2016

EEG Interictal Spike Detection Using Artificial Neural Networks

Howard J. Carey III
Virginia Commonwealth University

EEG INTERICTAL SPIKE DETECTION USING ARTIFICIAL NEURAL NETWORKS

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at Virginia Commonwealth University.

by

HOWARD J. CAREY III

B.Sc. Physics, James Madison University, 2011

Director: Milos Manic,

Professor, Department of Computer Science

Virginia Commonwewalth University

Richmond, Virginia

December, 2016

Acknowledgements

I would like to thank my adviser, Dr. Milos Manic, from the VCU Department of Computer Science. He gave me the opportunity to pursue my interests and guided me through the difficult process of my research. I would also like to thank my coworkers, Dumidu Wijayasekara, Kasun Amarasinghe, Daniel Marino, and Patrick Sivils. All of them helped create a very supportive working environment where collaboration and open discussion helped fuel my research interests. Finally, I would like to thank my wife Aimee, parents, friends and family for being supportive through my many late nights and missed events.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abstract

EEG INTERICTAL SPIKE DETECTION USING ARTIFICIAL NEURAL NETWORKS

By Howard J. Carey III

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at Virginia Commonwealth University.

Virginia Commonwealth University, 2016.

Director: Milos Manic,
Professor, Department of Computer Science

Epilepsy is a neurological disease causing seizures in its victims and affects approximately 50 million people worldwide. Successful treatment is dependent upon correct identification of the origin of the seizures within the brain. To achieve this, electroencephalograms (EEGs) are used to measure a patient's brainwaves. This EEG data must be manually analyzed to identify interictal spikes that emanate from the afflicted region of the brain. This process can take a neurologist more than a week and a half per patient. This thesis presents a method to extract and process the interictal spikes in a patient, and use them to reduce the amount of data for a neurologist to manually analyze. The effectiveness of multiple neural network implementations is compared, and a data reduction of 3-4 orders of magnitude, or upwards of 99%, is achieved.

CHAPTER 1

INTRODUCTION

1.1 Introduction to Epileptic Problem

Epilepsy is a neurological disorder that is caused by an abnormal firing of a cluster of neurons in the brain [17, 30]. According to the World Health Organization (WHO), it affects approximately 50 million people worldwide across all age groups [32]. People who have epilepsy experience debilitating seizures that come without warning, interrupting their daily life and potentially endangering themselves. Epilepsy typically comes in two forms, the most common being idiopathic epilepsy, which accounts for about 60% of the population with epilepsy. Idiopathic epilepsy has no known, direct cause. The second form is called symptomatic epilepsy. Potential causes include head injury, brain tumors, brain infections, and lack of oxygen to the brain. Epilepsy is treatable through various forms of medication, but has no known cure [32].

With the invention of electroencephalography (EEG), shown in Figure 1, doctors have been able to analyze the electromagnetic radiation given off by the brain, commonly called brainwaves. The EEG device consists of numerous electrodes that are placed in strategic locations around the patient's head. Each electrode is used to measure the voltage potential across the brain, giving a voltage over time readout, as shown in Figure 10 in Section 3.1.1. The EEG readings can be used to identify abnormal brainwave patterns, such as an epileptic seizure, or in the case of this study, an interictal spike. An ictal event is another name for a seizure event, and an interictal spike is a spike that occurs between ictal events (between seizures). An epileptic seizure is characterized in the EEG data by a period of very high amplitude, short

Fig. 1.: An EEG cap shown on a mannequin head. All wires connect to a central unit via the large white cable [46].

duration pulses. An interictal spike, however, is a high amplitude short duration pulse that occurs sporadically, as opposed to in a quick series.

These interictal spikes, while not seizures themselves, are generated by the same group of neurons that cause the patients seizures [6, 29]. Therefore, if neurologists can identify where in the brain these spikes are coming from, they have likely found the source of the patient's seizures as well. To identify these spikes, a neurologist must manually analyze the EEG output across multiple channels and be able to discern a spike from the noisy background brainwave signals. Such noise can include eye blinks, facial muscle movement, errant EM waves from electronic devices, etc. A

trained neurologist can differentiate between these signals to find the spike signal, however it comes with a great time commitment.

An EEG reading session can range from a few hours to dozens of hours, given the immense amount of data to analyze [43, 16, 1]. The interictal spike lasts about 100ms, therefore the neurologist must examine in fine detail a multiple-hour session second by second to find the spike pattern. Once enough spikes are identified, he or she can start to piece together which electrodes exhibit the spike pattern and identify where they are coming from within the brain. This process is incredibly time consuming, taking many hours of a neurologist's valuable time away from other important tasks.

A given patient may have hundreds of spikes in a single EEG reading. A patient with 300 spikes will yield only 30 thousand data records being related to interictal spikes. An entire multiple-hour EEG reading can have up to 20 million data records. This turns out to be a massively imbalanced dataset, with less than 1% of the total data being spike data, or the points of interest. The combination of the massive time required to manually solve this problem and a heavily imbalanced dataset lead to the contributions of this thesis, presented in the following section.

1.2 Contributions

The motivations for analyzing this problem are as as follows:

1. Manual interictal spike detection is very time consuming, and therefore costly.
2. Quick identification of interictal spikes can lead to quicker patient treatment.
3. Massively imbalanced dataset presents an interesting computer science classification problem.

The contributions of this thesis for addressing these motivations are as follows:

1. Reduce the data needed to be manually analyzed by the doctor.

2. Maintain a sufficient degree of accuracy in spike detection to identify the interictal spike source.

The overall goal for this thesis was to reduce the amount of data a doctor needs to manually analyze, thereby reducing the total time it takes to locate a patient's seizures within the brain. This was accomplished by utilizing supervised machine learning techniques to filter out EEG data that does not contain interictal spikes, leaving a much smaller dataset to be manually analyzed. To provide a ground truth to train the system, a doctor would be required to manually locate a small number of interictal spikes (on the order of 10 spikes), then let the algorithm analyze the rest of the dataset to find locations of potential spikes. These locations would be extracted and returned to the doctor to to determine the accuracy of the output. Keeping the doctor in the loop to provide the final determination on whether an extracted sample is a spike or not helps to maintain the accuracy and overall reliability of the system.

A very important point must be made about the second objective of this project. The goal was not to identify every instance of an interictal spike in the patient, but only to identify "enough" to solve the problem: where are the spikes coming from within the brain, or, which electrodes exhibit the spike pattern. Identification of every spike is not necessary to solve this problem. The neurologist consulted for this study only needed to identify about 100-150 well formed interictal spikes before being confident enough to draw a conclusion about the spike origin. With this in mind, as long as the methods examined were able to extract enough spikes from the model, the total size of the remaining data was the ultimate metric of effectiveness.

The rest of this thesis presents the methods used to analyze and test the identification of interictal spikes across five patients. The work was split into two distinct

parts. The first part, chapter 3, takes a single patient and analyzes interictal spike patterns, attempting to extract relevant information from that same patient. This chapter analyzed information from a single electrode, then studied three sets of multiple electrodes to compare the effectiveness of using EEGs with more electrodes. The second part, chapter 4, analyzes whether the interictal spikes of one patient can be used to identify spikes in other patients.

CHAPTER 2

BACKGROUND

2.1 Related Work

Machine learning and epileptic research have been closely tied for many years. Some of the first successful automated identification of both seizure and spike features started to come out in the mid 1970s. Automated seizure and spike detection has been researched thoroughly for the past two to three decades. Earlier automated systems looked at neural networks by themselves as a tool for analysis, Ozdamar et al. and Webber et al. [33, 42]. Nigam et al. and Gabor et al. [31, 13] utilized a self-organizing map (SOM) with neural networks to classify seizures. The early results had varying degrees of success, with more success in identifying seizures than interictal spikes.

Many researchers have utilized neural networks in conjunction with other statistical analysis methods, such as correlation methods by Ko et al. and Garg et al. [22, 14]. They used various filtering and thresholding techniques to identify abnormal regions of EEG data. Zerifia et al. [44] used a genetic algorithm to optimize these thresholding values to achieve accurate results on data that humans failed to accurately classify. James et al. [19] analyzed epileptic spikes in patients using a neural network in conjunction with a fuzzy logic system that added spatial information to their process.

Many researchers use wavelet transforms to break down the voltage signal into frequency values for further analysis. Various methods of implementing these wavelet transforms have led to successful results in classification. Approximate entropy of

wavelet transformed data was used by Kumar et al. and Akareddy et al. [25, 2]. In Kalayci et al., Subasi et al., and Aliabadi et al. [20, 37, 3] utilized neural networks with wavelet data as input for classification. Quiroga et al. [34] explored the use of superparamagnetic clustering of wavelet data while Chaibi et al. [9] analyzed the difference in accuracy between using continuous and discreet wavelet transforms (CWT vs DWT).

Recent years have led to newer methods of analysis. Ren et al. [35] examined EEG data using a convolutional deep belief neural networks, and Anusha et al. [4] explored unique feature vectors for neural network classification. The use of principal component analysis was compared by Kottaimalai et al. [23]. Slow waves patterns in EEG were examined by Liu et al. [27] and used along with the Adaboost classifier to identify spikes. Barkmeier et al. [7] validated the use of these automated systems by showing the accuracy of their method was comparable to the human reviewers they used for comparison.

A recent survey in automated epileptic feature detection by Tzallas et al. [40] discussed many successful methods in extracting the desired features from EEG recordings. However, as seen in the previously discussed works and in the survey paper, there is no method which clearly outperforms all others. No consensus has been found regarding which system, if any, to employ in a commercial environment, and there has been little to no penetration of these systems into the medical community.

2.2 EEG and Interictal Spikes

This sections serves to give background on what electroencephalography (EEG) is, how it works, and how it is used to help treat epilepsy. It will also go into detail on what an interictal spike is, and how doctors currently use EEG to detect these interictal spikes.

2.2.1 Electroencephalography

Every living person's brain generates pulses of electricity, creating an electromagnetic field around the brain. EEGs are devices capable of measuring this field by way of numerous sensors place around the brain. EEGs can have a number of different electrodes. Higher numbers of electrodes increase the spatial resolution of information read from the brain, as there are simply more electrodes recording data around the brain. Electrodes can also be placed directly onto the brain itself, allowing much more sensitive readings of the electromagnetic field, however this requires invasive surgery and is understandably dangerous. This thesis presents work done only with external EEGs.

Figure 9, in Section 3.1.1, shows a 256 electrode EEG on the head of a person. EEG measures voltage differences across the brain from numerous electrodes placed around the head. These electrodes are placed directly on the scalp using electrically conductive gel to help increase the sensitivity of the electrodes. The position of the electrodes must be precisely aligned to the patient's features to make sure each electrode is measuring the correct region of the brain.

Each electrode of the EEG network outputs its own voltage reading over time. Different thought patterns, muscle movements, and even emotions [21] can cause a voltage differential detectable by the device. Artifacts in the data can be caused by excessive muscle movements, like smiling or blinking, as well as external sources, such as high-voltage power lines, cell phones or anything that generates electromagnetic radiation. The 256 electrode EEG used in this study also has electrodes up to the patient's forehead and along the jaw. These electrodes allow software to filter out any artifacts caused by muscle movement from those areas. Figure 10, in Section 3.1.1, shows an example of raw EEG voltage data from a single electrode, with time along

the x axis, and voltage along the y axis.

2.2.2 Interictal Spikes

Epileptic seizures are characterized by excessive, synchronous abnormal firing of neurons in the brain [17, 30]. Monitoring the brain for extended durations can show various patterns that arise as the patient is observed. One of these patterns, the interictal spike, tends to occur in the region of the brain where seizures emanate from, but the spike itself is not a seizure [6, 29]. These interictal spikes are useful for doctors since they provide information about the regions of the brain where seizures originate. With this information, neurosurgeons attempt to pinpoint the region of the brain where seizures propagate from and surgically remove it.

While spikes may differ from patient to patient, the general characteristics include a sharp change in amplitude of the voltage in a relatively short amount of time, much like a high frequency pulse. Figure 2 shows an example of an interictal spike after being run through a low pass filter. The feature itself is approximately 100ms in duration, exhibits a sharp drop in amplitude followed by an equally sharp rise in amplitude. This sharp change in amplitude is uncharacteristic of the surrounding region, making it stand out enough for a doctor's eye to identify.

Figure 3 shows two interictal spikes as seen in the surrounding environment. They can be seen as the sharp, negative pulses just before samples highlighted by the red line. This plot also shows the subtlety of the spike in the surrounding signals and how hard it can be to detect with an untrained eye.

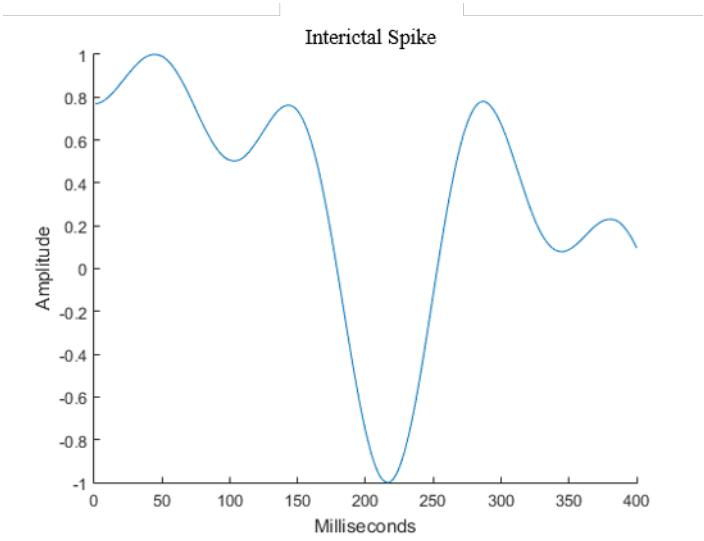


Fig. 2.: A sample interictal spike after filtering. Voltage values normalized.

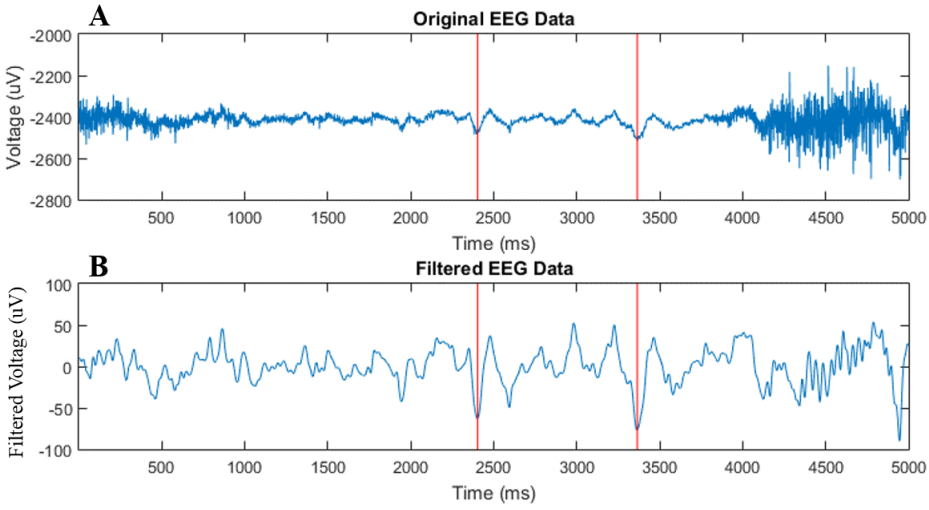


Fig. 3.: Original EEG data plotted with EEG data after being run through the bandpass filter.

2.2.3 Current Detection Methods

To detect these epileptic spikes, neurologists manually (by eye) analyze the time-domain voltage waveform across multiple electrodes (256 electrodes). This process is both time-consuming and potentially inaccurate [7, 43, 16, 1]. The doctor is prone to fatigue, boredom, and any number of psychological effects that occur when performing repetitive, monotonous tasks. Thus, the standard approach for spike detection is still performed by physician inspection of frequency filtered EEG data.

The patient data used in these studies came from the EGI Dense Array EEG. It has 256 electrodes that record the voltage differential across the brain from their specific location on the scalp. Visually identifying these spikes requires a trained observer who knows what a normal EEG reading looks like at any given time. The doctor must be able to differentiate patterns from noisy interference, such as muscle artifacts, and patterns from normal brainwave activity, such as sleeping or a normal alert state. Identifying the subtle spike pattern amidst the range of EEG patterns provides an opportunity to ease the workload of the doctor by automating part of this process. Furthermore, robust detection of epileptic spikes may improve the efficiency of surgical interventions by more accurately pinpointing the spike source.

The ultimate goal of finding these spikes is to determine where in the brain they are coming from. This is done by analyzing which electrodes exhibit the spike patterns and triangulating the region in the brain where they most likely originate from. A twenty hour EEG reading may yield hundreds or thousands of these spikes, however it is not essential to identify every single one to solve the problem. The doctor can reliably locate the origin of the interictal spike with something on order of 100-150 spikes depending on the patient and quality of the spikes examined. This lets the automated system miss some interictal spikes to further reduce the dataset.

2.3 Artificial Neural Networks

Artificial neural networks (ANNs) are biologically inspired mathematical constructs that can be used to approximate nonlinear, multidimensional functions. These networks are capable of learning the interdependencies between the inputs and outputs of a system. A neural network is composed of a network of artificial neurons, which attempt to mimic their biological counterparts. Each neuron produces an output based on the summation of its inputs and comparing it to a threshold function. The inputs and outputs of each neuron can be from/to other neurons respectively. Linking multiple neurons together allows for an output to be produced based on system of weighted inputs.

Each individual neuron is represented by an input x_q multiplied by a weight w_q. A network of neurons is then taken as a summation of these weighted inputs:

$$net = \sum_{q=1}^{n} w_q x_q \tag{2.1}$$

The output for each neuron is determined by an activation function that behaves like a threshold. The output of an individual neuron is:

$$o = f_s\left(\sum_{q=1}^{n} w_q x_q\right) \tag{2.2}$$

with $f_s(x)$ representing the activation function. A typical activation function for ANNs is the sigmoid function:

$$f_s(x) = \frac{1}{1 + e^{\lambda_s x}} \tag{2.3}$$

The hyperbolic tangent function is also frequently used in artificial neural networks:

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{2.4}$$

A network of neurons has numerous neurons interconnected, possibly in multiple layers as shown in figure 4. There is one input layer, one output layer, and any number of hidden layers in between.

