVM	L-SVM	S-SVM	P-SVM
RBF	0.058 ± 0.121	0.059 ± 0.120	0.075 ± 0.137	0.089 ± 0.137	0.064 ± 0.125	0.076 ± 0.122
ERBF	0.061 ± 0.125	0.060 ± 0.122	0.064 ± 0.130	0.089 ± 0.138	0.069 ± 0.130	0.083 ± 0.130

Table 2 – The average classification accuracy exhibited by each kernel machine taking into account the best-calibrated models produced for associated kernel machine configurations [18].

2.1.3 Collections of Papers

The literature search for state of the art publications was done using the PubMed database from the National Center for Biotechnology Information (NCBI). NCBI is the most respected resource containing citations for biomedical literature; their mission "is to develop new information technologies to aid in the understanding of fundamental molecular and genetic processes that control health and disease [30]." Publications found on the PubMed database are well known by the scientific community and clinicians alike. These are research methods that have been examined in the clinical setting and/or whose ultimate goal is to be integrated in hospital environments.

Throughout the literature search for SVM methods in epilepsy detection and localization, it was verified that all pertinent articles satisfied the criteria of being tested on iEEG time series and were published within the last 10 years. Various combinations of keywords were used to extract papers. These are, in no particular order, seizure detection, seizure localization, SVM, support vector machine, epilepsy, iEEG, intracranial EEG, automated seizure detection, classification, machine learning, automated epilepsy localization, and similar articles recommended by the PubMed database. The following were the most common combinations of searches: seizure detection SVM; seizure detection support vector machine; epilepsy localization support vector machine; iEEG seizure detection support vector machine; iEEG epilepsy localization support vector machine; SVM epilepsy localization, detection; SVM intracranial EEG epilepsy; SVM for epilepsy detection, automated seizure detection SVM, SVM classification epilepsy; machine learning intracranial EEG epilepsy, seizure detection machine learning, automated epilepsy localization; and similar articles recommended by PubMed.

All papers were saved into a Collections folder in PubMed; the union of the searches produced 274 distinct papers. Papers matching the criteria above were extracted and rigorously analyzed further. The resultant number of papers applicable to the assessment of the state of the art is 20 distinct papers. In other words, 7-8% of the SVM papers fit the conditions set forth in the introduction. It can be

concluded right away from the paper extraction that few SVM methods are tested and validated on intracranial EEG datasets.

2.1.4 Detection Methods

A multitude of SVM detection methods can be implemented based on various feature sets, SVM detector architectures, channel policies, and preprocessing and post-processing techniques. Here we describe the two main SVM detector architecture methods that have been successfully implemented in the literature. The first technique [10] separately preprocesses and splits EEG data into epochs. An epoch usually refers to a short time window or segment of EEG data used for processing. Then a feature vector is extracted from each epoch and sent as input to a single SVM classifier. The output of the SVM classifier can produce either a data pointwise decision, i.e. a -1 for non-seizure class and a 1 for seizure class, or a probability of seizure during the epoch. SVM outputs can be post-processed using various advanced filtering techniques, which will be discussed in detail later. Ultimately, the SVM outputs are compared to a threshold value, i.e. 10 channels in seizure, to determine a final multi-channel binary decision. The following papers followed the above methodology using a single SVM per channel [5, 6, 9, 10, 14]. A representation of the aforementioned SVM seizure detector is shown in Figure 5.

The second technique seen in the literature is described below. The detection system passes a 2-4 second epoch from each of the *N* intracranial EEG channels to a feature extractor.

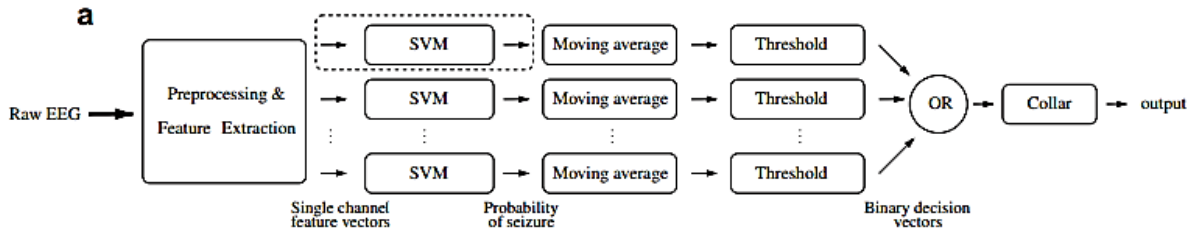


Figure 5 – Architecture of an SVM seizure detection system [10].

The feature extractor computes the best characterizing features representing the signals most efficiently and groups these features across all channels into one extended feature vector. Features from all channels are grouped into one large feature vector in order to observe spatial relationships between channels [19]. The feature vector is then passed through a trained SVM classifier and assigned to either a seizure or non-seizure class. An overall seizure is declared upon meeting a certain temporal constraint. For example, a seizure can be declared once 3 consecutive epochs are marked as seizure; requiring seizure activity to last for 6-12 seconds prior to declaring onset has been shown to decrease the false positive rate [19]. The papers implementing detectors with one SVM across all channels include [3, 4, 7, 8, 11, 12]. Figure 6 shows the second SVM detector architecture technique.

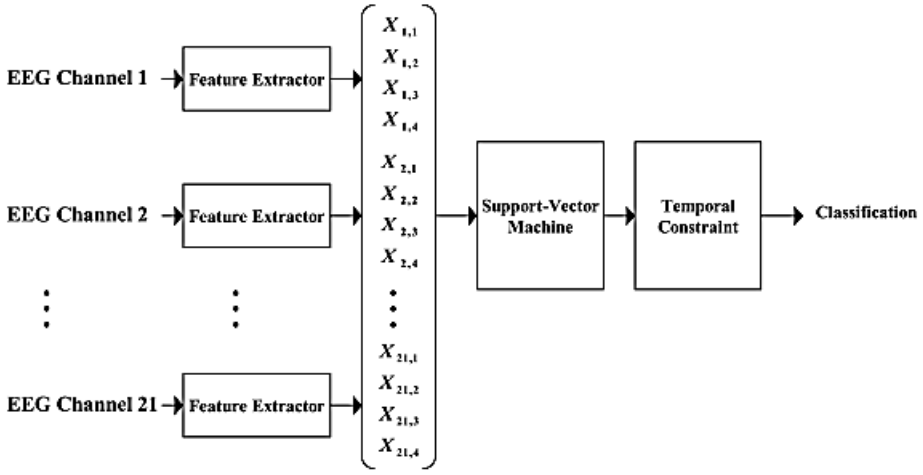


Figure 6 – Architecture of an SVM seizure detection system with one large feature vector [19].

Additionally, a few papers execute their methods on a single intracranial EEG channel. Reasons for analyzing a single channel may include increasing computational efficiency or that channel selection has been employed prior to the classification phase of detection. The approach in [2] uses wavelet decomposition along with a single channel to validate its algorithm. Channel selection is a critical problem in the design of reliable SVM detectors and is thus discussed next.

We first list the potential problems with channel selection. First, the computational cost of the seizure detection algorithm varies directly with the increasing number of channels. Second,

computational load can be reduced if decreasing the number of channels is done by studying pertinent features. Third, the levels of reliability and robustness decrease if using a single channel. Fourth, the problem of potential over-fitting of data can occur by using too many channels [8]. The first two issues are concerned with computational power and time, an important characteristic if another algorithm can accomplish the equivalent task with greater efficiency. Problem three ascertains that single channel selection may work well for some patients, but lack in accuracy for others. Finally, the fourth problem deals with classifier discrimination between data points. A greater number of channels imply a greater number of features for classification. This may cause a problem during seizure onset because if the seizure has not yet generalized across many channels, then features calculated from non-focal channels will have similar values to features found during the pre-ictal state [8]. This will cause overlaps between seizure and non-seizure data, potentially confusing the classifier and producing false negatives.