The output of a neuron i within a network layer $l+1$ is given by:

$$o_i^{l+1} = \sum_{j=1}^{S_l} w_{ij}^{l+1} o_j^l + b_i^{l+1} \tag{2.5}$$

with S_i being the number of neurons in a given layer l, w_{ij}^{l+1} being the weight of the connection from neuron j in a given layer l, b_i^{l+1} is the bias of neuron i and a_j^i is the output of neuron j in a given layer l.

Using this, the error for a given layer L can be calculated if the desired output is known by using:

$$E = \sum_{p-1}^{P} \sum_{m-1}^{M} (d_{pm} - o_{pm}^L)^2 \tag{2.6}$$

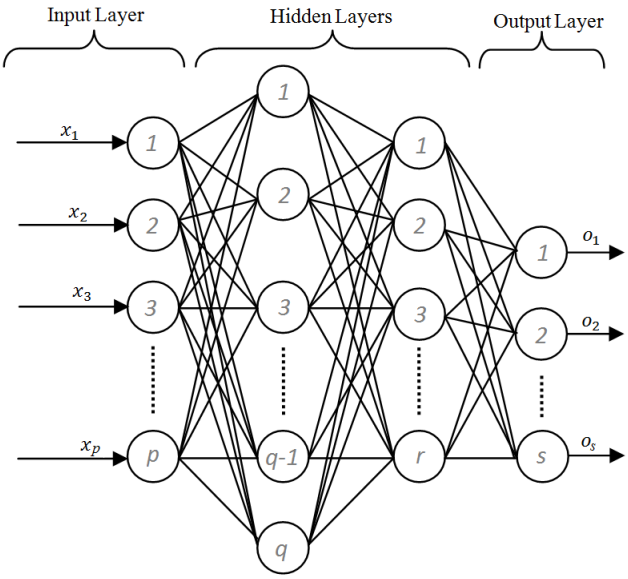


Fig. 4.: An example ANN architecture.

2.4 Convolutional Neural Networks

Convolutional neural networks (CNNs) are very similar to classical artificial neural networks in that they use a system of weighted inputs to produce an output. The primary difference in the two is in the way they partition the data to be fed into the network.

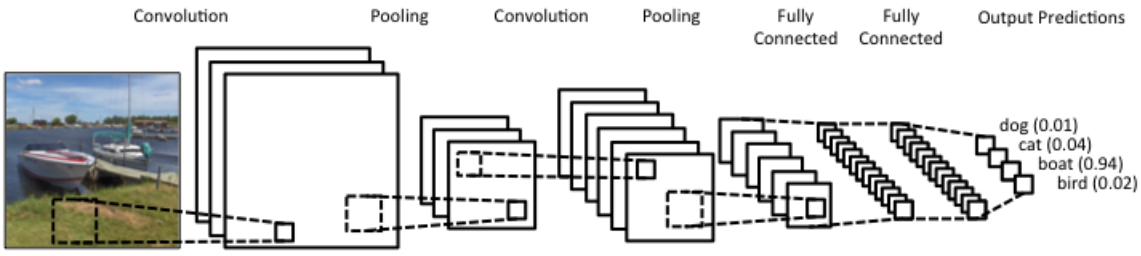


Fig. 5.: Overview of a convolutional neural network architecture [10].

The CNN is heavily inspired by the visual cortex in animal brains [18, 12]. The brain views an image as a whole through smaller filters called receptive fields. These

receptive fields are tiled over the entire view and allow spatially relevant information to be extracted from each region of the visual field. A CNN takes this tiling approach and applies it to an input set of data, treating it like an image and sliding a tiled window across the data. They have been used to increase accuracy of computer image recognition and processing [26, 24, 36].

A major motivation for utilizing CNNs comes from the problems that arise in traditional ANNs when dimensionality increases. With a fully connected ANN, a 32x32x3 image would generate an input layer with 3,072 weights. This can be manageable with power systems, however scaling to a larger image creates very real problems. The EEG in this study has 256 electrodes, and the interictal spike width is approximately 120 samples in time. This adds up to an input layer with 30,720 weights. An image of this size with multiple channels for color can have an input layer with nearly 100,000 weights. This size is simply unmanageable in terms of processing power and memory requirements.

A CNN utilizes the tiling method along with various types of stacked layers to increase the speed of training and provide spatial relevance of data within the image. Trading out the fully connected ANN, a CNN only passes locally connected neurons onto the next layer by the use of a smaller filter. Figure 5 shows an example of a CNN architecture. Multiple layers are involved, including convolutions, pooling, and fully connected layers.

Figure 6 gives a more in depth look at a convolution layer. This figure shows an input layer splitting into a convolution layer through three filters. Figure 7 shows how the filter passes information to the next layer. The 3x3 filter in the image will slide along the entire image, stepping by a variable number of elements. This step size is known as the stride.

One of the layers typically used in CNNs is called a pooling layer. A pooling

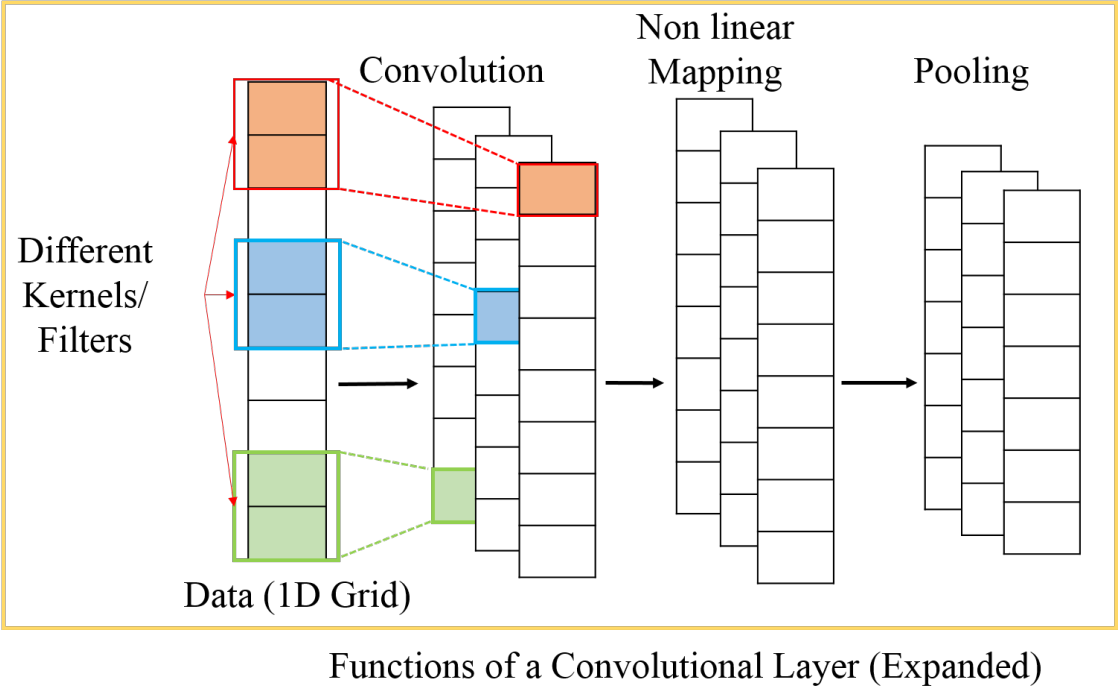


Fig. 6.: A detailed look at a typical convolution layer within a network.

layer aggregates information from a 'pool' of elements into a layer of reduced size. Pooling typically behaves as a sub-sampling tool, as it reduces the input image size by a factor of the pool dimensions. This reduction in spatial size of the input image allows for fewer numbers of parameters to compute and helps reduce overfitting. In Figure 8, a max pool of size 2x2 is used. This takes the maximum value of the 2x2 area and passes it on to the next layer. Max pooling is the most common function to use in a pooling layer, however mean pooling is also used.

The final major difference in a CNN is the activation function used between convolution layers. The rectified linear unit (ReLU) function is typically used in CNN architectures, and is defined as:

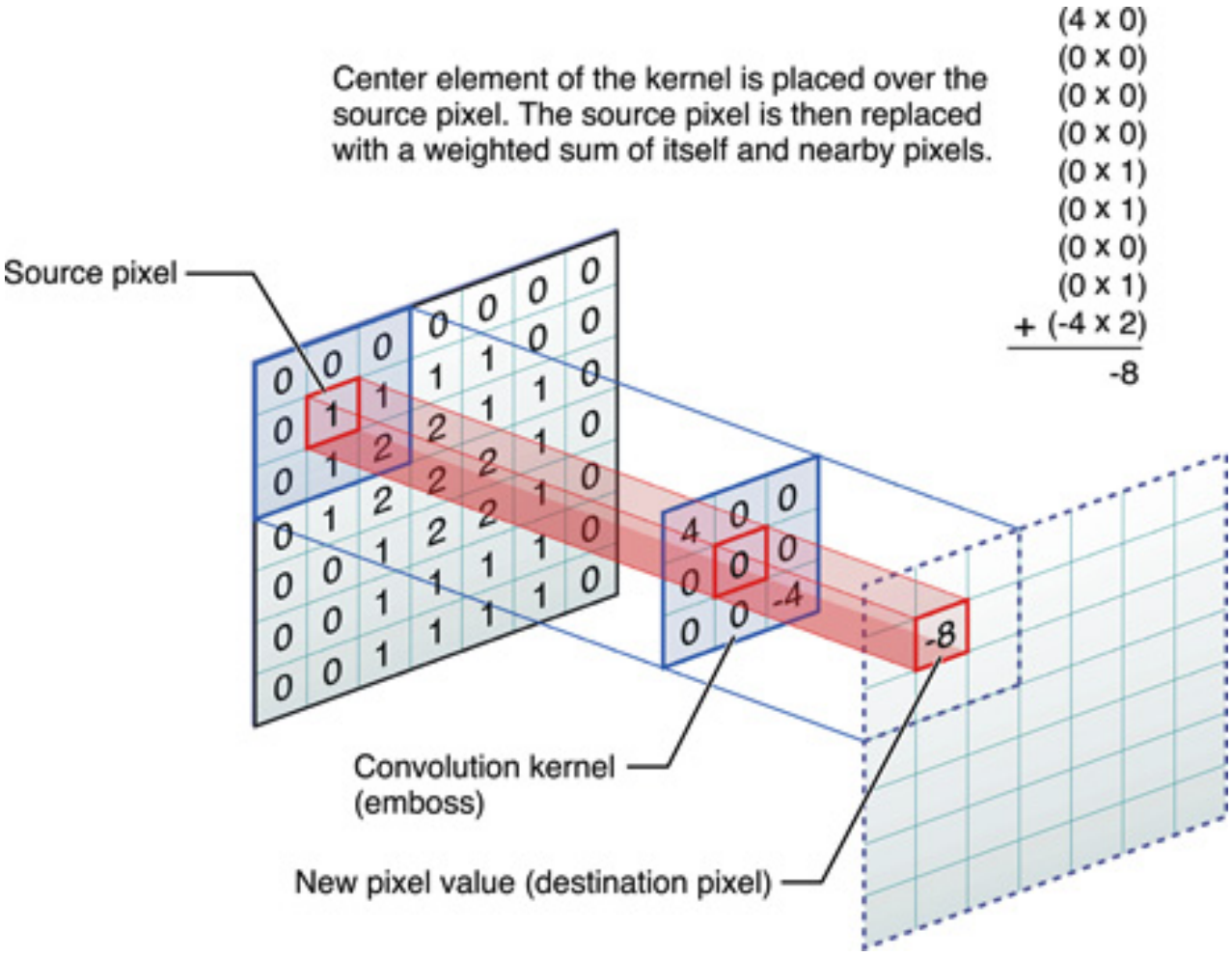


Fig. 7.: Shows how the filter of a convolution passes information on to the next layer. The filter size in this picture is 3x3, passing a single value on to the next layer [5].

$$f(x) = \begin{cases} x & \text{if } x > 0 \\ 0 & otherwise \end{cases} \tag{2.7}$$

ReLUs have been shown to have a number of advantages over activation functions like the sigmoid in Equation 2.3, or the hyperbolic tangent function in Equation 2.4. It is computationally faster, having no exponential calculations, only comparisons, summations, and multiplications. It provides a more sparse activation, only neurons

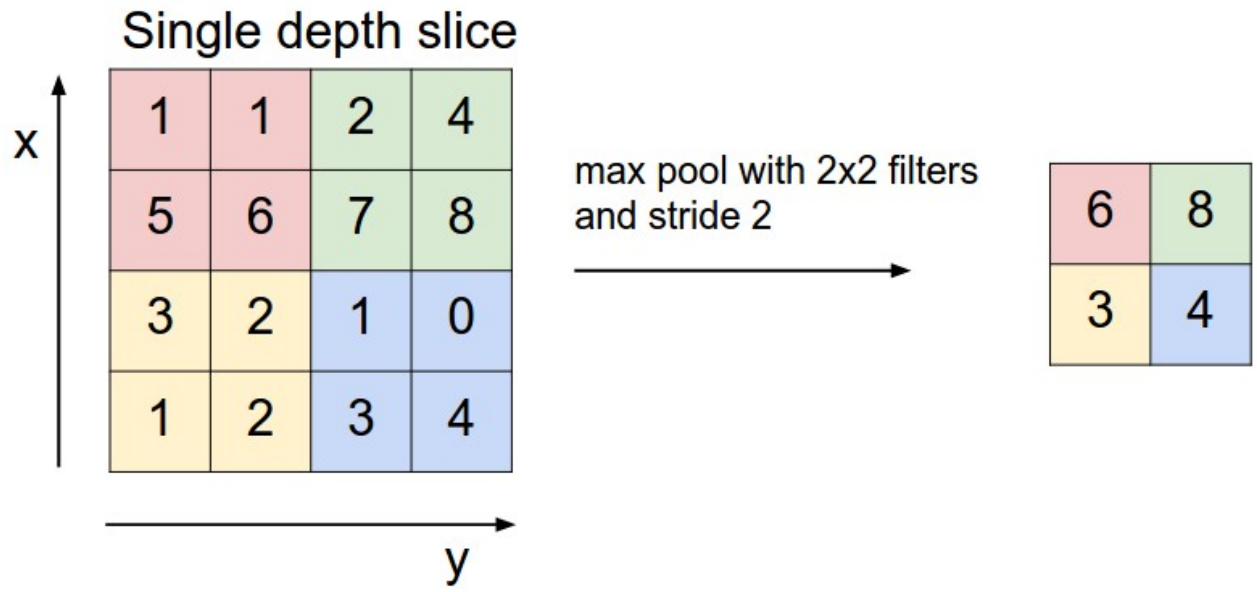


Fig. 8.: Shows a max pool of size 2x2 with a stride of 2. Reduces the input image from 4x4 to 2x2 [41].

with non-zero outputs will fire. They are scale invariant, and been shown to reduce the likelihood of a vanishing gradient in deep layered networks [45, 28].

Chaining multiple layers together creates the architecture for the CNN. Each layer serves its own purpose, as described above, and it is up to the user to determine which type of architecture best solves his or her problem.

CHAPTER 3

SINGLE PATIENT ANALYSIS

3.1 Preliminary Data Processing

3.1.1 Raw Data

The data used in this experiment comes from an EGI Dense Array EEG [11] used in a clinical study at Virginia Commonwealth University Health's (VCU Health) Department of Neurology. EEG data from five different patients was analyzed for this study. A single patient reading could last between 5 and 24 hours or more. The Dense Array fits around the head much like a hair net, as shown in Figure 9. The cap must be aligned very precisely around the patient's head to make sure the electrodes are over the correct regions of the brain and face. This is a time consuming process that makes taking an EEG reading a non-trivial matter.

The sampling rate of the EEG was 1000 Hz, or one sample every millisecond. This translates to roughly 20 million data points for each electrode on the lower end of the time scale. With 256 electrodes and 20 million data points, the amount of the data is roughly 25 GB per patient for the raw data alone. For this reason, only small portions of the data were actually analyzed. The size of the analyzed sections ranged from approximately 70,000 to 250,000 samples, or roughly one to five minutes of continuous EEG data. The selected regions were chosen due to their high frequency of doctor-annotated interictal spike activity.

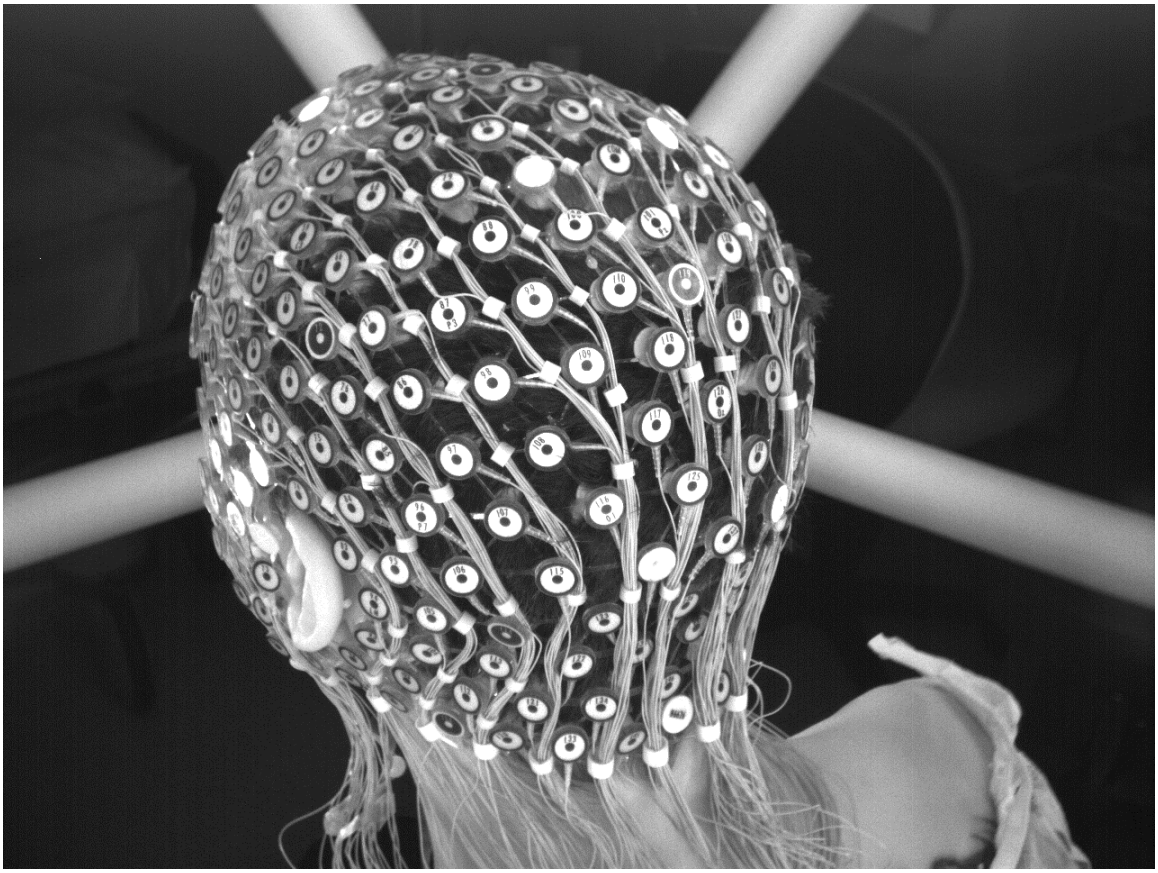


Fig. 9.: A patient wearing the EGI Dense Array EEG.

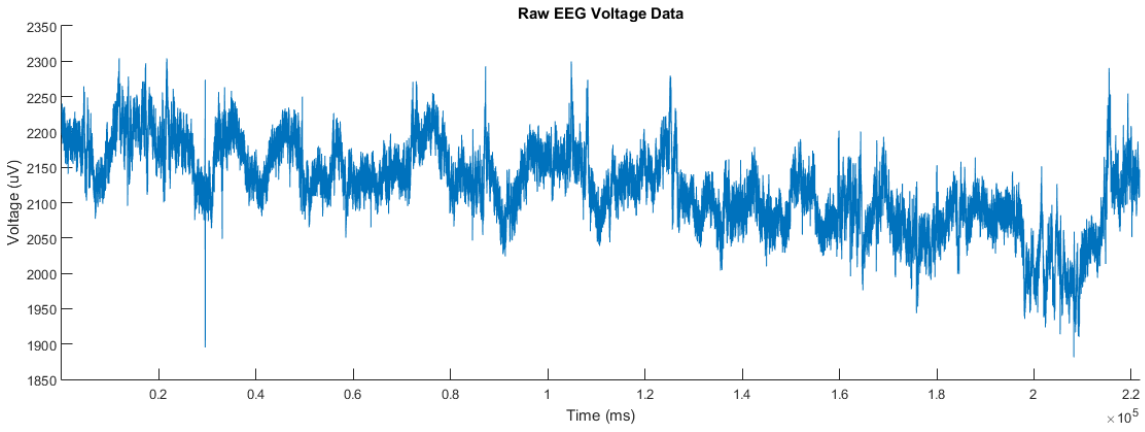


Fig. 10.: Single channel of raw EEG data showing high frequency noise.

3.1.2 Signal Processing

The information output from the EEG data is a voltage signal propagating over time. Figure 10 shows the raw voltage as a function of time from a single electrode of EEG data. The signal is visibly noisy over the entire duration. This noise may come from muscle artifacts, a strong, errant EM wave from some source within the hospital, or something similar. Any patterns exhibited in this noise would be nearly impossible for a human eye to discern. To reduce the high frequency noise, the raw data is run through an infinite impulse response (IIR) Butterworth band pass filter from 5-30 Hz [38]. In effect, this band pass filter cuts off most frequencies below 5 Hz and above 30 Hz. To best approximate the neurologist's filtering process, the parameters chosen are displayed in Table 1.

The high pass filter cutting off any frequencies below 5 Hz acts as a de-trending tool. Any low frequency shift that occurred across the electrodes over time is eliminated by this filter. Figure 3A-B shows a comparison between the raw voltage values and the filtered voltage values before normalization. To standardize the range of the data, the max and min values were normalized to 1 and -1 respectively.

Figure 11 shows another look at filtered data, but from the viewpoint of the software used by the doctors at VCU Health for analysis. This screenshot is about $\frac{1}{3}$ of the width of the screen, cut to fit on page. The vertical blue bars represent a single second in time, while the horizontal blue bars separate out different electrodes. This snippet shows 34 electrodes on screen and approximately five seconds of EEG data. Multiple interictal spikes can be seen as sharp, relatively large amplitude dips on electrodes F4-C4, Cz-C4, F8-T8, and maybe others. Eyeblinks can be seen as longer duration, larger amplitude dips on electrodes Fp1-F9, Fp1-F7, Fp1-F3, and maybe others. The doctor must manually scroll through the hours of EEG data, second

Table 1.: Filter Settings

Passband Frequency Low	5 Hz
Passband Frequency High	30 Hz
Passband Ripple	0.5 dB
Sample Rate	1000 Hz

by second analyzing 256 electrodes by eye, attempting to locate as many interictal spikes as possible. At the bottom of the image is a yellow bar, this represents the time scale for the entire dataset. There is a small gray box on the far left, this is the time window box that is displayed on the full screen. This serves to give a scale for the amount of data the doctor has to scroll through. The gray box starts at the far left of the yellow section and slides across the entire bar to the end. On screen at any given time is a few seconds of data to maybe dozens of seconds. Scrolling through and analyzing multiple hours of data by chunks of seconds takes a significant amount of time.

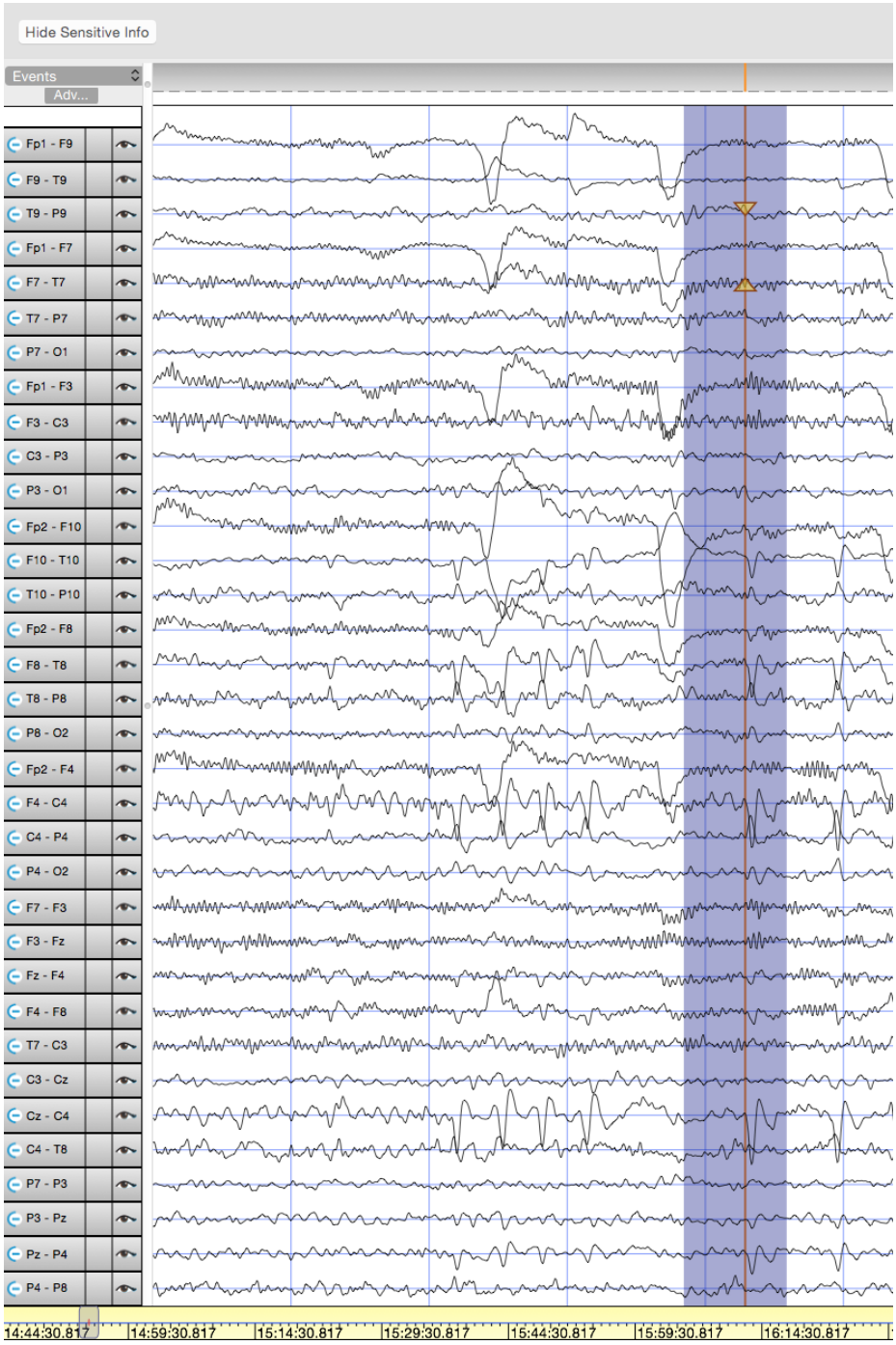


Fig. 11.: Example of filtered EEG data as viewed by the doctor's software. The faint, blue, vertical bars represent one second in time. Each horizontal signal represents a different electrode. Interictal spikes, eyeblinks, and other brain activity can be seen in this snippet.

3.2 Single Electrode Analysis

The initial analysis of this problem was performed only on a single electrode. This allowed for faster prototyping of methods that could be expanded upon in the multiple electrode approach. This section details the single electrode method.

3.2.1 Why Examine a Single Electrode?

The interictal spike is a point source signal emanating from within the brain, propagating outward. The signal is detected by numerous electrodes close enough to that point source. It is inherently a multi-electrode, or multidimensional, problem, and the reasoning behind examining only a single electrode must be explained.

The single electrode analysis serves as a proof of concept for the problem at large. In the test patients, it is known where the interictal spikes are in time, thus the electrode closest to the point source can easily be found. This electrode will have the strongest interictal spike signal, and therefore have the best definition of what the spike looks like in the most optimal of circumstances. These spikes are very subtle, as discussed in section 2.2.2, and the framework for detecting spikes should at least be able to discriminate between a spike and a nonspike when shown the optimal spike figure. This section documents the work done to verify the viability of a neural network in identifying and filtering interictal spikes as presented in Carey et al. [8]. Determining whether neural networks can be used to accurately identify interictal spikes is the goal of this section.

3.2.2 Data Preparation Framework

This section presents the framework designed to train an artificial neural network to discriminate between EEG patterns. Figure 12 makes reference to Raw EEG data

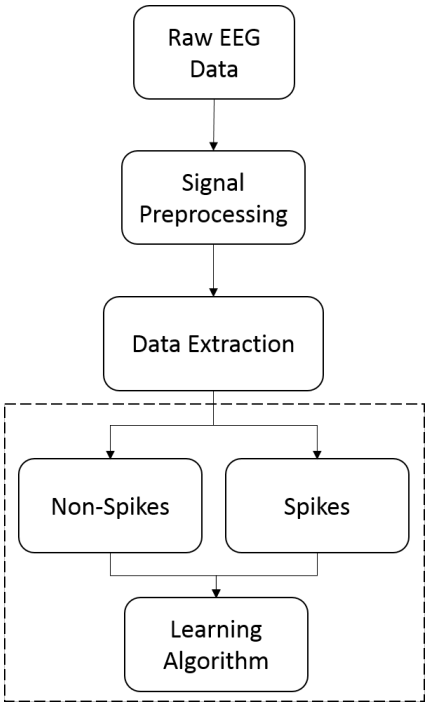


Fig. 12.: Flow diagram documenting the overall design of the process

and Signal Processing as that were explained in Section 3.1. The rest of the diagram sections are explained here.

Out of the multiple hours of EEG reading and subsequent millions of data points, proportionally very few features were identified by a neurologist as interictal spikes. However, these spike features did not occur on every electrode at every labeled time step. These annotations were spread out across the entire data sets, with clusters of spikes occurring in various regions of data.

Figure 13 shows a visual representation of the data extraction process. To capture the spike, a sliding window technique was used to analyze each section of data. The sliding window was set to the size of the spike (120 samples). Sliding this window across all the samples for each patient allowed for a direct comparison of each timestep to the spike.

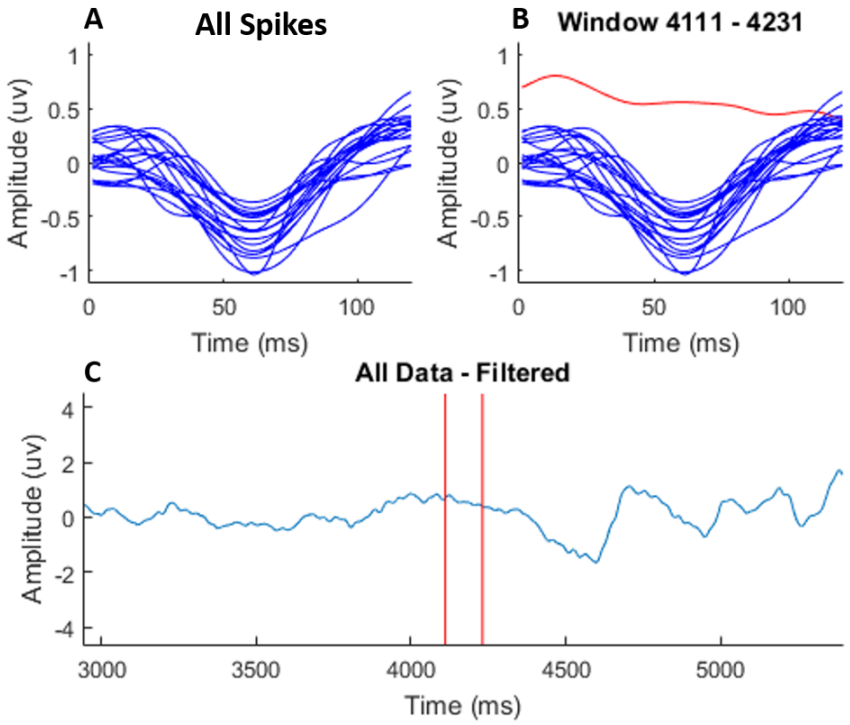


Fig. 13.: A shows all labeled spikes from patient 1. B shows all the features plotted along with a sample of non-spike data. C shows the filtered data with the bars representing the area plotted in B. All voltage values normalized.

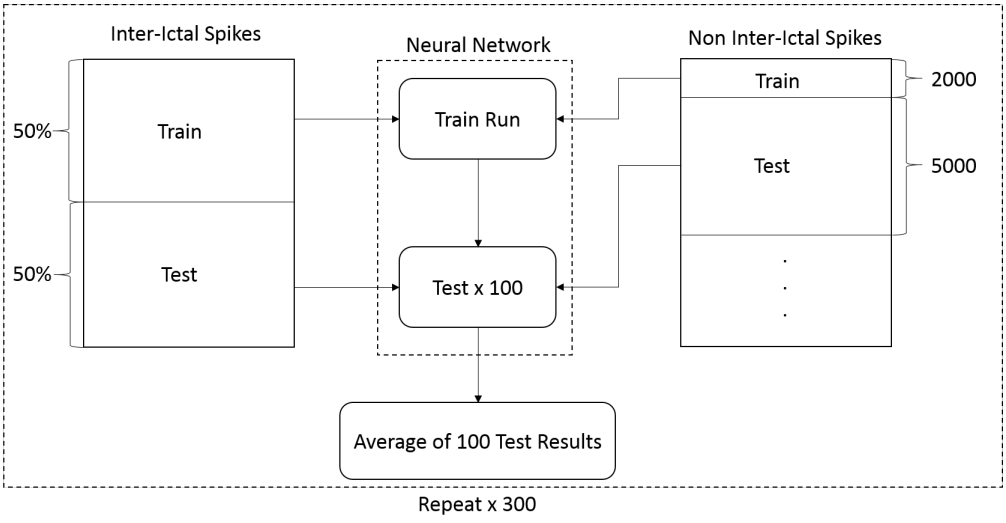


Fig. 14.: Framework showing method used to train and test single electrode analysis.

Figure 14 shows the process used to select training and testing features for the neural network. The data was first split into both spike sets and non-spike sets. All windows that included spike data were removed from the non-spike data to prevent contamination of the training set. Of the annotated spike windows, a random 50% were used for training, and the other 50% for testing. A continuous set of 2000 random non-spike samples were used for training, and a continuous set of 5000 random non-spike features were used for testing.

For the patient datasets examined in this approach, anywhere from 6 to 14 interictal spikes existed within the portion of data being analyzed. With each spike window being 120 samples in time, this represents from 720 to 1680 samples of ground truth. Since the windows of spike and nonspike data were extracted and separated from eachother, this led to a massive imbalance of 6-14 spikes and 2000 nonspikes for training. However, half of the spikes were used for testing, leaving 3-7 spikes for training, further increasing the imbalance.

In a learning algorithm such as neural networks, this imbalance in represented data is a serious challenge. If these numbers were used as-is, the learning algorithm would increase its accuracy simply by classifying all of the samples as that of the largest set, in this case the nonspikes. Classifying 2000 nonspike correctly, and only missing 3-7, would yield a 99% accuracy. However, this would not be very useful in finding interictal spikes at all. To help alleviate this problem, the spikes used for training were upsampled to be a larger portion of the overall data percentage. This allows the neural network to see the spike samples more frequenly and adjust its weights toward classifying those patterns correctly. In effect, this method weights the spike samples higher than the nonspike samples by showing it the same few spikes many more times.

Once the neural network was trained, it was tested 100 times on separate con-

tinuous sets of 5000 random non-spike features along with the remaining test spike features. This process of training and testing over 100 regions was repeated a total of 300 times, with each network being trained on randomized spikes and randomized non-spike features. The total number of tested samples was 500,000 for each trained network.