Previous works in channel selection have utilized the clinician's judgement or on calculated features, instead of on raw iEEG data. Basing the relevant channels on the clinician's verdict may produce a bias in the decision and highly depends on their level of attentiveness [8]. One possible simple automatic channel selection method employed in the literature is to compare all channels based on a simple feature [8]. This method compared selection of channels based on different criteria such as clinician's choice of channels, variance among channels, difference in variance of the channels, entropy of the channels, and random selection of channels. It is observed that channel selection based on variance produced the best results. The n channels with maximum variance are chosen for further analysis and feature extraction. It can be observed from Table 3 below that automatic channel selection based on variance is comparable to the clinician's judgement.

Method	*Focal channels*	*doctor*	*var*	*dvar*	*ent*	*non-foc*	*Randomly*
Sensitivity (%)	92.4 ± 8.8	95.7 ± 9.0	95.4 ± 11.2	95.3 ± 14.9	91.7 ± 16.6	62.2 ± 5.7	61.8 ± 24.1
FDR (/h)	0.16 ± 0.005	0.18 ± 0.006	0.15 ± 0.005	0.18 ± 0.051	0.12 ± 0.003	0.139 ± 0.006	0.35 ± 2.48

Table 3 – Comparison of channel selection methods for choosing optimal channels. Results obtained after 30 distinct training models and reported as mean ∓ SD. [8].

Another paper that uses channel selection is [21], in an attempt to minimize the number of channels for continuous monitoring of EEG signals in an ambulatory setting. This added technique may allow patients to return home with a personal monitoring device. By reducing the number of channels, this method manages to the lower power consumption of the algorithm. It was shown that the computational burden was reduced by 65% with no negative effect on system performance.

Overall, channel selection (also known as sensor or feature selection) is a powerful tool that can improve the real-time component of a seizure detection system while requiring less computational resources. Furthermore, channel selection provides information on the localization of seizure onset, crucial knowledge for the patient and clinician alike. From this thorough literature review it is concluded that the majority of papers focus on seizure detection and very few have implemented robust techniques for seizure localization. More research is needed in this area, both in selecting channels and attempting to keep the feature vectors small.

2.1.4.1 Datasets

Various datasets were used for testing and validation of the proposed methods. The main datasets found in the literature were the Freiburg Database [2-5, 9, 11], the BONN Database [1, 6, 13, 16-18], and the Flint Hills Scientific (FHS) Database [8]. A few other datasets including recordings from the Neonatal Intensive Care Unit Cork University Maternity Hospital, Cork, Ireland [11], from surgical patients requiring invasive monitoring at Massachusetts General Hospital [7], and from recordings collected for clinical

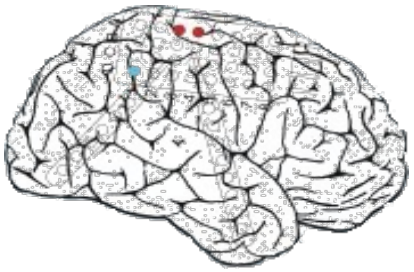


Figure 7 – Representation of grid-, strip-, and depth- electrodes for iEEG recording in the brain [31].

investigation of an implantable neurostimulator [12]. It has been observed that according to the pool of analyzed papers, roughly 30% utilize the Freiburg Database, roughly 40% utilize the BONN Database, and about 30% of the remaining papers use other independent patient data. It is important to note the 70% of the proposed algorithms employing SVM complete testing and validation on public datasets. This implies that proposed algorithms can be compared to one another with minimal biases.

In order to better understand the Freiburg, BONN, and Flint Hills Databases, we give a brief overview of the data stored in each one. All three are publicly available datasets for researchers in the epilepsy domain. The Freiburg database [31] includes intracranial EEG recordings of 21 patients suffering from medically intractable focal epilepsy. All patient data was recorded prior to an invasive surgical procedure. The epileptic focus was located in different brain regions across patients; eleven patients contracted epilepsy in neocortical brain structures, eight patients in the hippocampus, and two across both regions. Intracranial grid-, strip-, and depth- electrodes were used to acquire records directly from focal areas with a higher signal-to-noise ratio (SNR) and less artifacts. The EEG data consists of 128 channels, 256 Hz sampling rate, and a 16 bit analog-to-digital converter. It is important to note that notch or any preprocess filtering has not been applied. The data was collected from three focal electrodes and three non-focal electrodes per patient, resulting in recordings of 2-6 seizures per patient with at least 50 minutes of pre-ictal data. For 13 patients, there are 24 hours of interictal recordings; for the remaining subset of patients, less than 24 hours of interictal recordings were joined together to come up with at least 24 hours of nonseizure data per patient.

Although the Freiburg database ranks as one of the most comprehensive databases available, there are a few weaknesses worth noting. First, about one day of recording time per patient is not sufficient for a rigorous analysis of proposed seizure detection algorithms. Second, about 40% of the Freiburg database does not contain continuous iEEG recordings. This discretized fusion of

time series data can increase artifacts within the data and cause detection algorithms to produce a greater number of false positives. Third, there is only limited clinical metadata and annotation information available at this time [33].

The University of BONN database [32] comprises 5 sets, labeled A-E, of EEG data. Each set in the database contains 100 single-channel EEG 23.6 second segments with each segment comprising 4096 data points. The data has a sampling rate of 173.61 Hz, a 12 bit analog-to-digital converter, and the time series have a spectral bandwidth equal to that of the acquisition system, which is 0.5 Hz – 85 Hz. Sets A and B were recorded from surface EEG electrodes of five healthy volunteers with eyes open (A) and eyes closed (B). The other three datasets were obtained from intracranial EEG from patients during a pre-surgical evaluation procedure. Both sets C and D contained recordings from interictal (non-seizure) and ictal (seizure) intervals, while set E contained only ictal iEEG recordings. Set C was collected from the hippocampal area of the brain, set D was collected from areas opposite the epileptogenic zones, and set E contained data from all recording sites.

A number of drawbacks observed from the BONN database include the following. First, the data contains scalp EEG recordings, which are not the focus of this literature review. Thus, an attempt to choose papers that only employed the intracranial subset of EEG data was made. Second, individual datasets are discontinuous and contain about 40 minutes of EEG data [33].

The Flint Hills Scientific, L.L.C., Public ECoG Database, supported by the NIH, consists of 1419 hours of continuous intracranial EEG recordings for 10 patients with 59 total seizures. The data is sampled at 249 Hz. The database includes metadata information about seizure events and all electrode locations, which range from 48-64 per patient [33].

Overall, scalp EEG is the least invasive type of EEG and are relatively simple to obtain. However, they provide high distortion and attenuation of high frequencies due to interference from the skull [20]. Intracranial EEG measurements have high SNRs and possess few artifacts.

However, iEEG data is difficult to obtain due to the complex invasive presurgical evaluation required. As such, iEEG data is limited because of ethical issues and the highly intrusive nature of the technique [20].

There is an obvious need for a well-rounded, comprehensive, and generalized epilepsy database in addition to the mere datasets currently available [33].

Some common frequency sub-bands examined in EEG epilepsy diagnosis are shown in Table 4 below.

Name	Frequency Range (Hz)	Associations
Delta	0 – 4	Deep sleep, serious brain disease or anesthesia
Theta	4 – 8	Emotional stress; drowsiness (abnormal in wakefulness)
Alpha	8 – 13	Quiet wakeful state (eyes closed)
Beta	13 – 30	Intense mental activity, stress and tension
Gamma	30 +	Observed at the cellular level; previously not of interest

Table 4 – Commonly analyzed frequency sub-bands in EEG [20].