The neural network design, shown in Figure 15, followed the standard feed forward algorithm using the back propagation algorithm for a multilayer perceptron (MLP), and was implemented using Matlab's Neural Network Toolbox [39]. The neural network contained a single hidden layer with 10 neurons. The output layer consisted of one binary neuron using a sigmoid activation function to determine whether the individual feature being analyzed was a spike or not a spike. The input layer of the neural network contained 120 neurons. The input to these neurons was the 120 values from the windowed datasets. This allowed the neurons to learn on the individual values that represented each window. Figure 13-A shows a sample of windows that contained spike features. While the distribution of amplitudes is not tightly bound, there is a clear pattern in the amplitude drop. Figure 13-B shows a windowed feature from a non-spike region. The non-spike pattern clearly does not exhibit the drop in amplitude of the spikes. This difference in pattern is what the neural network learns, and classifies each sample accordingly.

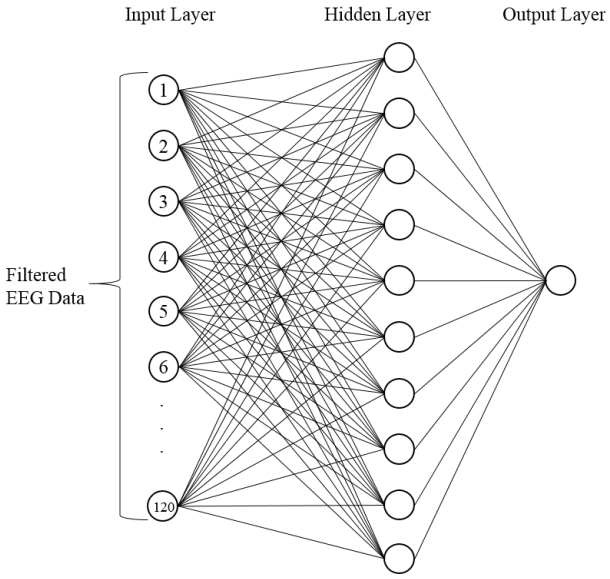


Fig. 15.: Design of the ANN used to train the single electrode analysis.

Table 2.: Number of annotated spikes, total samples, and percentage of spike data per patient

Patient	Number of Spikes	Total Samples (ms)	Samples Annotated as Spikes (%)
1	6	221,710	.32%
2	7	179,305	.47%
3	8	171,799	.56%
4	14	134,927	1.25%
5	9	49,282	2.19%

3.2.3 Results

The purpose of the single electrode analysis was to determine whether or not a neural network could discern the interictal spike under optimal circumstances. The results of interest are presented in Figures 16-18, with discussion as to why each metric is important. These results are the average of each train/test phase of each patient. Table 2 shows how many annotated interictal spikes each patient had, as well as the total number of samples per patient. Since each spike window in this analysis was 120 samples, the total number of data that is considered a spike can be seen by:

$$\text{Percent Spikes} = 100 * \frac{\text{Number of Spikes} * \text{Spike Window Size}}{\text{Total Samples}} \tag{3.1}$$

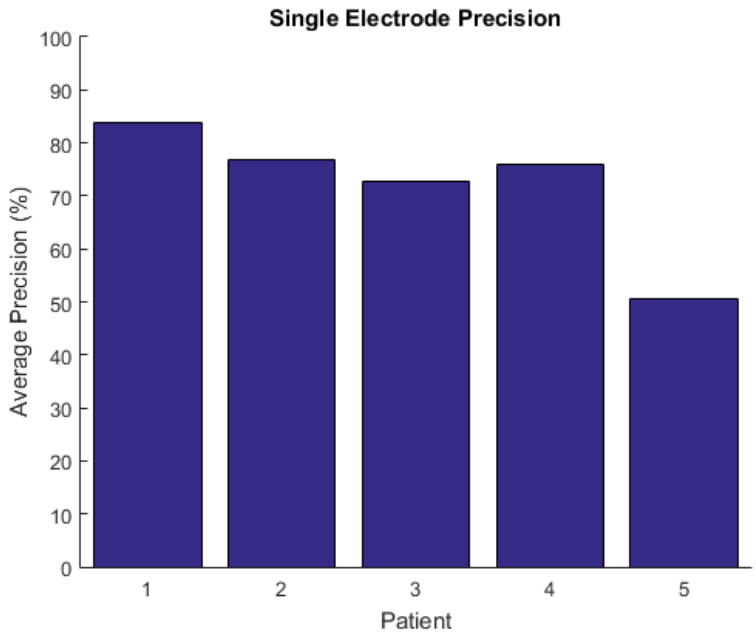


Fig. 16.: Precision averages for all five patients.

Three metrics used in the analysis of the results are precision, recall, and F1 score defined in the following equations:

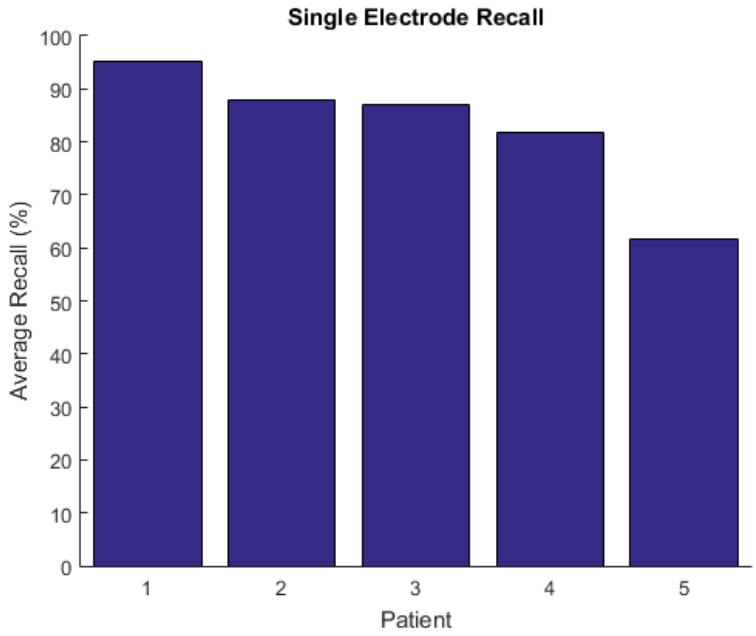


Fig. 17.: Recall averages for all five patients.

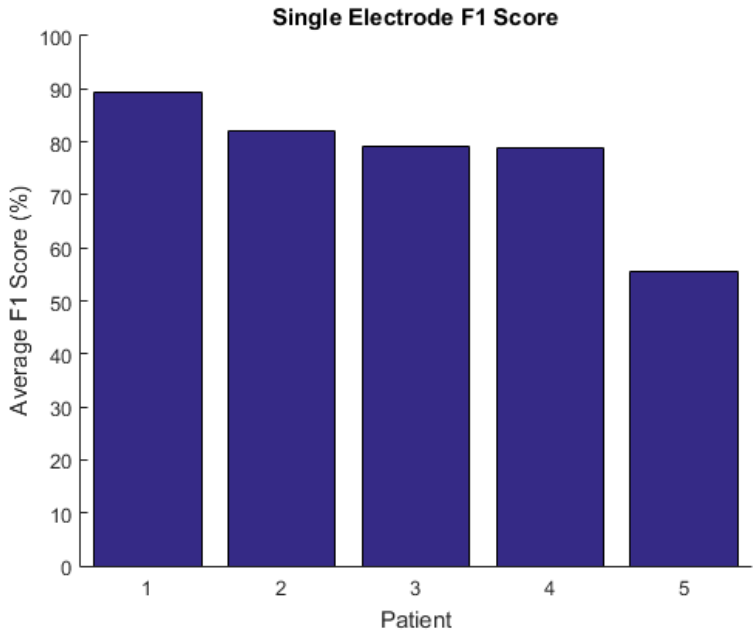


Fig. 18.: F1 score averages for all five patients.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{3.2}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{3.3}$$

$$\text{F1 Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3.4}$$

Precision determines how much of the extracted information is relevant, or how many extracted spikes are actually spikes. It measures the number of true positives relative to false positives of the system. Recall determines how much relevant data was extracted, or how many of the expected spikes were extracted. It measures the number of true positives relative to false negatives of the system. The F1 score is a weighted average of precision and recall, allowing a single value to represent the accuracy of the system.

Analysis of these metrics provides a good understanding of how well the single electrode system performed. Figure 16 shows the system had greater than 50% precision across all patients. This means the system extracted more true positives than false positives for the tested region. When analyzing a final system. Figure 17 shows greater than 60% recall across all patients. This means the system extracted more than 60% of the expected spikes during testing. It is important to note that the objective of this project was not to identify every potential spike in the patient, making the recall metric less important. The goal was to reduce the total amount of data for the neurologist to manually analyze. Reducing the number of false positives, measured by increasing the precision score, will drastically reducing the total amount of data needing to be analyzed. If these results are extrapolated out to the size of a full EEG reading, the total amount of data can be analyzed using the equations

below:

$$\text{False Positives} = \frac{(\text{Total EEG Data Size} * \text{False Positives})}{\text{Test Sample Size}} \tag{3.5}$$

$$\text{True Positives} = \frac{(\text{Total EEG Data Size} * \text{True Positives})}{\text{Test Sample Size}} \tag{3.6}$$

Below is an example using patient 1 with data from the confusion matrix table 22 in the appendix. This math is assuming an EEG reading of 20 million values, with each value being a millisecond in time, which equates to about a 5.5 hour EEG reading:

$$\text{False Positives} = \frac{(20,000,000 * 0.55)}{5,000} = 2199.07 \tag{3.7}$$

$$\text{True Positives} = \frac{(20,000,000 * 2.86)}{221,710} = 257.99 \tag{3.8}$$

In a live environment, this system would extract into the final dataset anything it classified as a spike, so the total number of true positives plus false positives. The doctor doesn't analyze each sample one at a time, but in chunks, analyzing the waveforms and determining whether they exhibit the interictal spike characteristics. It would be disingenuous to simply consider the entire dataset as a whole, so instead the total dataset should be divided by the size of the spike window. The equation below shows how the total dataset reduction is calculated:

$$\text{Data Reduction Percentage} = 100 - (100 * \frac{\text{True Positives} + \text{False Positives}}{\frac{TotalEEGDataSize}{SpikeWindowSize}} \tag{3.9}$$

Table 3 shows the total number of true positives and false positives as well as the percentage of total dataset reduction if the results from the study are extrapolated out to an EEG reading size of 20 million datapoints. A doctor would only need to analyze a few thousand potential spike windows in the best case to a little over 10 thousand in the worst case. A doctor need only analyze as much as 10% or as little as

1.5% of the previous number of samples. While not perfect, it could save many hours. Table 4 shows how much of the remaining data is actual spike information. This is important because it determines how useful the extracted data truly is. Using patient 1 for example, the doctor only has to manually analyze 1.47% of the original data, however only 11.73% of the that extracted information contains spikes. This means there will be a very large number of potential false positives to filter out manually. However, the total amount of data needing to be analyzed is still significantly lower compared to current practices.

It is important to note that while the number of false positives is significantly higher than true positives, the total amount of data for manual analysis is still drastically lower than before, with a substantial increase in the percentage of data being spikes versus nonspikes. The identification of the point source for the spike activity does not require every possible spike to be identified. According to the doctor involved with this study, accurately extracting 150 spikes could allow for an accurate model to be used for treatment. The results from table 3 show those kind of true positives could be extracted if the density of spikes across the duration of the reading is relatively consistent with the tested region.

The single electrode analysis showed that reduction of the original EEG data can be done using ANNs with a fairly simple architecture. This directly addresses the goals of this section, determining whether ANNs can be used to identify the interictal spikes of the patients used in this study from the noise of an EEG reading. There is much room for improvement, and the multiple electrode analysis will expand upon these results in the next section.

Table 3.: Extrapolated true positives, false positives, and dataset reduction percentage for all patients. Assuming 20 million total EEG reading dataset.

Patient	True Positives	False Positives	Data Reduction %
1	257.99	2,199.07	98.53%
2	392.82	4,283.61	97.19%
3	405.11	5,258.66	96.60%
4	849.35	7,259.47	95.13%
5	1,252.66	12,114.67	91.98%

Table 4.: This table shows how much of the reduced data is spike data, referencing the figures from Table 3.

Patient	Data Reduction %	Data Remaining %		Spike Data %
1	98.53%	1.47%	$\rightarrow$	11.73%
2	97.19%	2.81%	$\rightarrow$	9.17%
3	96.60%	3.40%	$\rightarrow$	7.70%
4	95.13%	4.87%	$\rightarrow$	11.70%
5	91.98%	8.02%	$\rightarrow$	10.34%

3.3 Multiple Electrode Analysis (MLP)

3.3.1 Why Examine Multiple Electrodes?

The single electrode test validated that the data could be classified accurately in the most optimal circumstances. On a new patient, the spike locations are unknown, and the electrode closest to the source must be found. The doctor's manual approach is to view multiple electrodes at once and discern patterns in the signals that occur across multiple electrodes. The key to this approach is the correlation of the spike pattern across multiple electrodes. Since the interictal spike behaves like a point source, the electrodes closest to the source will see the largest voltage spike, with the value decreasing radially outward. Capturing a large number of electrodes exhibiting the same pattern rarely occurs naturally in the brain outside of large amplitude muscle artifacts, and is likely an interictal spike. The goal of this approach is to combine the electrode correlation aspect with the pattern similarity between the interictal spikes.

3.3.2 Template Matching Baseline

Machine learning algorithms are computationally expensive, requiring training time, testing, and tuning of the hyperparameters that are the design of each individual network. This thesis focuses on designing various neural networks to analyze the features of manually annotated interictal spikes, and classify the remaining test material accordingly. Signal analysis is not a new field, and machine learning algorithms are not the only method available to match signals to one another. A more straight forward approach is to use simple template matching.

As a baseline for comparison, a spike template for each patient was constructed from the average of half of a patient's annotated spikes. This template was the width of the spike (120 samples). This average template was slid across the entire dataset

for each patient, with the mean-squared-error (MSE) being recorded for each window. To determine whether a given signal was an interictal spike, a threshold was set so that any window with an MSE lower than this threshold would be classified as a spike. Five thresholds were tested, ranging from the standard deviation of the spikes used to create the average set to the variances, as defined by:

$$\text{Variance} = \text{Standard Deviation}^2 \tag{3.10}$$

The five thresholds were spread evenly between the standard deviation and the variance, with the best threshold being recorded.

Table 5.: Hardware specifications used to process the template matching analysis in 3.3.2

Hardware	Details
CPU	Intel Xeon X5690 @ 3.33GHz

This threshold can be changed to best fit the patient's spikes, reducing the number of false positives and increasing the number of true positives, however choosing this threshold is a difficult task, and one that cannot be easily generalized across patients. There is no simple, easy process to choose a good threshold. This is where the benefit of learning algorithms come in. A learning algorithm uses its training data to select a number of thresholds for each feature of the input space, and classifies each sample accordingly. The system specifications used to analyze the template matching system are shown in Table 5. A single core was used with no parallel or hyperthreading performed. The results from the template matching will be further explained in the results section of this chapter.

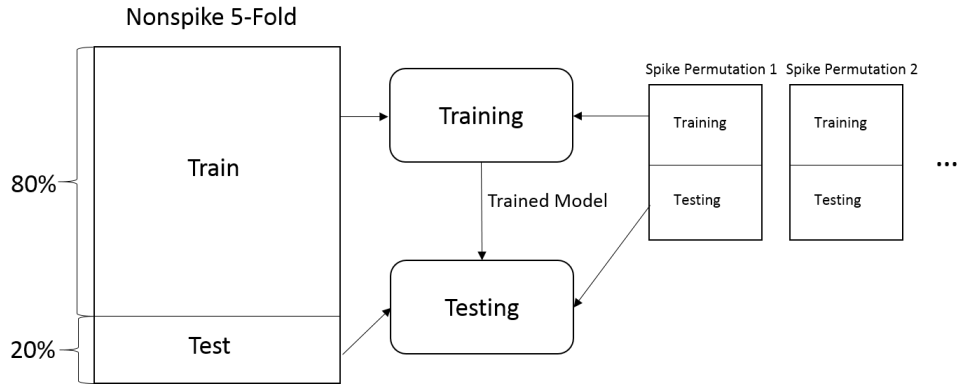


Fig. 19.: Framework showing method used to train and test multiple electrode analysis.

3.3.3 Multilayer Perceptron

The MLP for the multiple electrode test is very similar to the MLP used in the single electrode analysis. It used a single hidden layer with 100 neurons and a single output neuron using the sigmoid activation function for classification. The differences are expanded upon below.

The multiple electrode analysis followed the same methods for raw data handling and signal processing as described in Section 3.1. Figure 19 shows the framework used to train and test the standard classifiers. It is similar to the single electrode approach shown in Figure 14, but has a few key differences. Instead of training on a random 2000 chunk of nonspike samples, a 5-fold cross validation is used to split the data into training and testing samples. However, with as few as six spikes on a single patient, the spike data is too small to be effectively folded across. To address this, five random permutations of spike data were generated to be used in conjunction with each fold of nonspike data, with half to be used for training and half for testing. The training used a batch size of 50 samples, with a 50% chance of each sample being a spike from the training set or nonspike from the training fold. This equal weighting of spike and

nonspike helps alleviate the problem of the massively imbalanced dataset. Since a given training size of spikes is anywhere between 3-7 spikes, each batch will contain multiples of the same spike, essentially weighting the spikes higher than a nonspike.

In the single electrode approach, a single window of 120 features were extracted, representing 120 milliseconds from the EEG. These 120 values were fed directly in to the input layer of the MLP as input values. For the multiple electrode approach, scaling this up to 256 electrodes would create an input of 30,720 values. This is not viable as it would introduce massive memory and processing requirements. To get around this, the values were downsampled by taking every 10th data point from every window.

Table 6.: Hardware specifications used to train the MLP in Section 3.3.3

Hardware	Details
CPU	Intel Xeon E5-2670 v2 @ 2.50GHz
GPU	NVIDIA Tesla K20c

Table 6 shows the system specifications used to train the MLP for this analysis. The bulk of the work was done on the GPU using the Google TensorFlow API [15]. The GPU allowed for much faster training, and therefore prototyping of the architecture of the MLP.