2.1.4.2 Data-based vs. Model-Based Detection

Within seizure detection, there are two main paradigms used for construction an algorithm. One is called a data-based approach and the other is called a model-based approach. In the data based approach, EEG data is viewed as a large time series; here, linear and non-linear features are extracted via calculations from signals and systems theory. In the model based approach, a model, usually a state-space representation model is used to mathematically model the EEG data. From this model all resultant features and signal characteristics are extracted. In this literature review, 90% of the papers use a data-based approach [2, 3, 5-10, 12-15], while 10% used a model-based approach [4, 11]. Thus, we see a stronger inclination for methods to use data-based approaches. However, for those papers that did use a model-based approach, the EEG data was viewed as a non-linear

dynamical system [4] and as an autoregressive (AR) model [11]. All other papers viewed the EEG data as either a time-frequency distribution or a non-linear, nonstationary time series.

Viewing the EEG data as either data-based or model-based ultimately culminates in the extraction of features to be input into an SVM classifier. Each of the studies analyzed in this work greatly differed in the type of features extracted. However, there are some differences worth noting. First, the Discrete Wavelet Transform (DWT) is popular in feature extraction because it can represent both time and frequency information simultaneously. Another repetitive feature is signal power and/or Power Spectral Density (PSD); these features are used in epilepsy detection because EEG data increases in Power and Energy during seizure onset. A complete listing of features extracted for a subset of 15 papers is shown below.

Papers [Citation Number]	**Features**
1	ApEn, Hurst exponent, Detrended fluctuation analysis
2	DWT, Total Power, Log of product of absolute feature values
3	EMD algorithm extracts IMFs, variance of IMFs
4	Lyapunov exponent of Discrete Wavelet Packet Transform, Shannon Wavelet Entropy
5	Wavelet decomposition - 5 scales/3 frequency bands, Relative Energy, Relative Amplitude, Coefficient of Variation and Fluctuation Index
6	Teager Energy, Power, Lempel-Ziv Complexity for sub-bands
7	Spectral Distribution/Energy, Short-term Temporal Evolution; Artifiact Rejection System works alongside SVM to decrease FP
8	Wavelet Analysis
9	Spectral Power - Raw or Bipolar and/or Time-differential signals (only linear features)
10	Table of Frequency Domain, Time Domain, and Information Theory features included
11	AR coeffcients, Smoothed with Moving Average Filter
12	Spectral Power in 5 frequency bands via Wavelet and FFT
13	Mean of absolute value of coefficients, Average Power of Wavelet coefficients, SD of coefficients, Ratio of absolute mean values in sub-bands (via DWT); reduced via PCA, LDA, ICA
14	Relative Spectral Powers in specific frequency bands, Spectral Power Ratios, Feature Selection by Classification and Regression Tree (CART)
15	Wavelet features, ApEn, LLE, Maximum Value, Minimum Value, Mean, SD

Table 5 – A table of features values for a subset of 15 papers analyzed in the literature review.

The goal in the feature extraction phase of the seizure detection algorithm is to choose maximally relevant and minimally redundant features. If the feature set proves too large, redundancy among features may arise, ultimately increasing memory requirements and decreasing computational speed. Thus, a thorough analysis of the feature vector size in each algorithm was studied and interesting conclusions result. To begin, on average 26 features were chosen per vector to send into the SVM detector across all papers. The size of the feature vector varied among methods, with the smallest feature vector containing 3 features [3] and the largest feature vector containing 55 features [10]. It is important to note that a number of papers use dimensionality reduction techniques to reduce the dimension of their feature vector. The main dimensionality reduction techniques used were PCA, ICA, and LDA. There feature transformation techniques have the ability to minimize the within class scatter and maximize the between-class scatter, ultimately significantly improving classification accuracy [13]. A key observation with regards to feature vector dimension is almost no correlation exists between the number of features and the performance of the classifier. For example, a sensitivity of 100% and a FP rate of 0.10 FP/hr were achieved with only 6 features per feature vector [4]. Additionally, a detection accuracy of 98.72% was achieved in [6] with only 7 features per feature vector. Thus, removing the redundancy in the feature set is a key objective for increased classification results via SVM.

To further improve the accuracy of the detectors, post-processing techniques are employed to increase the robustness and reliability of detection. SVM classifiers are very sensitive to changes in data; therefore, the performance is greatly affected by post-processing. Algorithms which would have otherwise had 80% sensitivity without post-processing had 95% sensitivity with post-processing. The most notable, i.e. producing the highest sensitivity and detection accuracy, post-processing techniques included Kalman filtering [9, 11], smoothing, multi-channel decision and collar technique [5, 10, 12, 14], and maintaining a temporal constraint in addition to thresholding [3,

8]. Smoothing, multi-channel decision and collar technique was effective when SVM outputs were mapped, via the sigmoid function, to the probability of either being in a seizure or non-seizure class.

Lastly, to determine how an algorithm actually ranks compared to others in the field, a validation step must be executed. The majority of SVM algorithms proposed in the literature require a large subset of the clinical data to be used for training purposes. The goal is to develop seizure detection algorithms that are trained on a minimal set of data. However, roughly 70% of the papers analyzed use *n*-fold/leave-one-out cross validation technique, where *n* is usually 1 or 2, to estimate the performance of their classifiers. The cross validation technique works as follows, first, let *N* equal the number of data segments. To estimate the detector's sensitivity, specificity, latency, and false alarm rate, the classifier is trained on *N-1* segments per patient. The classifier is then tested on the N^{th} withheld data segment. This process is continued *N* times so that each data segment is tested. Usually each round results in a training of *N-1* seizures because there is roughly one seizure per data segment [7]. It was observed that a larger training set of data resulted in higher sensitivities with values around 90%. However, one exception was when Wavelet Analysis was used. In [8], 54 minutes of training data per patient were used with a testing set containing 639 hours across all patients. The resulting sensitivity was 96% with a FP Rate of 0.14 FP/hr.

3 Methods

The objective of this research project is to design, implement, and validate a support-vector-machine (SVM)-based software tool for the automatic detection of onset time of seizure events in subjects with drug-resistant epilepsy (DRE). The tool operates on time series of feature vectors computed from multichannel intracranial EEG (iEEG) recordings collected continuously across one week from DRE epilepsy patients. This work builds upon mathematical and computational tools for interpreting iEEG signals to answer the question: given promising features, can seizures be detected at the initiation stage in real-time?

3.1 Materials

3.1.1 Experimental Setup

Real patient data is drawn from existing databases, eliminating the need for human subjects and physical EEG equipment. Twenty-six DRE patients were monitored 24/7 via iEEG across several days, ranging from 2 to 10 days, before receiving surgical resection of the epileptogenic zone. The resulting multichannel iEEG recordings were sampled at 500 Hz, stored for offline review, de-identified, and then made available to the scientific community through the IEEG Portal. Data for each patient is obtained in the form of iEEG signals from multiple channels, with each channel corresponding to one recording intracranial electrode placed beneath the patient's scalp. The IEEG Portal, known as the International Epilepsy Electrophysiology Portal, is funded by the National Institutes of Neurological Disease and Stroke [42]. Notes about the clinical history of the

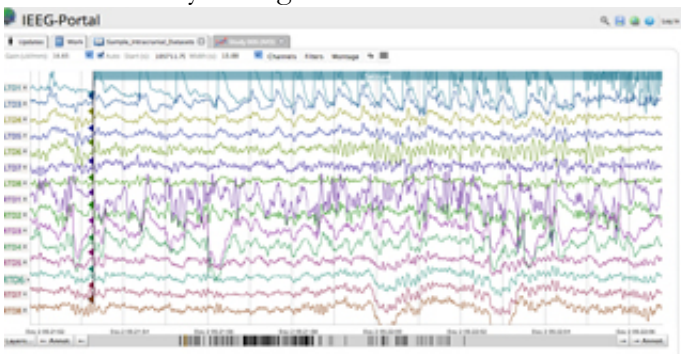


Figure 8 – EGG channel data, with each channel representing the electrical activity at respective areas of the brain. Image provided from IEEG.org [42].

subjects and annotations of seizure events (onset and offset times) are provided by board-certified epileptologists.