To capture the additional electrodes, the downsampled EEG values were concatenated together to create a single, large input set. This can be seen in figure 20, where 21 electrodes are concatenated together to create a single window of 252 input values. Every twelve values represents a single electrode, shown being separated by the vertical red bars. Figure 20 shows all of the annotated spikes from a single patient overlaid on top of each other. This makes the similarity of the spikes immediately

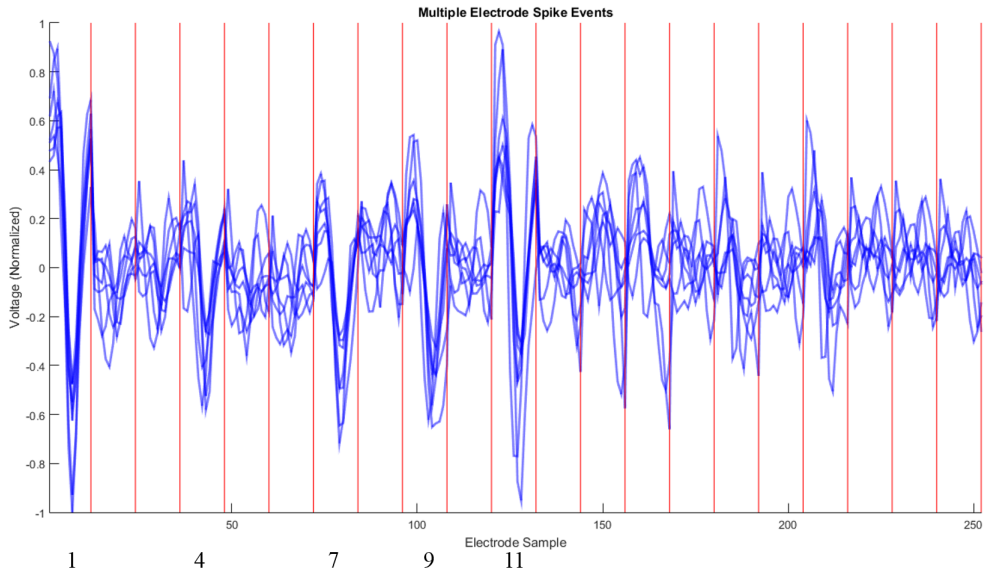


Fig. 20.: Multiple electrode input values showing 21 concatenated electrode windows of interictal spike data.

visible, as well as the correlation across electrodes. The 1st, 4th, 7th, 9th, and 11th electrodes exhibit a very strong correlation in their spike patterns. These five electrodes are all experiencing the interictal spike which is clearly visible, where as there is no correlation between the other electrodes that are not experiencing the spike.

Figure 21 shows seven randomly chosen windows of non-interictal spike data, from the same patient as figure 20, plotted on top of each other. It is clear that the five electrodes from 20 are not exhibiting any kind of correlation during a non-spike event. The combination of this correlation and the spike shape is what makes these events classifiable.

Reducing the number of inputs by an order of magnitude still leaves quite a large input layer for most classifiers. To this point, not every electrode was analyzed at once. Different test sets with increasing numbers of electrodes were analyzed to compare the effectiveness of adding electrodes to the analysis. Figure 22 shows the

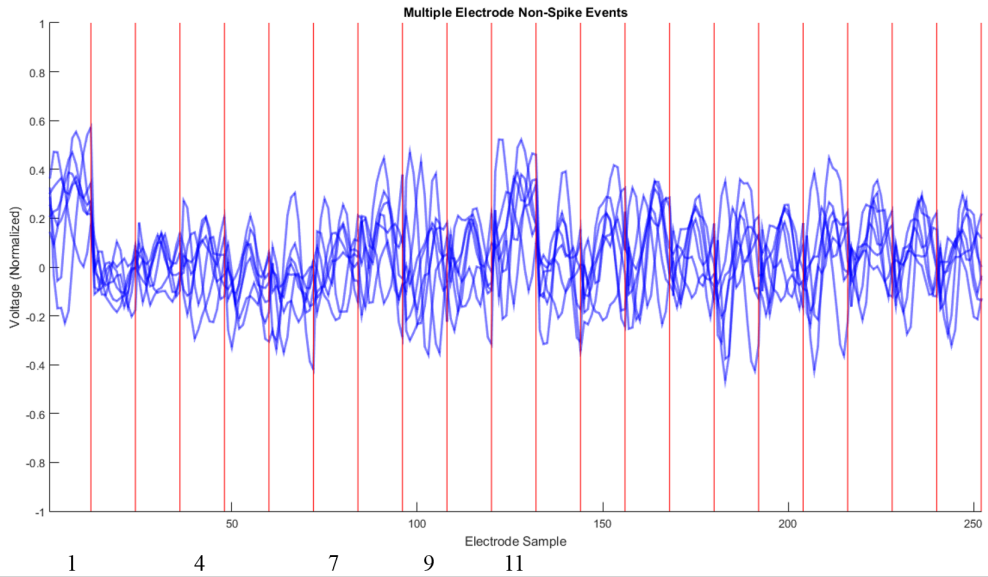


Fig. 21.: Multiple electrode input values showing 21 concatenated electrode windows from 7 random non-interictal spike data samples.

electrode numbering used for the EGI Dense Array. The electrodes in each set were chosen in a spread-out fashion to capture as many regions of the brain as possible. Figure 35, Appendix B, shows the electrode numbers used for the three tests; Figure 22 can be used for reference to determine the location of each electrode. Table 7 shows the total number of electrodes in each set, as well as the number of neurons in the input layer to the MLP. Each spike window was 120 sample long, downsampled by an order of magnitude, leaving 12 samples per electrode. So the input layer is simply 12 multiplied by the number of electrodes used in the test.

Table 7.: Number of electrodes tested and size of corresponding input layer into the neural network for each test set.

Test Number	Number of Electrodes	Input Layer Size
1	21	252
2	39	468
3	64	768

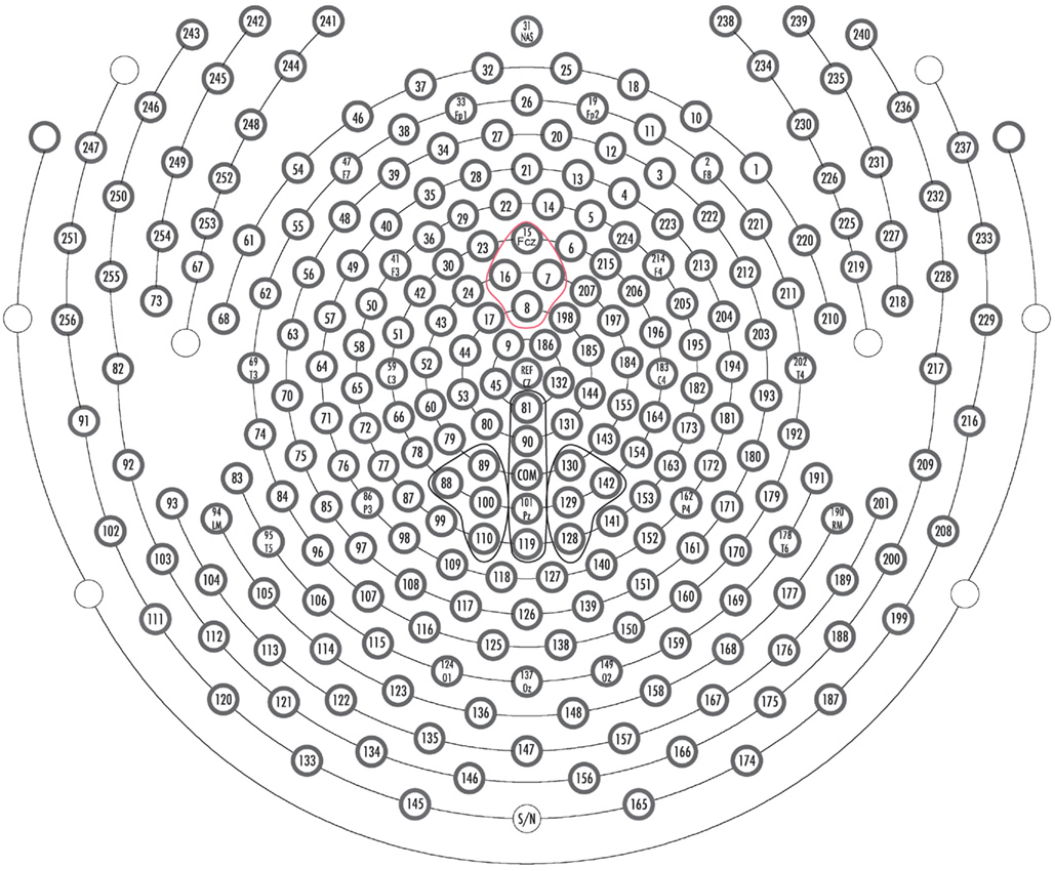


Fig. 22.: Electrode numbering for the EGI Dense Array.

3.3.4 Convolutional Neural Network

Convolutional neural networks operate differently than traditional ANNs, as described in Section 2.4. The data extraction and training/testing samples are created using the same technique as in the MLP analysis, as described by Figure 19. The only major exception to the data extraction process is the lack of downsampling and an increased spike width of 150 ms. The same patients and annotated spikes were analyzed, but the spike window was increased to give the convolution patches a wider window to analyze potential regions of interest.

Instead of concatenating all electrode data end-to-end to form a single vector for the input layer, the CNN takes as input the electrodes and windowed features as a 2-d image. The columns of the image being the voltage data of the spike as it propagates through time, and each row being a separate electrode. The following list enumerates the different layers used in the CNN analysis:

1. Input layer: # Electrodes x Window Size (150)
2. First Convolution Layer: 2x2 filter size, 1 filter, ReLU activation function.
3. Dropout Layer: 50% chance a neuron's output is kept.
4. Second Convolution Layer: 2x2 filter size, 1 filters, ReLU activation function.
5. Dropout Layer: 50% chance a neuron's output is kept.

Table 8.: Hardware specifications used to train the CNN in Section 3.3.4

Hardware	Details
CPU	Intel Xeon X5690 @ 3.33GHz
GPU	NVIDIA GeForce GTX TITAN X

6. Densely Connected Layer: Takes input from previous pooling layer, outputs fully connected layer with 1024 neurons, ReLU activation function.

7. Output Layer: One output layer, sigmoid activation function for classification.

The convolution layers use a smaller filter size to convolve over this 2-d image of the EEG data. They learn features from sub-samples of the image, and pass that on to the next layer. The dropout layer helps prevent overfitting by giving a 50% chance for each neuron's value to be dropped when passing on to the next layer. The densely connected layer essentially brings the system back to the standard MLP, with every neuron from the previous neuron connected to every neuron in the next layer. The sigmoid activation function is used as the final output neuron to classify between a spike or nonspike.

3.3.5 Results

3.3.5.1 Template Matching Results

Similar to the single electrode analysis, Figures 23 - 25 show the precision, recall, and F1 scores for the template matching analysis. Immediately obvious is the poor precision for all patients except patient 1. Patient 1's performance is likely due to the quality of the annotated spikes. The more similar a set of annotated spikes are, the better a template matching system will work. This is because the standard deviation used to calculate the threshold value yield a more precise value, cutting out more false data. The poor precision means the template matching system does not discriminate between true positives and false positives very well. In a live system, this would yield a very large dataset for the doctor to analyze.

Adding more electrodes did not yield better results in most patients. A template matching system that relies on a threshold will likely perform worse when adding

more electrodes unless the majority of the electrodes added contain spikes. The F1 score, Figure 25, shows the overall performance of the template matching system. Patient 1 performs well, while the remaining 4 do not perform nearly as well.

Tables 23-25 in the appendix show the original confusion matrix for the three electrode sets. Those results are used to create tables 9-11, showing how this system would perform if these results are extrapolated out to an EEG reading with 20 million values. Patient 1, for instance, could still achieve a reduction of the dataset by 99% across all three electrode sets. Patient 2, while performing poorly on the metrics discussed above, still manages to reduce the dataset by 91%. However, only 5.65% of the extracted data contains useful information. This would still require the doctor to spend a large amount of time manually annotating spikes. The results for patients 3-5 are poor across the board when extrapolated out. With Patients 4 and 5 exhibiting no data reduction on some electrode sets.

Analyzing the time it takes to perform these tests is important. If there is no time speedup in achieving these results, then there is little benefit to implementing them. Figure 26 shows the time it took for each patient on each electrode set. Patient 1, having the largest amount of data, took just over 160s, or about 2.5 minutes, on the largest of the three electrode sets.

Table 5 shows the hardware used to analyze the template matching system. A single processor with no hyperthreading was used. It is possible to scale this method up to as many processing nodes as needed to increase performance, as analysis on each section is independent from the rest.

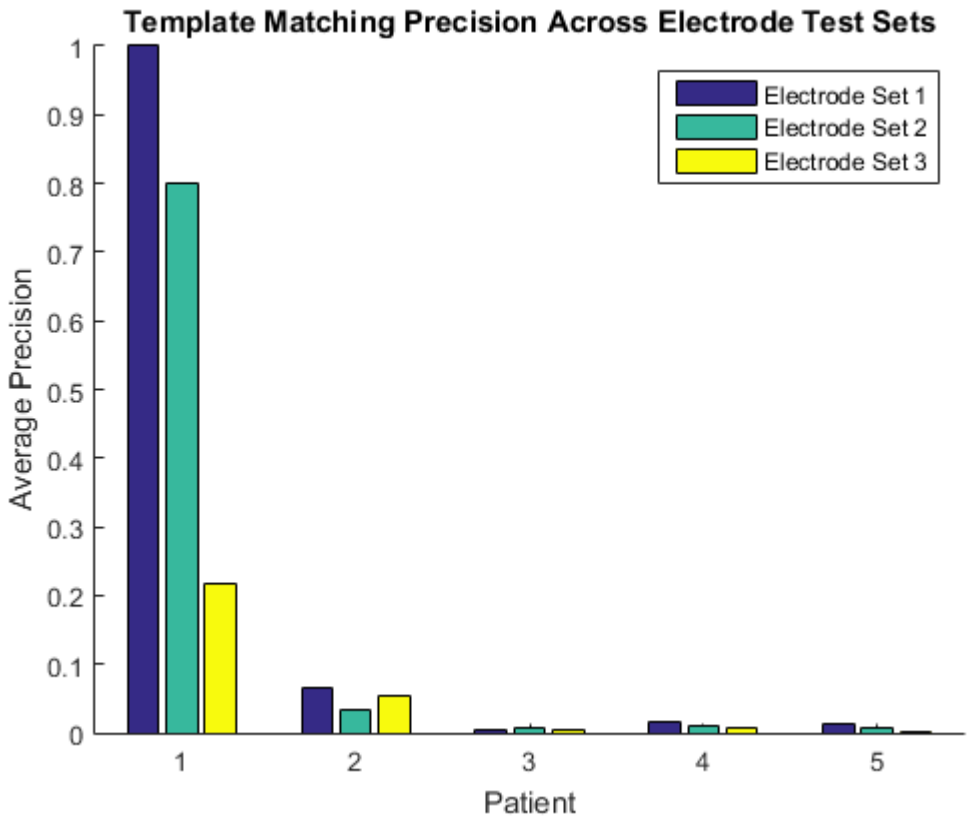


Fig. 23.: Template matching precision averages across all three tests for all five patients.

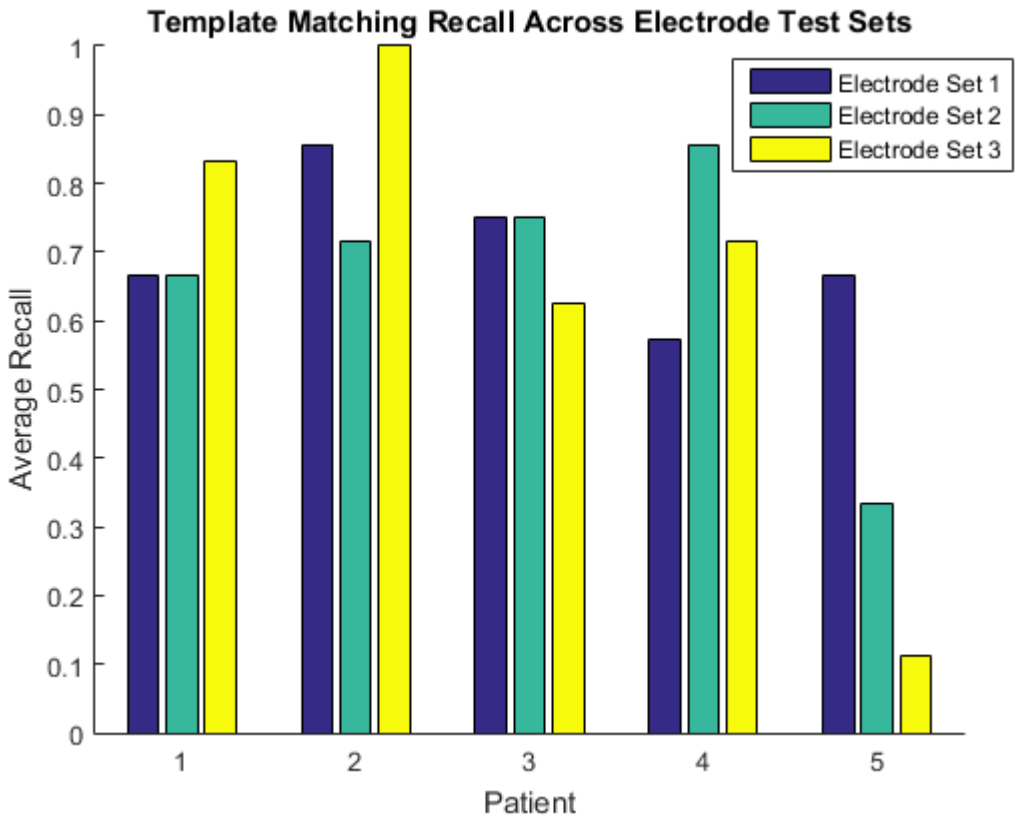


Fig. 24.: Template matching recall averages across all three tests for all five patients.