3.1.2 Data Repository

For each patient, the original multichannel iEEG recordings were divided into time series (one time series per channel) and each time series was divided into 2-second-long windows with a 1-second overlap. For each window, the iEEG signal was band-pass filtered in four distinct frequency sub-bands, resulting in four band-limited signals. For each band-limited signal, two features were computed, line-length and standard deviation. These two features were chosen based on their low computational burden and their ability to capture changes in iEEG system dynamics. The mathematical representation of each feature is shown below.

$$\text{Signal Line Length: } L.L. = \sum_{i=1}^{M} |x[i-1]\text{-}x[i]|$$

$$\text{Signal Standard Deviation: } S.D. = \sqrt{\frac{1}{M-1}\sum_{i=0}^{M} (x[i]-\mu)^2},$$

where x is the iEEG signal measured in microvolts, M is the number of samples, and μ is the mean.

As a result, an 8x1-feature vector was computed for each iEEG channel per second for each patient. Patient data sets can be as large as 100 channels, consisting of both electro-corticographical and intracranial data, thus forming a truly big-data problem. All analyzed data was made available on a local high capacity server at the UConn Storrs Campus. The main frequency bands of interest considered in this study are: [15, 30] Hz (beta), [35, 70] Hz (gamma), [80, 250] Hz (high-gamma), and [10, 250] Hz (full-band).

3.2 Design

The prediction of an impending seizure relies on classification techniques made possible by machine learning, such as SVM. Classification allows a descriptive label to be chosen given a particular input data set [44]. Labels are usually determined in advance, for our purposes defined as suffering a seizure or not. A classifier model is termed *supervised* if it is developed based on tested features containing the appropriate label for each input [44]. The flowchart of a *supervised* classifier is shown in Figure 9 above.

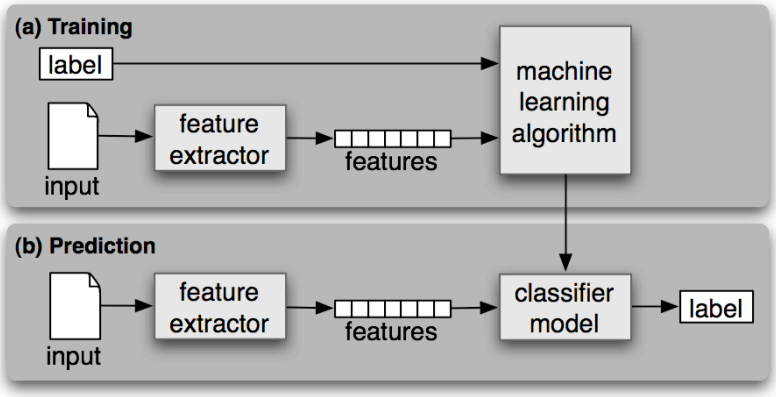


Figure 9 – Model of Supervised Classification. During training, the feature extractor converts the input into a set of features which are fed into a machine learning algorithm to develop a model. During prediction, an identical feature extractor is used to convert unknown inputs into feature sets. The classifier model is used to classify/label the given input. Image provided from Natural Language Processing with Python textbook [44].

Thus, the domain specific aim is making informed automated decisions, which in this application is differentiating between normal electrical activity in the brain and electrical signals which indicate an imminent seizure [23].

3.2.1 Algorithm Implementation

The algorithm implemented in this study is designed to input a continuous stream of multichannel iEEG data and output a decision every second based on the patient's brain wave morphology. This task is accomplished by splitting the multichannel iEEG data into time series data, which is then further segmented into overlapping time windows. A generated decision is calculated during each time window.

Within each time window, the algorithm executes through 3 consecutive stages, as shown in Figure 10 below. The first stage constitutes preprocess filtering, where each channel passes through a bank of band-pass filters. The band-pass filter pass-band and stop-band frequencies are chosen

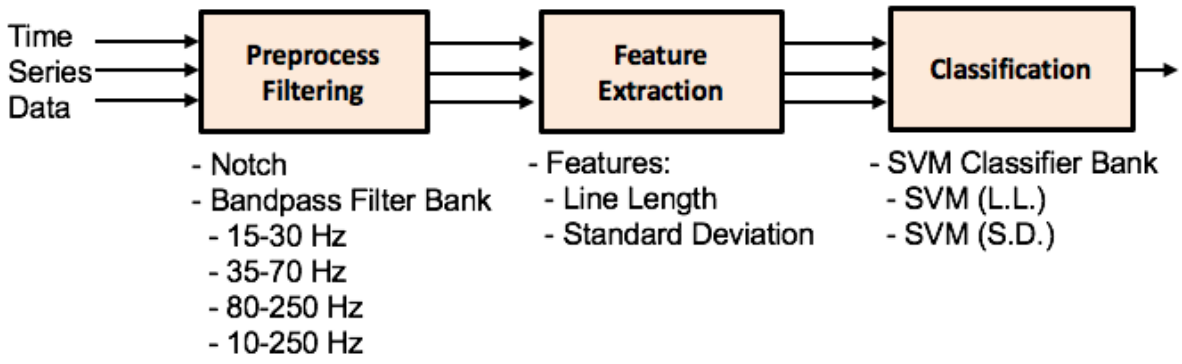


Figure 10 – Stages of the detection algorithm.

such that resulting band-limited signals are clinically relevant for seizure detection [5]. It is important to note that the number of filtered band-limited signals will be four times the number of incoming iEEG channels. The second stage comprises the calculation of two mathematically relevant features from each band-limited frequency signal. The quantitative expressions for line-length and standard deviation features are given in Section 3.1.2. From the Review of Literature (Section 2), line length and power are shown to be successful in the detection of seizures [5,7,19]. In most cases, it is the deviation of power from the mean that captures most important signal characteristics. Thus, it is hypothesized that taking the standard deviation with power instead of amplitude will aid in seizure detection. The line length feature is also known to be sensitive to variations in amplitude and frequency modulation [45], therefore it is also hypothesized to exhibit changes in behavior between the band-limited frequency signals.

An important additional consideration for the signal features is computational cost. The two features used in this study were selected for both capturing significant signal information and their computational efficiency. Observing the mathematical relations for line-length and standard deviation, their calculation requires the use of one summing operation. Thus, only a single iteration

of the signal is required for computation, resulting in a *O(n)* runtime. Comparing our computation to the FFT (Fast Fourier Transform), which requires a *O(n*log(n))* runtime, the chosen features are computationally efficient.

The third stage of the algorithm involves classification via a support-vector-machine (SVM), selected for its adaptability to large varying datasets and its precision in discriminating between two classes. The SVM ultimately outputs a discrete decision, i.e. a seizure or non-seizure class label. All training, testing, and validation was completed in MATLAB through the Statistics and Machine Learning Toolbox and the SVM Library.

3.2.1.1 SVM Classifier Architecture

The SVM architecture designed in this study assumes that the multichannel iEEG data is linearly separable in some higher dimensional feature space. As such, a linear kernel was chosen in the SVM implementation after initial testing. Our architecture combines both aspects of the two state of the art architectures discussed in the Detection Methods (Section 2.1.4). We pass all *n* channels through a feature-specific SVM, as shown in Figure 11 below, producing an *n-dimensional* feature space. For each patient, the two feature-specific SVMs produce a discrete output class label for each time window and the resulting labels are correlated to produce a final binary decision.

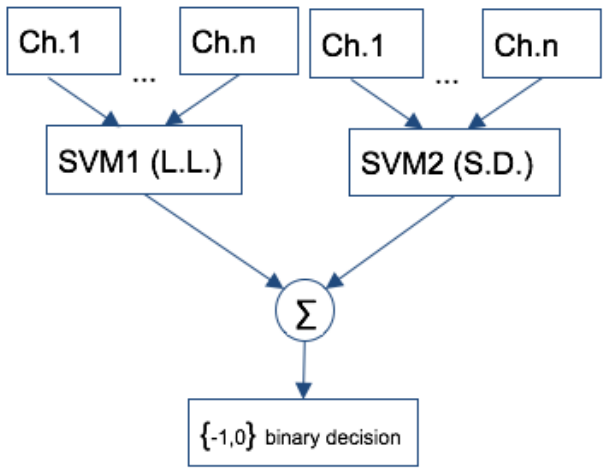


Figure 11 – SVM architecture for seizure detection problem.

When building the model, an SVM algorithm searches for two specific criteria, one, a hyperplane with the largest possible boundary, and two, a hyperplane that separates as many different class instances as possible. Akin to Heisenberg's Uncertainty Principle, where an observer can either calculate an electron's position or

velocity with 100% certainty, so too can the SVM optimize only one of the aforementioned criteria. The key to producing a successful classification is choosing constraints and parameters such that there is an elegant balance. The kernel method accomplishes the task of mapping the dataset into a higher dimensional feature space. This mapping is done implicitly, without explicit knowledge of each value, by simply computing the inner product between the two vectors in the original space. Then, the linear bounds are separated as much as possible in the higher dimensional feature space. The boundary constraint was chosen to be the default MATLAB value, $C = 1$, which produces a large margin within the hyperplane.

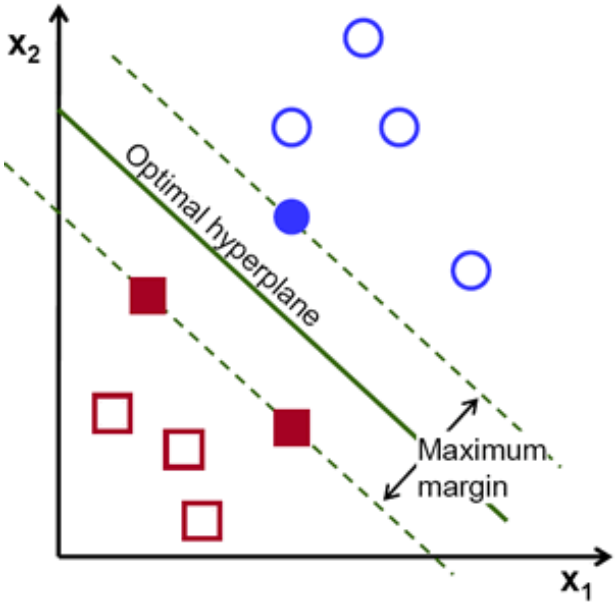


Figure 12 – Representation of the linear kernel used for the SVM implementation.

3.2.1.2 Classification Criteria

In order to classify the dataset based on the designed SVM Classifier Model, a few classification criteria and a detection rule must be put in place. It is important to mention that patient data is split into two categories, a training set and a testing, or validation, set[1]. First, for each testing block of data (~ 1 hour in length) in a single patient, the time windows are classified and then class labels from each feature specific SVM are summed together. A detection warning is issued if and only if the following two conditions are satisfied, one, both features are simultaneously in a 'seizure' class, and two, at least N consecutive 'seizure' class labels are observed. The parameter N is patient specific and designed to be less than the length of the smallest seizure. N was experimentally chosen such that to minimize False Positives and maximize True Positives.

[1] Note that training dataset used to build the classifier model will not be used in the testing dataset.

4 Results

The first step in understanding the applicability of the algorithm to various patients is by observing the behavior of the extracted signals features around seizure data. The line-length and standard deviation were examined in specific channels responsible for seizure initiation, as documented in the clinical annotations. In order for the SVM classifiers to make meaningful decisions, it is crucial for noticeable changes in the signal features to take place. To show robustness of the signal features, their graphs around seizure data for two randomly selected patients are shown in Figures 13-14 below. In particular, the beta-band line-length and standard deviation are plotted, with the red markers indicating clinical seizure onset and offset times.

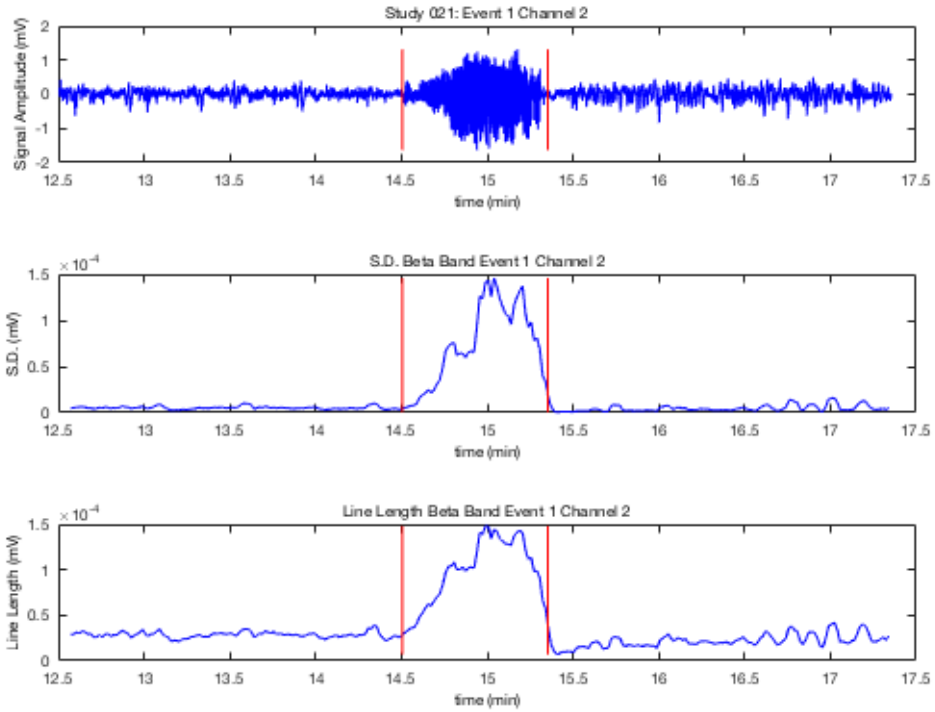


Figure 13 – Study 021 Seizure 1 in ECoG channel 2. Shown from top to bottom, beta-band signal, beta-band standard deviation, beta-band line-length.