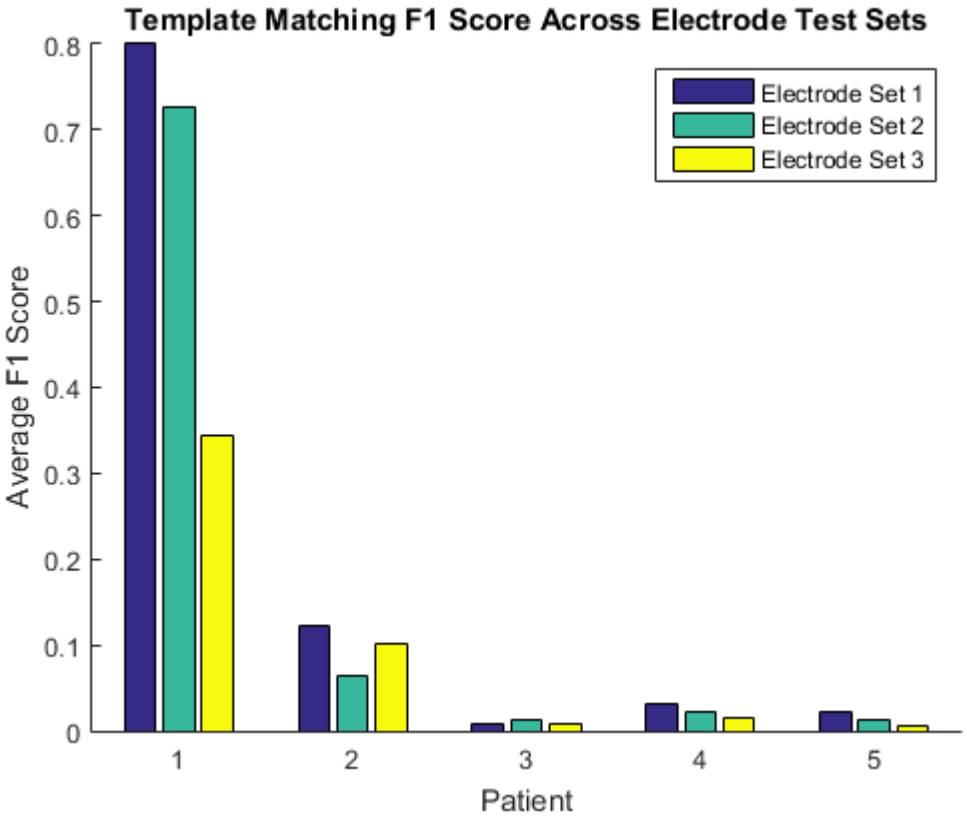


Fig. 25.: Template matching F1 score averages across all three tests for all five patients.

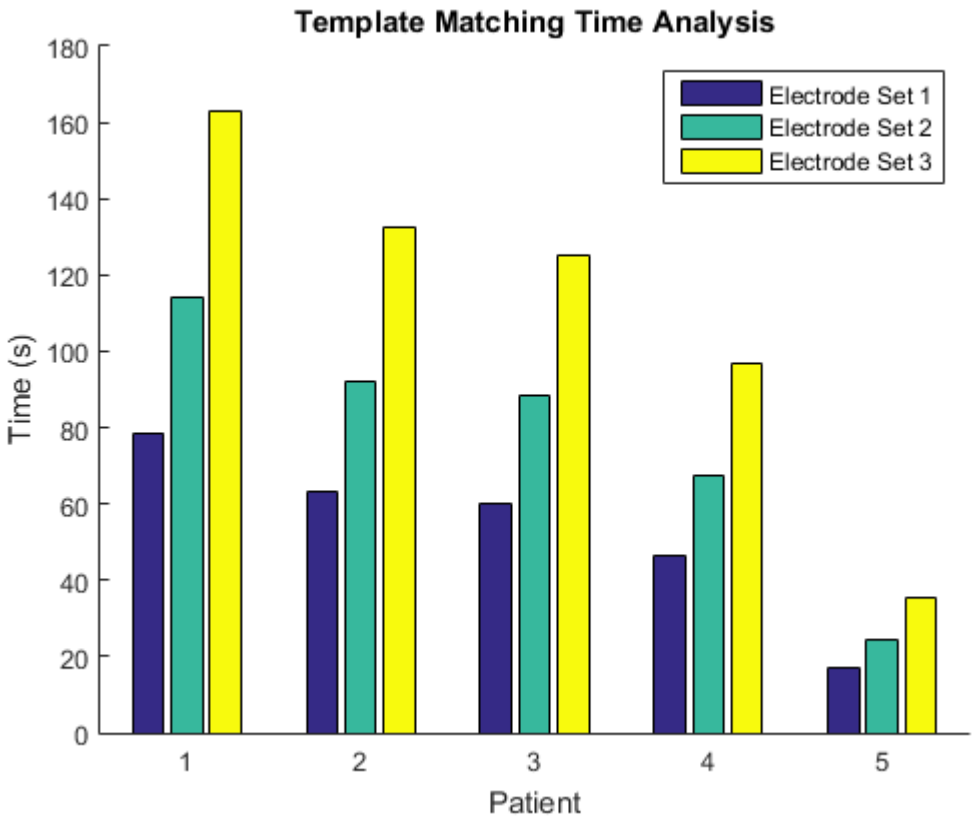


Fig. 26.: Template matching times across all three tests for all five patients.

Table 9.: Template matching extrapolated true positives, false positives, and dataset reduction percentage for all patients of the first electrode test set. Assuming 20 million total EEG reading dataset.

Electrode Set 1				
Patient	True Positives	False Positives	Dataset Reduction %	Spike Data %
1	360.83	0.00	99.78%	100.00%
2	669.25	9481.05	93.90%	6.59%
3	698.49	153085.87	7.73%	0.45%
4	1185.83	72038.95	56.07%	1.62%
5	2434.97	203319.67	0.00%	1.18%

Table 10.: Template matching extrapolated true positives, false positives, and dataset reduction percentage for all patients of the second electrode test set. Assuming 20 million total EEG reading dataset.

Electrode Set 2				
Patient	True Positives	False Positives	Dataset Reduction %	Spike Data %
1	360.83	90.21	99.73%	80.00%
2	557.71	16285.10	89.89%	3.31%
3	698.49	113621.15	31.41%	0.61%
4	1778.74	162754.67	1.28%	1.08%
5	1217.48	188709.87	0%	0.64%

Table 11.: Template matching extrapolated true positives, false positives, and dataset reduction percentage for all patients of the third electrode test set. Assuming 20 million total EEG reading dataset.

Electrode Set 3				
Patient	True Positives	False Positives	Dataset Reduction %	Spike Data %
1	451.04	1623.74	98.76%	21.74%
2	780.79	13831.18	91.23%	5.34%
3	582.08	136671.34	17.65%	0.42%
4	1482.28	180986.76	0.00%	0.81%
5	405.83	143257.17	13.80%	0.28%

3.3.5.2 Multilayer Perceptron Results

The interictal spike is ultimately a multidimensional phenomenon. While the single electrode showed that the neural network can be used to detect the spikes in the patients successfully, to truly locate the point source of the spikes, multiple electrodes must be examined. Figure 27 shows the average precision across the 5 folds used for testing, giving a useful indication of how well the MLP can distinguish between true and false positives. Figure 28 shows the average recall across the 5 folds used for testing, giving a good indication of how well the MLP can distinguish between true positives and false negatives. Figure 29 shows the F1 score averages across all the 5 folds for all 5 patients, giving a good general score of accuracy for the system.

The three electrode set sizes are shown grouped together for each patient on the corresponding graphs. The reason for testing increasing sizes of electrode sets was to determine whether adding more electrodes would increase the accuracy of the spike detection. By analyzing the precision, recall, and F1 scores, it is seen that increasing the number of electrodes, in general, decreased performance of MLP. This is likely due to MLP not being able to extract any extra useful information out of the new electrodes. The increased complexity and noise introduced by the new electrodes outweighed any extra useful information contained in the electrodes.

The single electrode tested only on 5,000 samples, so the results were based on a much smaller testing set. The multi-electrode analysis tested on much larger regions of data due to the cross validation approach instead of random sampling. Table 12 shows the total training/testing size per fold for each patient.

The confusion matrix averages across all 5 folds for each patient of the MLP multi-electrode approach are found in Tables 26-28 in the appendix. Using Equations

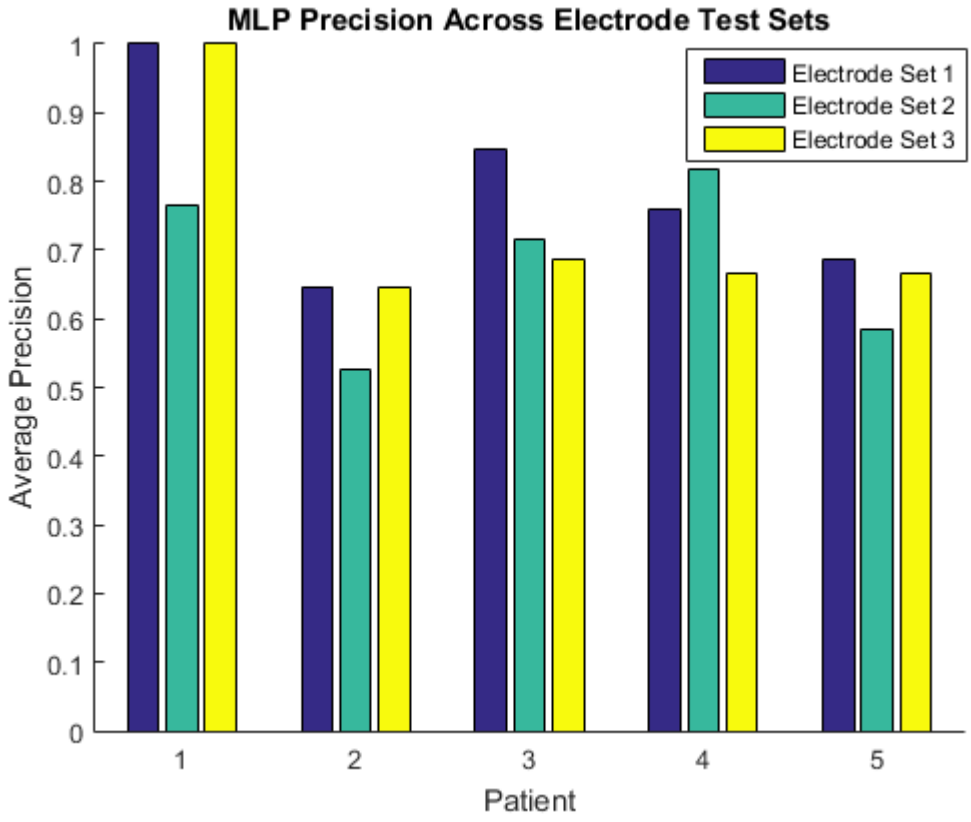


Fig. 27.: MLP precision averages across all three tests for all five patients.

Table 12.: Shows the size of the training and testing regions for each of the five folds across the five patients.

Patient	Total Samples	Training Size	Testing Size
1	227,710	177,368	44,342
2	179,305	143,444	35,861
3	171,799	137,436	34,359
4	134,927	107,940	26,985
5	49,282	39,424	9,856

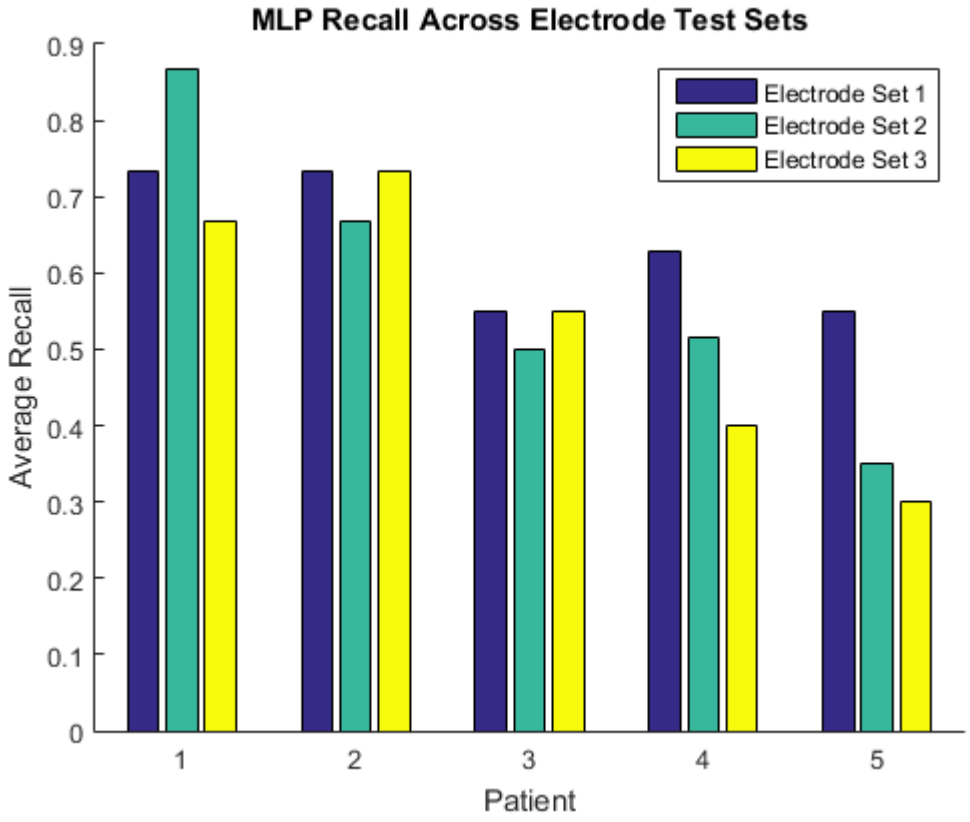


Fig. 28.: MLP recall averages across all three tests for all five patients.

3.1-3.6 along with the confusion matrix averages, the extrapolated data reduction was calculated and shown in Tables 13-15.

All three electrode test set sizes showed at least than a 98% reduction in the size of the dataset. The final column, the spike data percentage, is simply the number of true positives divided by the number of false positives. This gives an indication of the ratio of spikes to nonspikes that a doctor would encounter in a given dataset. As a final note, the total number of samples that would be extracted by the system is the number of true positives plus false positives. For all three electrode test sets, this number is no greater than 3,000. This means a doctor would only have to manually analyze 3,000 windows that contain potential spikes, instead of a continuous stream

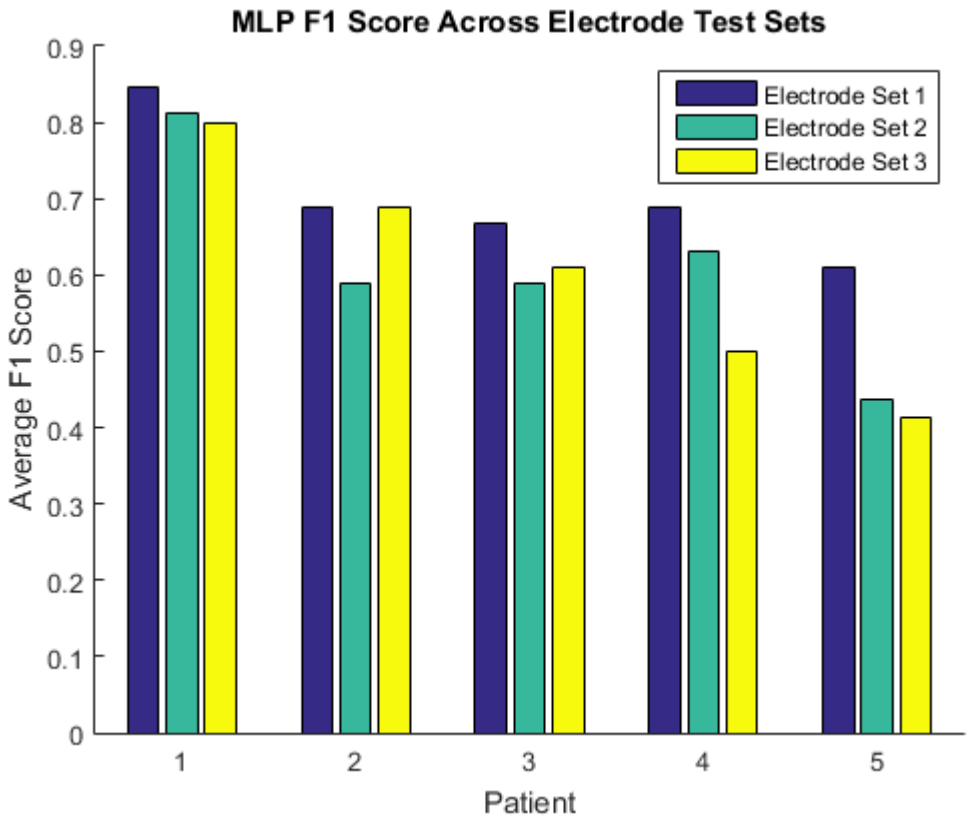


Fig. 29.: MLP F1 scoreaverages across all three tests for all five patients.

of 20 million, or more, samples. With respect to the goal of this problem, the MLP was able to achieve substantial data reduction, while maintaining sufficient accuracy.

Reducing the size of the dataset makes the doctor's manual identification job easier, however the training time of a neural network must be taken into consideration. Figure 30 shows the amount of time, in seconds, that each network took to train for each electrode set. Patient 1 had the most amount of data, as shown in Table 12, with each following patient having smaller datasets. The longest training time was for electrode set 3 for Patient 1 at 2362.8s, or just under 40 minutes. This time is still far under the week and a half the doctors currently use to manually analyze a single patient. As computer hardware increases in efficiency and speed, the training

times will decrease. Table 6, from Section 3.3.3, shows the hardware used to train the neural network. The scalability of the MLP system is more difficult than that of a CPU bound template matching system from section 3.3.2. Increasing the power of the GPU is the most straightforward way to increase performance, while splitting the training into multiple GPUs simultaneously is a much more difficult and time consuming task.