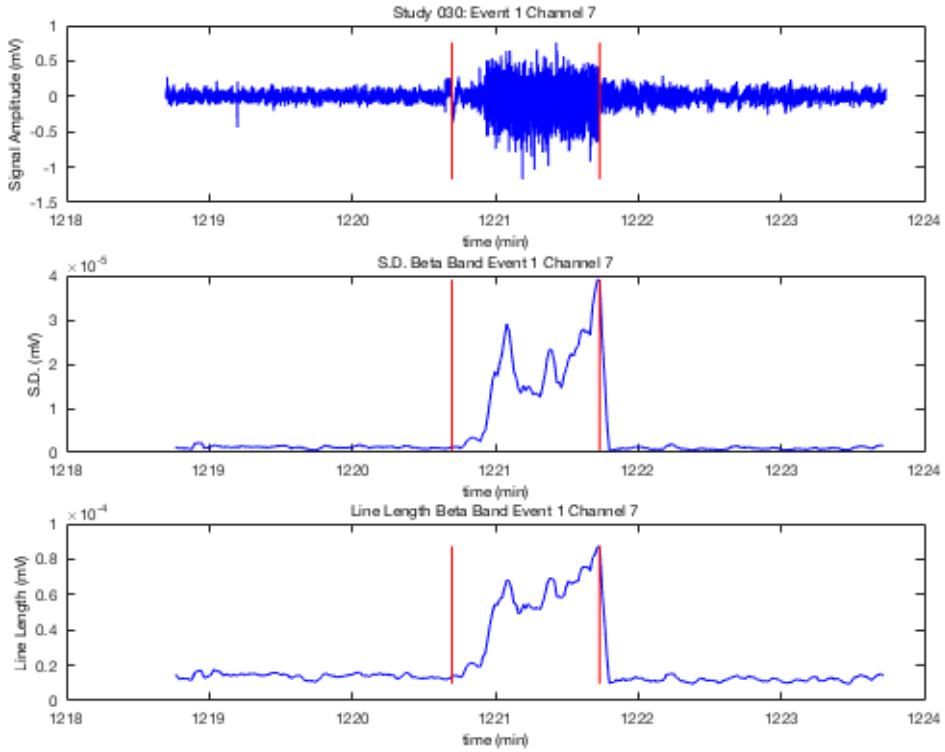


Figure 14 – Study 030 Seizure 1 in ECoG channel 7. Shown from top to bottom, beta-band signal, beta-band standard deviation, beta-band line-length.

We clearly see from the two figures above that the chosen features increase in amplitude during seizure onset. This result demonstrates that both features are responsive to seizure events across different patients. Therefore, by selecting line-length and standard deviation, the algorithm will be successful discerning seizure and non-seizure data among patients. Additionally, the two features are highly correlated, i.e. exhibit a similar shape, however, they do showcase slight differences. These slight nuances in feature shape can be instrumental in classifying incoming patient data. For example, from Figure 13 we see that the line length is more sensitive during the beginning of a seizure, a quality that can aid in earlier detection. The added redundancy of a standard deviation feature is also helpful in the detection process and such similarity among features can prove positive [16].

4.1 Classifier Validation

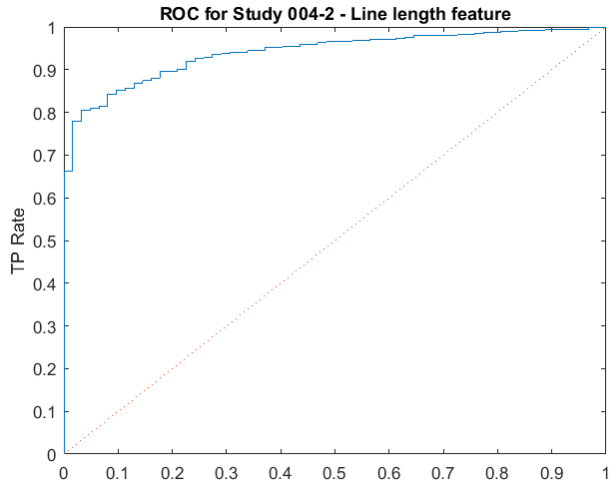
In order to determine whether or not the software tool can be used in a clinical setting, it must

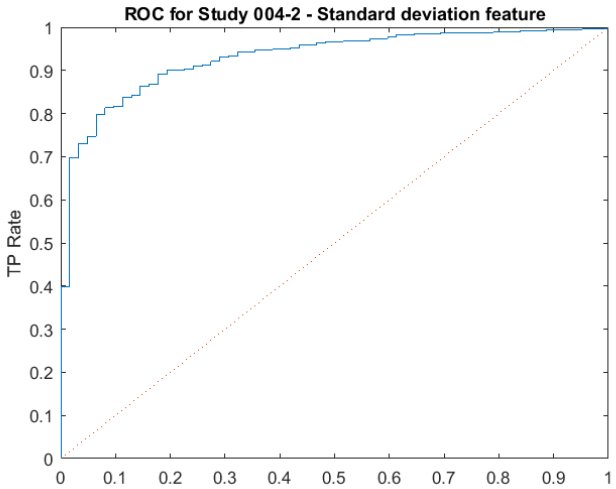
be validated on different patients experiencing varying seizure morphologies. As mentioned in footnote 1, it must be guaranteed that the algorithm is validated on data not included in the training set. Table 6 below shows the proposed algorithm trained on 2 hours of training data around seizure and validated on all remaining data from the same patient for 4 patients. The average False Alarm Rate across all patients is 0.00 FP./hr and the Average delay time across all patients is 4 seconds.

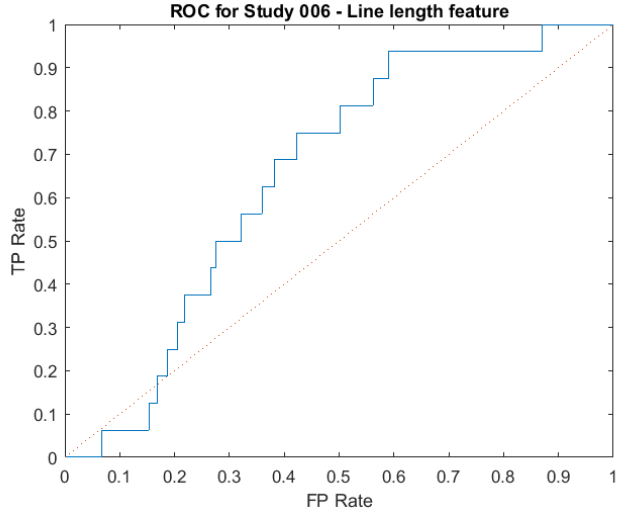
Patient	Total Seizures	Detected Seizures	False detection	Total testing EEG (h)	False alarm rate (# FP/h)	Average delay time (s)
Study 004-2	3	3	0	184	0.00	2
Study 006	5	4	0	25	0.00	14
Study 010	3	3	0	95	0.00	0
Study 011	3	2	0	55	0.00	0

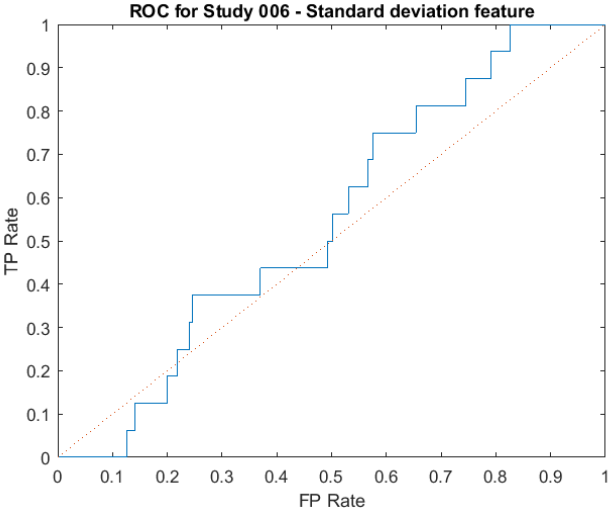
Table 6 – Table showing algorithm validation on a subset of 4 patients.

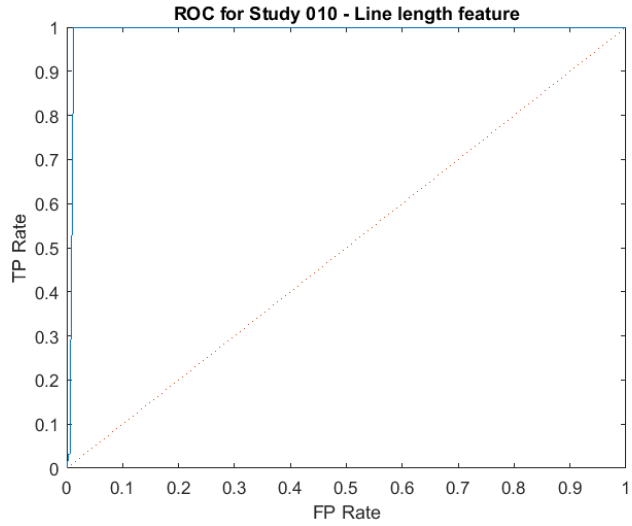
Another validation metric used for SVM is the ROC (Remote Operating Characteristic) curve generated from MATLAB testing data. The ROC curves were calculated using the *perfcurve* method in MATLAB, taking as parameters a vector of classifier predictions, given true class labels, and the positive class label. These curves were generated during training and validation when populating the table above. The algorithm performs significantly above chance as is shown in the figures below.

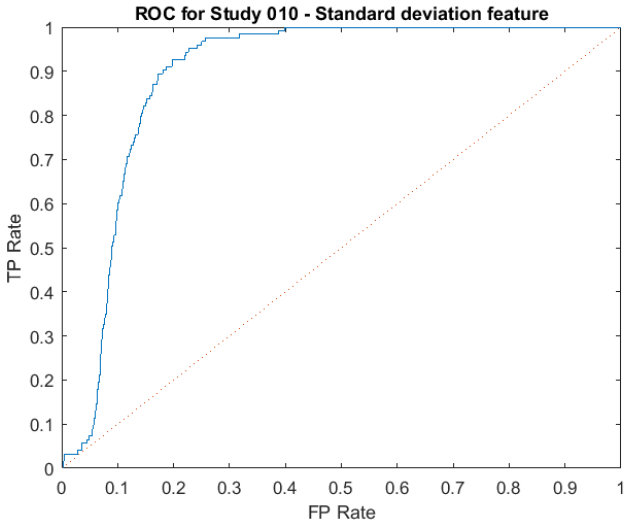

ROC for Study 004-2 - Line length feature
TP Rate
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1


ROC for Study 004-2 - Standard deviation feature
TP Rate
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1


ROC for Study 006 - Line length feature
TP Rate
FP Rate
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1


ROC for Study 006 - Standard deviation feature
TP Rate
FP Rate
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1


ROC for Study 010 - Line length feature
TP Rate
FP Rate
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1


ROC for Study 010 - Standard deviation feature
TP Rate
FP Rate
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

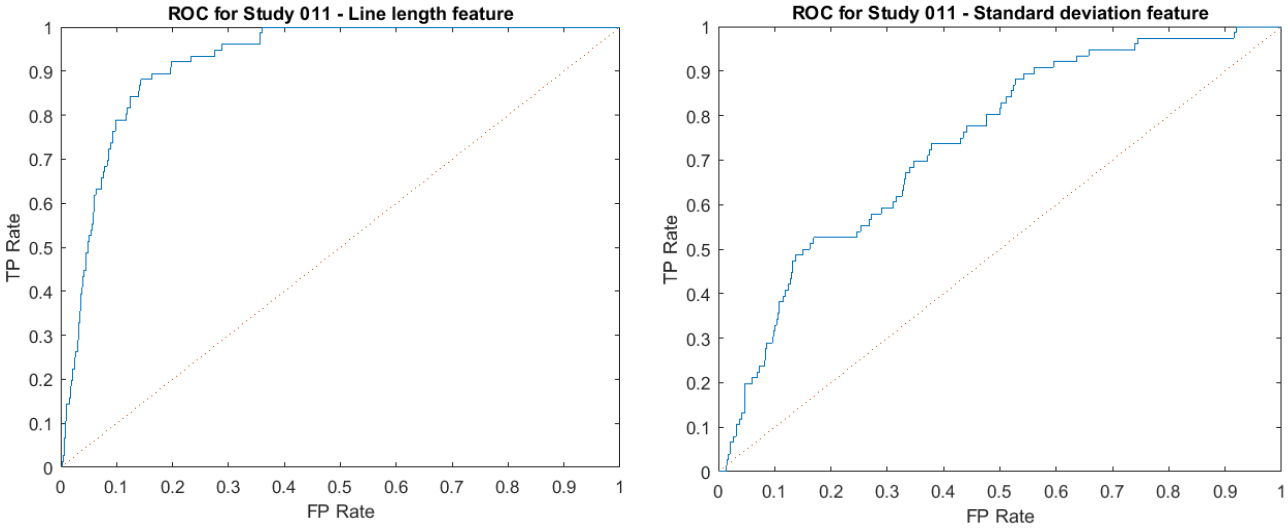


Figure 15 – ROC curves for each feature-specific SVM per patient.

Finally, it was observed that using all iEEG channels improved detection results, as more information was captured by the algorithm. It was also observed that higher frequency band-limited signals do not improve classifier performance and low frequency signals are more advantageous in the seizure detection problem. Through testing it was determined that beta-band (low frequency) features showed redundancy with full-band features, thus not contributing additional useful information.

Overall, the proposed software tool shows a significant improvement above chance and is proven to be robust across a subset of patients.

5 Conclusions and Future Work

5.1 Closed Loop Intervention Systems

A natural progression forward would be to improve the classifier seizure detection method and extend seizure localization capabilities. Current research is focused heavily on seizure detection methods and very few versatile techniques have been implemented in the area of seizure localization, i.e. in the form of channel selection methods and keeping feature vectors very small. The challenge in this area has been in selecting appropriate channels and understanding where to place detection electrodes. Correct selection and placement of electrodes will allow electrical stimulation to be applied to a local area of the brain, without the patient consciously perceiving intervention [22]. Such advances will allow the design of a low-power, low cost, automatic seizure detection algorithm capable of warning of imminent seizures and taking preventative measures to suppress them.

Another direction to consider is optimizing and training this algorithm offline so to deploy it on a microcontroller based system without impending memory usage. A significant advantage of the SVM is that it can be trained offline, thus requiring a simple set of matrix operations (dot product and vector sum) to reach a decision in real-time.

6 References

[1] Yuan, Q., Zhou, W., Li, S., & Cai, D. (2011). Epileptic EEG classification based on extreme learning machine and nonlinear features. Epilepsy research, 96(1), 29-38.

[2] Zhang, Z., & Parhi, K. K. (2014, August). Seizure detection using wavelet decomposition of the prediction error signal from a single channel of intra-cranial EEG. In Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE (pp. 4443-4446). IEEE.

[3] Zheng, Y. X., Zhu, J. M., Qi, Y., Zheng, X. X., & Zhang, J. M. (2015). An Automatic Patient-Specific Seizure Onset Detection Method Using Intracranial Electroencephalography. Neuromodulation: Technology at the Neural Interface, 18(2), 79-84.

[4] Nesaei, S., & Sharafat, A. R. (2014). Real-time Detection of Precursors to Epileptic Seizures: Non-Linear Analysis of System Dynamics. Journal of medical signals and sensors, 4(2), 103.

[5] Liu, Y., Zhou, W., Yuan, Q., & Chen, S. (2012). Automatic seizure detection using wavelet transform and SVM in long-term intracranial EEG. Neural Systems and Rehabilitation Engineering, IEEE Transactions on, 20(6), 749-755.

[6] Tang, Y., & Durand, D. M. (2012). A tunable support vector machine assembly classifier for epileptic seizure detection. Expert systems with applications, 39(4), 3925-3938.