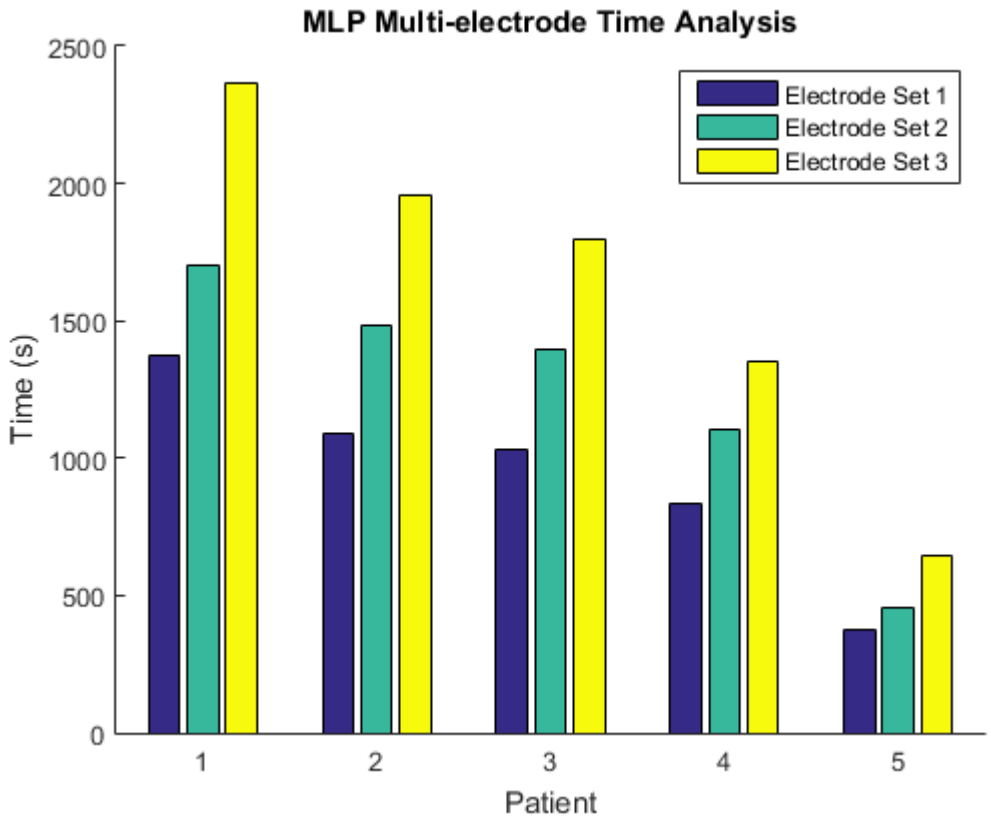


Fig. 30.: Shows the time it took to train each of the neural networks.

Table 13.: MLP extrapolated true positives, false positives, and dataset reduction percentage for all patients of the first electrode test set. Assuming 20 million total EEG reading dataset.

Electrode Set 1				
Patient	True Positives	False Positives	Dataset Reduction %	Spike Data %
1	270.62	541.25	99.51%	33.33
2	334.63	3234.71	97.86%	9.38%
3	465.66	2095.47	98.46%	18.18%
4	978.31	3557.48	97.28%	21.57%
5	1298.65	4869.93	96.30%	21.05%

Table 14.: MLP extrapolated true positives, false positives, and dataset reduction percentage for all patients of the second electrode test set. Assuming 20 million total EEG reading dataset.

Electrode Set 2				
Patient	True Positives	False Positives	Dataset Reduction %	Spike Data %
1	270.62	721.66	99.40%	27.27%
2	334.62	2230.84	98.46%	13.04%
3	442.38	1513.40	98.83%	22.62%
4	948.66	2519.88	97.92%	27.35%
5	649.32	6899.07	95.47%	8.60%

Table 15.: MLP extrapolated true positives, false positives, and dataset reduction percentage for all patients of the third electrode test set. Assuming 20 million total EEG reading dataset.

Electrode Set 3				
Patient	True Positives	False Positives	Dataset Reduction %	Spike Data %
1	270.62	902.08	99.30%	23.08%
2	290.01	1784.67	98.76%	13.98%
3	372.53	1513.40	98.87%	19.75%
4	622.56	2371.65	98.20%	20.79%
5	892.825	2840.79	97.76%	23.91%

3.3.5.3 Convolutional Network Results

Analysis of the CNN's performance follows the same as the previous two methods. The false positives and true positives across the same electrode sets as defined in Table 35, Appendix B, are analyzed below.

The CNN is able to analyze smaller windows of information, learning about subsections of the input image. This gives it the ability to learn about different features of the spike, and match these more refined details between data points. However, the results of the CNN are significantly lower than that of the MLP in the previous section. Examining the results from Tables 16-18 show the CNN is unable to reduce the dataset by any significant amount in almost all cases, in contrast to the MLP results in Tables 13-15. In the tests where it managed to make a noticeable reduction in data, the accuracy of spike detection was so low that there was essentially no useful data extracted. The CNN was unable to accurately differentiate between spike data and nonspike data.

CNNs are very difficult to optimize in terms of architecture design. Adding more layers significantly increases the number of parameters to optimize (in terms of weights for each neuron/filter). During testing of the CNN, the training accuracy would routinely reach 100% accuracy, however testing data would always result in every sample being classified as either a spike or a nonspike. This means the network was overfitting to the training data, simply memorizing the spike samples by adjusting the weights to perfectly classify the spikes during training. This had the effect of not generalizing to the remaining spikes in the data.

The way this is normally countered is with a validation set during training. While training, the training data is tested to see how well the optimization is performing. The validation set is extra data that is not used during training, but used to test

how well the architecture generalizes. The validation set is tested at regular intervals during training, and if the accuracy of the validation set starts to decrease, that means the system has started to overfit to the training data. When this occurs, training is usually stopped and the model is considered as trained as it can be for that specific configuration.

With the problem at hand, there is currently not enough data to create a validation set for training. Patient 1 had 6 total spikes, leaving 3 for training and 3 for testing. If a validation set were created, that would leave only 1 or 2 for training/validation. This is not enough data to make any reliable claims for generalization if only 1 sample is considered in the validation or training set. Validation was handled manually for the CNN, by testing various lengths of training and architectures. The presented architecture was one of the better performing networks. Future work will include expanding the number of spikes per patient to decrease the effect of this problem. With respect to the goal of this section, the CNN architecture was not able to achieve accurate data reduction.

Table 16.: CNN extrapolated true positives, false positives, and dataset reduction percentage for all patients of the first electrode test set. Assuming 20 million total EEG reading dataset.

Electrode Set 1				
Patient	True Positives	False Positives	Dataset Reduction %	Spike Data %
1	18.04	124757.57	25.13%	0.01%
2	133.85	1835866.26	0.00%	0.01%
3	186.26	2592448.15	0.00%	0.01%
4	296.46	2652842.91	0.00%	0.01%
5	649.32	1902520.19	0.00%	0.03%

Table 17.: CNN extrapolated true positives, false positives, and dataset reduction percentage for all patients of the first electrode test set. Assuming 20 million total EEG reading dataset.

Electrode Set 2				
Patient	True Positives	False Positives	Dataset Reduction %	Spike Data %
1	90.21	165260.9	0.79%	0.05%
2	89.23	2564792.84	0.00%	0.00%
3	209.55	3357179.03	0.00%	0.01%
4	88.94	1911700.40	0.00%	0.00%
5	649.32	1319345.81	0.00%	0.05%

Table 18.: CNN extrapolated true positives, false positives, and dataset reduction percentage for all patients of the first electrode test set. Assuming 20 million total EEG reading dataset.

Electrode Set 3				
Patient	True Positives	False Positives	Dataset Reduction %	Spike Data %
1	18.04	85968.16	48.41%	0.02%
2	111.54	2844092.47	0.00%	0.00%
3	325.96	2921088.02	0.00%	0.01%
4	59.29	1087995.73	0.00%	0.01%
5	405.83	764579.36	0.00%	0.05%

CHAPTER 4

MULTIPLE PATIENT ANALYSIS

4.1 Why Examine Multiple Patients?

The interictal spikes of a patient are inherently unique to the patient. However, the interictal spike exhibits similar characteristics across multiple patients. The large amplitude, short duration waveform is a common feature across patients. Upon analysis of the five patients in this study, it was confirmed that there is strong similarity in the shape and duration of the interictal spikes. Figure 31 shows five centered spikes from the five patients in this study. There is a clear pattern in the shape, with variation of the amplitude. This section serves as an initial look at generalizing a data reduction system by being able to identify spikes from one patient by training a neural network on spikes from separate patients.

For multiple patient analysis, analyzing a single electrode does not make much sense. Each patient will more than likely have their interictal spikes originating from a different region of the brain, and therefore have different electrodes containing the spikes. A system that can work with multiple patients must be able to identify the spike pattern independent from the electrode. To put more simply, training a system to recognize spikes from patients that only had their seizures near electrode 47 would make it difficult or impossible to recognize spikes from electrode 213. The electrodes analyzed follow the same set organization from the single patient analysis in the previous chapter. See Table 35 in Appendix B for the electrode numbers that correspond to the three sets tested.

Being able to generalize the spike detection problem could lead to a number of

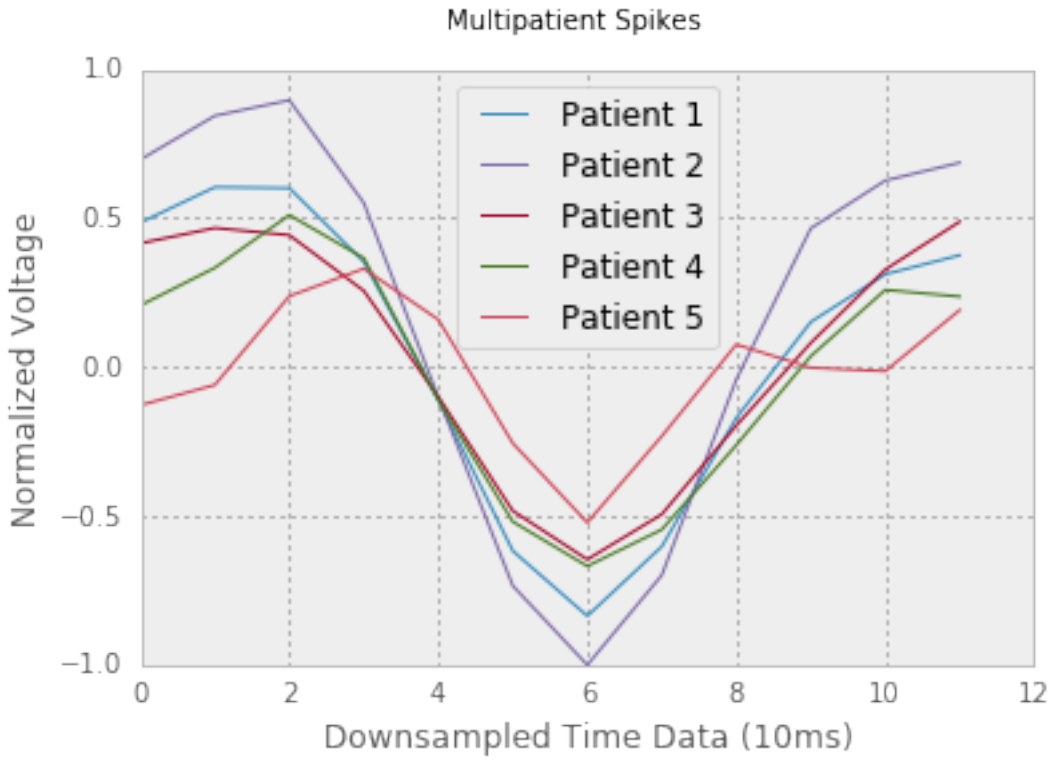


Fig. 31.: Spikes from all five patients plotted together. They exhibit very similar characteristics, despite being from different people.

benefits. The most obvious is more data to train a network with. If a doctor did not have to manually analyze a few spikes for each patient, as required by the system in the single patient analysis, it could lead to drastically decreased analysis time per patient as the system could simply use the pre-trained network from previous patients on a new patient. This chapter examines whether four of the five patients can be used to train a network that can accurately identify spikes in the fifth patient. Each patient is tested, with the other four being trained on.

4.2 Multilayer Perceptron

The multiple patient analysis followed the same methods for raw data handling and signal processing as described in Sections 3.1.1 and 3.1.2 respectively. This approach uses the same multiple electrode techniques discussed in Section 3.3 for concatenating electrode windows together. The neural network comprised of a single hidden layer with 100 neurons and a single output layer using the sigmoid activation function for classification and a learning rate of 0.001.

Figure 32 shows the framework of the data extraction process used for the multiple patient analysis. Four of the five patients are used to train a network, while the fifth is used for testing. The spikes and nonspikes from each patient used for training are all pooled for batches similar to the multiple electrode analysis in section 3.3.3. There is a 50% chance for the batch to select a spike and 50% to pick a nonspike from the pooled data with equal chance to pool from each patient.

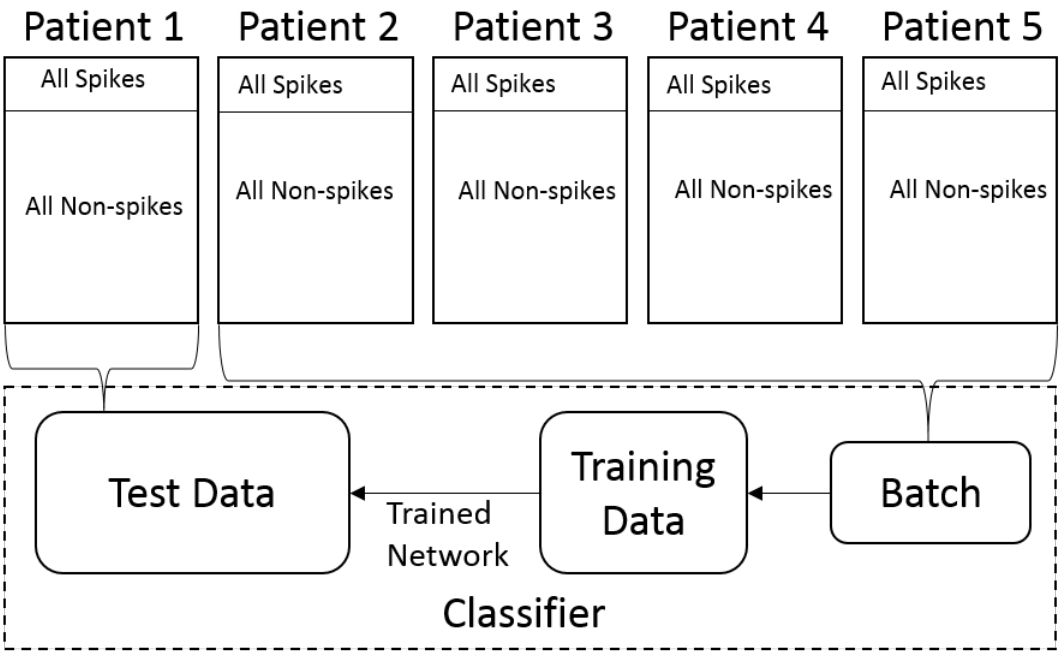


Fig. 32.: Framework showing method used to train and test multiple patient analysis.

4.3 Results

The results for the multi-patient experiment were subject to the same analysis as the single patient analysis. Figures 33-35 show the precision, recall, and F1 scores for all five patients across the tested electrode sets. Again, similar to the single patient analysis, the addition of extra electrodes seems to reduce the overall effectiveness of the system.

Patients 1-3 perform significantly better than patients 4 and 5. This is likely due to the spatial location of the interictal spikes within the brain. The spike activity for Patients 1-3 occurs near the same electrodes, while for Patients 4 and 5 it occurs near electrodes on the opposite side of the brain. When testing a network on Patient 4, it was trained on data from Patients 1-3 and 5. So, the spatial relation for spikes on Patients 1-3 will not be relevant for Patient 4. This is a severe limitation for

the accuracy of an approach with a standard neural network. An advantage of the convolutional network is the capability to be spatially invariant in learning its features.

With the goal of data reduction in mind, the Tables 19-21 display the total data reduction amounts if the results on the testing data are extrapolated out to an EEG reading of 20 million data points. While the overall scoring metrics discussed in the previous figures may be low for some patients, the total amount of reduction achieved is still 99% for Patients 1-3 and the 94% for Patients 4 and 5. With Patient 1 performing the best, the total amount of samples the doctor would need to analyze is about 720, while patient 5 would require about 10,000 samples. Even with differing spatial relations of spikes between patients, there is still a substantial reduction in the amount of data needed to be analyzed, keeping enough accuracy to be useful. Exploration of methods to increase the accuracy of generalization across patients is an avenue to be explored in future work.

Table 19.: Multi-patient MLP extrapolated true positives, false positives, and dataset reduction percentage for all patients of the first electrode test set. Assuming 20 million total EEG reading dataset.

Electrode Set 1				
Patient	True Positives	False Positives	Dataset Reduction %	Spike Data %
1	541.25	54.12	99.64%	90.91%
2	736.18	423.86	99.30%	63.46%
3	651.92	745.06	99.16%	46.67%
4	1037.60	13459.13	91.30%	7.16%
5	1136.31	2678.46	97.71%	29.79%

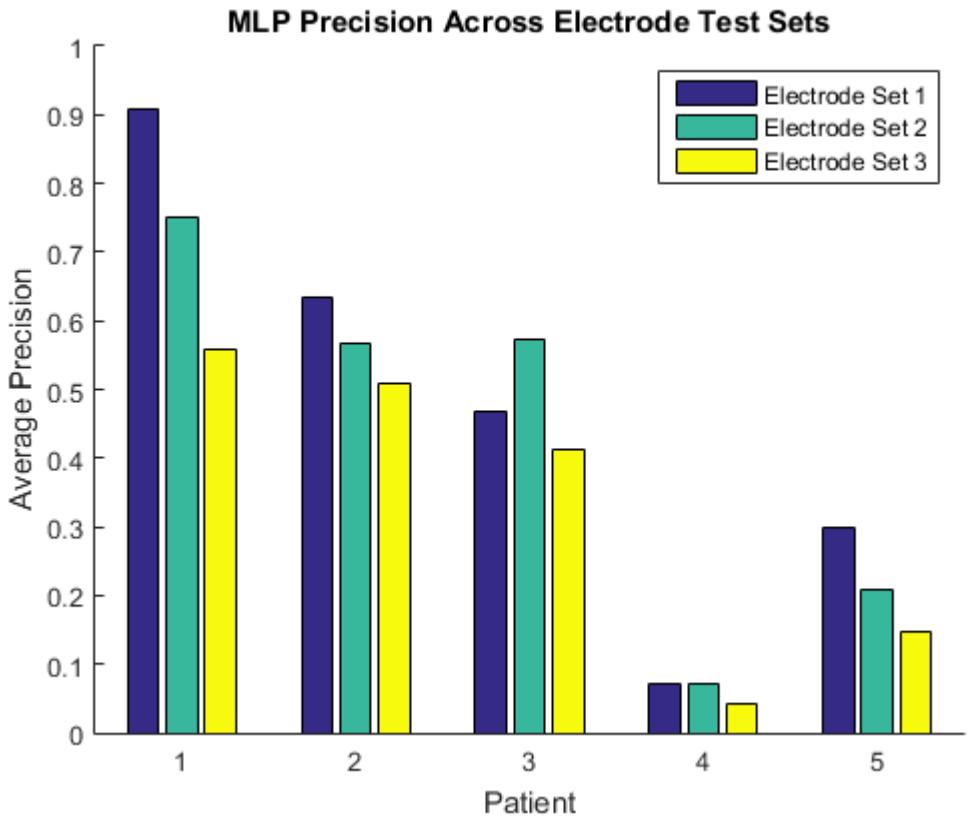


Fig. 33.: MLP precision averages across all three tests for all five patients.