[7] Kharbouch, A., Shoeb, A., Guttag, J., & Cash, S. S. (2011). An algorithm for seizure onset detection using intracranial EEG. Epilepsy & Behavior, 22, S29-S35.

[8] Duun-Henriksen, J., Kjaer, T. W., Madsen, R. E., Remvig, L. S., Thomsen, C. E., & Sorensen, H. B. D. (2012). Channel selection for automatic seizure detection. Clinical Neurophysiology, 123(1), 84-92.

[9] Park, Y., Luo, L., Parhi, K. K., & Netoff, T. (2011). Seizure prediction with spectral power of EEG using cost-sensitive support vector machines. Epilepsia, 52(10), 1761-1770.

[10] Temko, A., Thomas, E., Marnane, W., Lightbody, G., & Boylan, G. (2011). EEG-based neonatal seizure detection with support vector machines. Clinical Neurophysiology, 122(3), 464-473.

[11] Chisci, L., Mavino, A., Perferi, G., Sciandrone, M., Anile, C., Colicchio, G., & Fuggetta, F. (2010). Real-time epileptic seizure prediction using AR models and support vector machines. Biomedical Engineering, IEEE Transactions on, 57(5), 1124-1132.

[12] Chan, A. M., Sun, F. T., Boto, E. H., & Wingeier, B. M. (2008). Automated seizure onset detection for accurate onset time determination in intracranial EEG. Clinical Neurophysiology, 119(12), 2687-2696.

[13] Subasi, A., & Gursoy, M. I. (2010). EEG signal classification using PCA, ICA, LDA and support vector machines. Expert Systems with Applications,37(12), 8659-8666.

[14] Zhang, Z., & Parhi, K. K. (2015, August). Seizure prediction using polynomial SVM classification. In Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE (pp. 5748-5751). IEEE.

[15] Murugavel, A. M., & Ramakrishnan, S. (2015). Hierarchical multi-class SVM with ELM kernel for epileptic EEG signal classification. Medical & biological engineering & computing, 1-13.

[16] Riaz, F., Hassan, A., Rehman, S., Niazi, I. K., & Dremstrup, K. (2015). EMD based temporal and spectral features for the classification of EEG signals using supervised learning.

[17] Shen, C. P., Chen, C. C., Hsieh, S. L., Chen, W. H., Chen, J. M., Chen, C. M., ... & Chiu, M. J. (2013). High-performance seizure detection system using a wavelet-approximate entropy-fSVM cascade with clinical validation. Clinical EEG and neuroscience, 44(4), 247-256.

[18] Lima, C. A., & Coelho, A. L. (2011). Kernel machines for epilepsy diagnosis via EEG signal classification: A comparative study. Artificial intelligence in medicine, 53(2), 83-95.

[19] Shoeb, A., Edwards, H., Connolly, J., Bourgeois, B., Treves, S. T., & Guttag, J. (2004). Patient-specific seizure onset detection. Epilepsy & Behavior, 5(4), 483-498.

[20] Gardner, A. B. (2004). A novelty detection approach to seizure analysis from intracranial EEG (Doctoral dissertation, Georgia Institute of Technology).

[21] Faul, S. D. (2010, August). Dynamic channel selection to reduce computational burden in seizure detection. In Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE(pp. 6365-6368). IEEE.

[22] Mormann, F., Andrzejak, R. G., Elger, C. E., & Lehnertz, K. (2007). Seizure prediction: the long and winding road. Brain, 130(2), 314-333.

[23] van Mierlo, P., Papadopoulou, M., Carrette, E., Boon, P., Vandenberghe, S., Vonck, K., & Marinazzo, D. (2014). Functional brain connectivity from EEG in epilepsy: Seizure prediction and epileptogenic focus localization. Progress in neurobiology, 121, 19-35.

[24] Pati, S., & Cole, A. J. (2014). How focal is generalized epilepsy: A distinction with a difference?. Epilepsy & Behavior, 34, 127-128.

[25] Zeki, S., & Shipp, S. (1988). The functional logic of cortical connections. Nature.

[26] Iasemidis, L. D., Sackellares, J. C., Zaveri, H. P., & Williams, W. J. (1990). Phase space topography and the Lyapunov exponent of electrocorticograms in partial seizures. Brain topography, 2(3), 187-201.

[27] Vapnik, V. N., & Vapnik, V. (1998). Statistical learning theory (Vol. 1). New York: Wiley.

[28] Burges, C. J. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery, 2, 121-167.

[29] Zhang, Y., Zhou, W., & Yuan, S. (2015). Multifractal analysis and relevance vector machine-based automatic seizure detection in intracranial EEG. International journal of neural systems, 25(06), 1550020.

[30] General Introduction. (n.d.). Retrieved April 29, 2016, from http://www.ncbi.nlm.nih.gov/home/about/mission.shtml

[31] EEG Database. (n.d.). Retrieved April 29, 2016, from http://epilepsy.uni-freiburg.de/freiburg-seizure-prediction-project/eeg-database

[32] EEG Database. (n.d.). Retrieved April 29, 2016, from http://epileptologie-bonn.de/cms/front_content.php?idcat=193&lang=3

[33] Ihle, M., Feldwisch-Drentrup, H., Teixeira, C. A., Witon, A., Schelter, B., Timmer, J., & Schulze-Bonhage, A. (2012). EPILEPSIAE–A European epilepsy database. Computer methods and programs in biomedicine, 106(3), 127-138.

[34] Koçer, S., & Canal, M. R. (2011). Classifying epilepsy diseases using artificial neural networks and genetic algorithm. Journal of medical systems,35(4), 489-498.

[35] Yuan, Q., Zhou, W., Li, S., & Cai, D. (2011). Epileptic EEG classification based on extreme learning machine and nonlinear features. Epilepsy research, 96(1), 29-38.

[36] Song, Y., & Zhang, J. (2016). Discriminating preictal and interictal brain states in intracranial EEG by sample entropy and extreme learning machine. Journal of neuroscience methods, 257, 45-54.

[37] Han, M., Ge, S., Wang, M., Hong, X., & Han, J. (2014). A Novel Dynamic Update Framework for Epileptic Seizure Prediction. BioMed research international, 2014.

[38] Yang, J., Singh, H., Hines, E. L., Schlaghecken, F., Iliescu, D. D., Leeson, M. S., & Stocks, N. G. (2012). Channel selection and classification of electroencephalogram signals: an artificial neural network and genetic algorithm-based approach. Artificial intelligence in medicine, 55(2), 117-126.

[39] Übeyli, E. D. (2010). Lyapunov exponents/probabilistic neural networks for analysis of EEG signals. Expert Systems with Applications, 37(2), 985-992.

[40] Yuan, Q., Zhou, W., Liu, Y., & Wang, J. (2012). Epileptic seizure detection with linear and nonlinear features. Epilepsy & Behavior, 24(4), 415-421.

[41] Singh, D., Singh, P., & Kumar, G. Detection of Epileptic Seizures using ANN and SVM.

[42] "IEEG-Portal Webconsole." *IEEG-Portal Webconsole*. Web. 4 Sep. 2016. <https://www.ieeg.org/>.

[43] Perucca, P., F. Dubeau, and J. Gotman. "Intracranial Electroencephalographic Seizure-onset Patterns: Effect of Underlying Pathology." *Brain* (2014): 183-96. Print.

[44] Bird, Steven, and Ewan Klein. *Natural Language Processing with Python*. Beijing: O'Reilly, 2009. Print.

[45] L. Guo, D. Rivero, J. Dorado, J. R. Rabunal and A. Pazos. Automatic epileptic seizure detection in EEGs based on line length feature and artificial neural networks. J . Neurosci. Methods 191(1), pp. 101-109. 2010. DOI: 10.1016/j.jneumeth.2010.05.020.