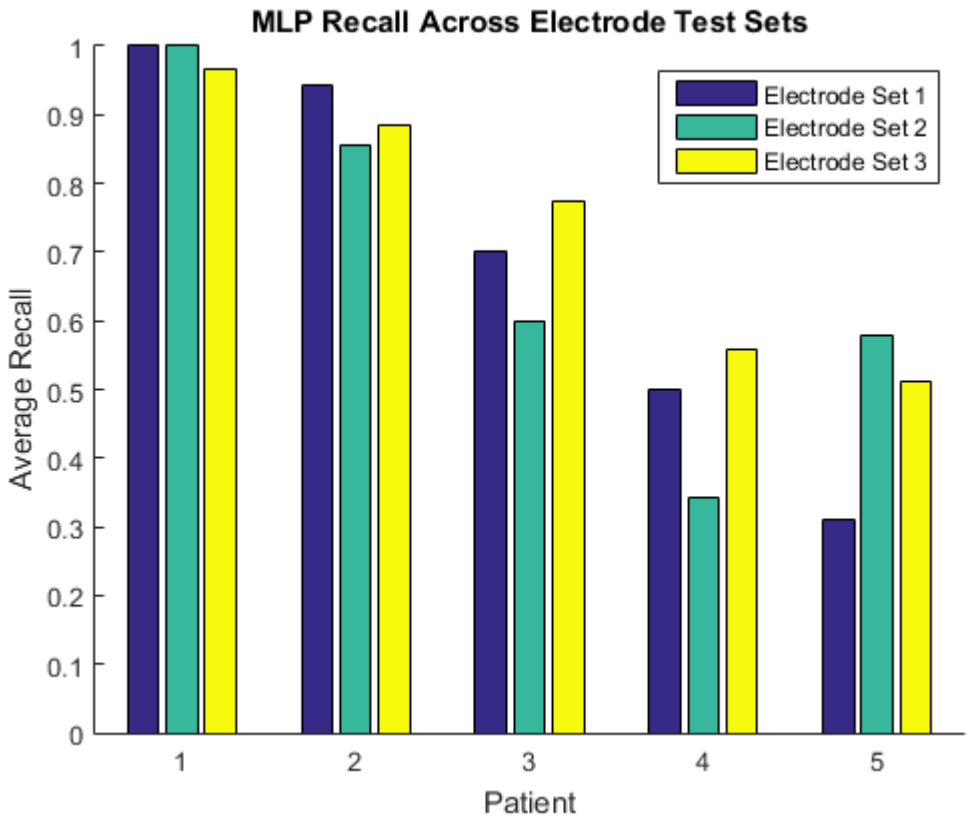


Fig. 34.: MLP recall averages across all three tests for all five patients.

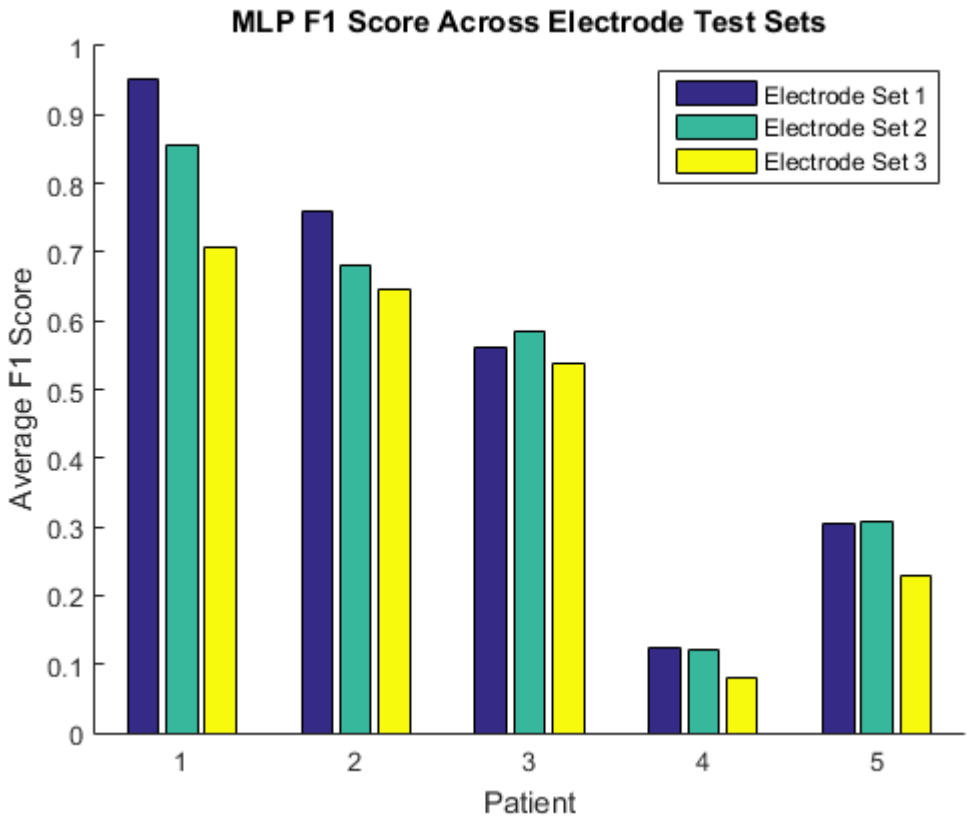


Fig. 35.: MLP F1 score averages across all three tests for all five patients.

Table 20.: Multi-patient MLP extrapolated true positives, false positives, and dataset reduction percentage for all patients of the second electrode test set. Assuming 20 million total EEG reading dataset.

Electrode Set 2				
Patient	True Positives	False Positives	Dataset Reduction %	Spike Data %
1	541.25	180.42	99.57%	75.00%
2	669.25	513.09	99.29%	56.60%
3	558.79	419.09	99.41%	57.14%
4	711.50	9101.22	94.11%	7.25%
5	2110.30	7954.22	93.96%	20.97%

Table 21.: Multi-patient MLP extrapolated true positives, false positives, and dataset reduction percentage for all patients of the third electrode test set. Assuming 20 million total EEG reading dataset.

Electrode Set 3				
Patient	True Positives	False Positives	Dataset Reduction %	Spike Data %
1	523.21	414.96	99.44%	55.77%
2	669.25	669.25	99.18%	50.82%
3	721.77	1024.45	98.95%	41.33%
4	1156.18	25317.39	84.12%	4.37%
5	1866.80	10713.85	92.45%	14.84%

CHAPTER 5

COMBINED RESULTS ANALYSIS

This chapter compares the effectiveness of all three methods together. The results of the CNN are not shown here, as its performance was negligible in comparison to the other methods. The results for electrode sets 1, 2, and 3 are shown in separate graphs for each metric: data reduction, F1 score, precision, recall, and time analysis.

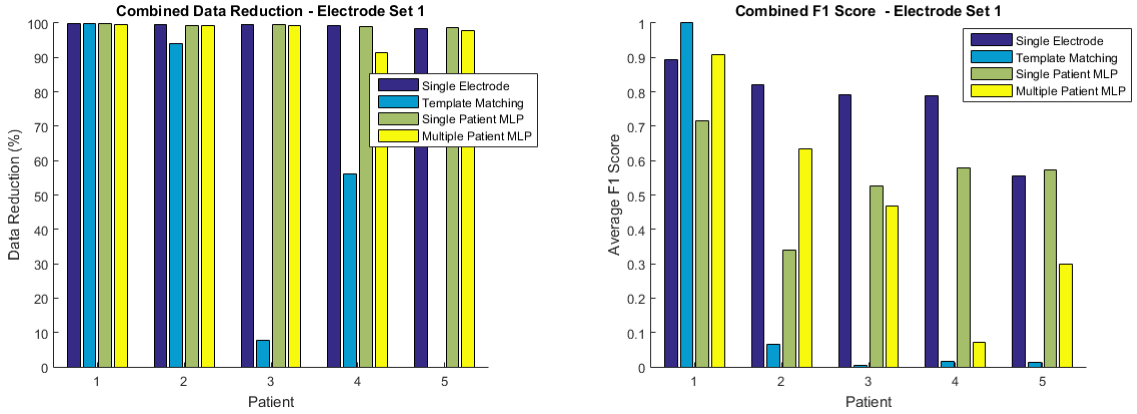
The template matching served as a baseline, requiring much less computational power than both the classical MLP and CNN. The template matching performed worse than the MLP on almost every instance, with Patient 1 on the first electrode set being the only exception. The template matching had massive reductions in data for Patient 2 across all electrode sets, however the accuracy of the data extracted was minimal, making the results much less useful.

The single electrode analysis was included in the combined results for comparison against the known region in the brain where the spikes were occurring. In a live scenario, a single electrode approach would not be useful, as the interictal spike phenomenon is inherently multi-dimensional in nature. While the single electrode performs better than the multiple electrode approach in most cases as far as accuracy of spike extraction goes, the data reduction values are very similar.

A common theme seen in the results is a reduction in the accuracy of spike classification with the addition of extra electrodes. This is likely due to the extra electrodes not carrying any further useful information. The electrodes added were spread equally around the brain, so electrodes that may be localized to the region where the spike is occurring would likely be outnumbered by electrodes further away

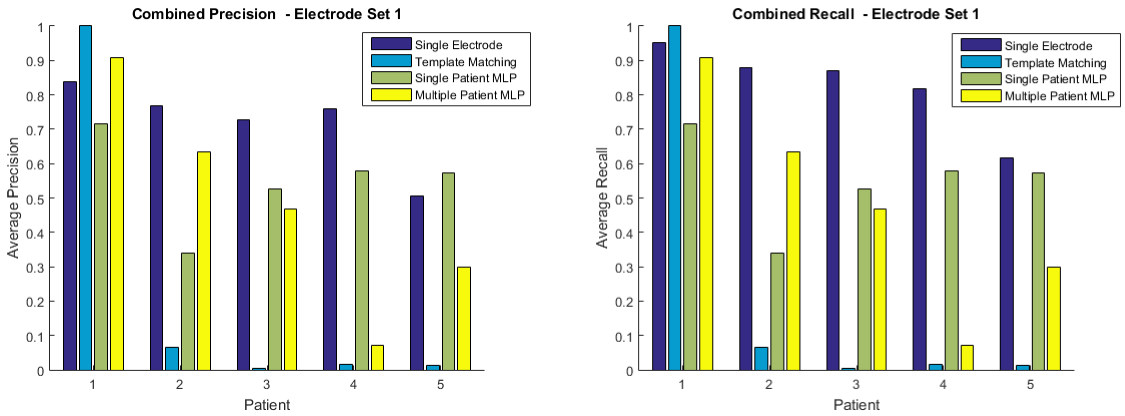
from the source. That could potentially be alleviated by running a first pass analysis to find which electrodes initially exhibit spike features, then running a separate analysis on multiple electrodes closer to those original electrodes. This could function as a sort of feature extraction and serve as a pre-processing step.

The time analysis shows the template matching system taking less time than the other algorithms. The accuracy and data reduction increases gained by using the learning algorithms outweighs the loss in time. Comparing to the original 10 days per patient that the neurologists took, the time multiple minutes for the algorithms tested could save days of valuable doctor time.



(a) (b)

Fig. 36.: Combined reduction percentage averages (a) and F1 score averages (b) for electrode set 1 across all patients.



(a) (b)

Fig. 37.: Combined precision averages (a) and recall averages (b) for electrode set 1 across all patients.

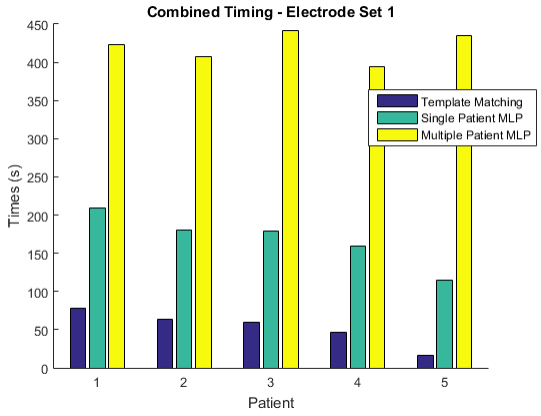
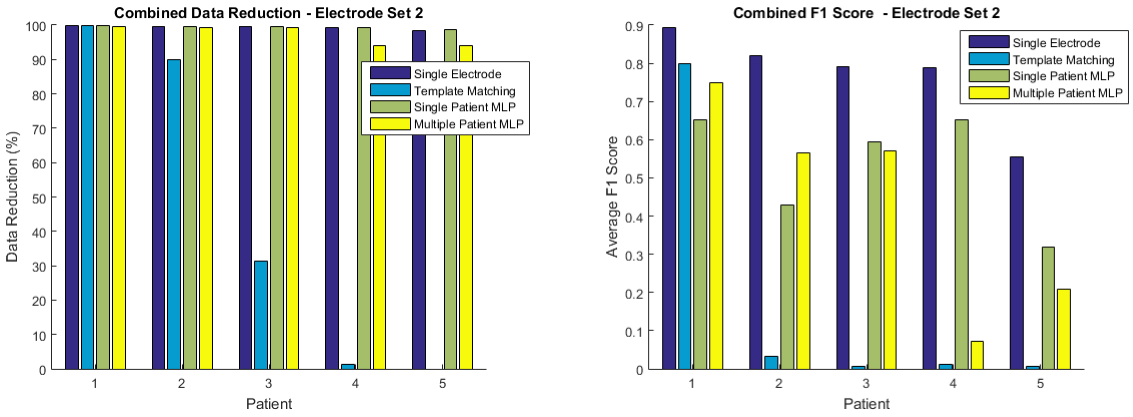
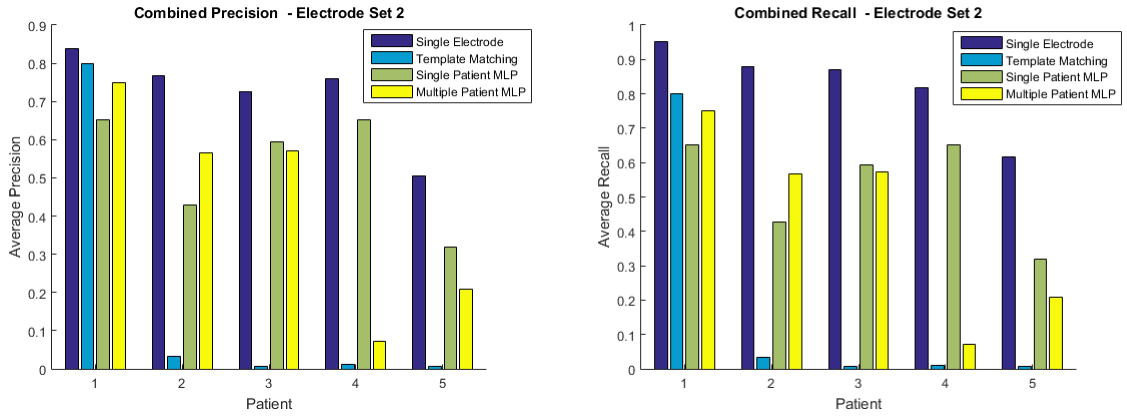


Fig. 38.: Combined time analysis for electrode set 1 across all patients.



(a) (b)

Fig. 39.: Combined reduction percentage averages (a) and F1 score averages (b) for electrode set 2 across all patients.



(a) (b)

Fig. 40.: Combined precision averages (a) and recall averages (b) for electrode set 2 across all patients.

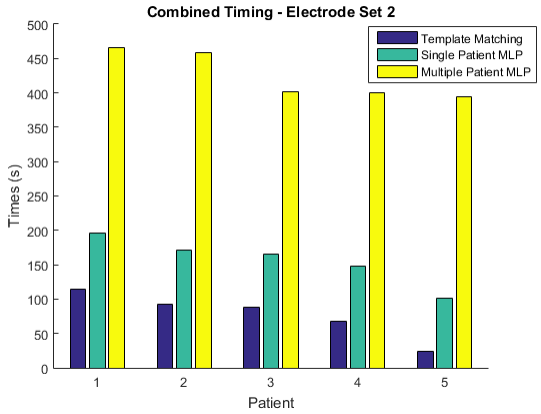
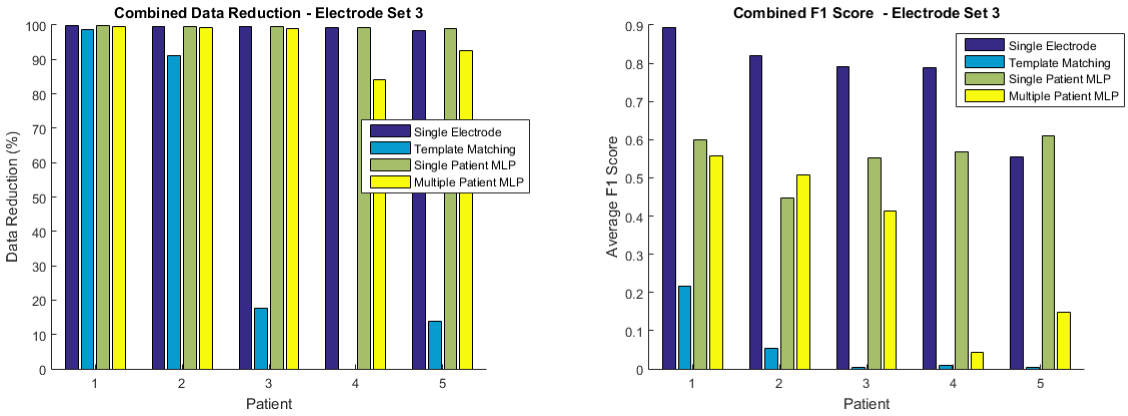
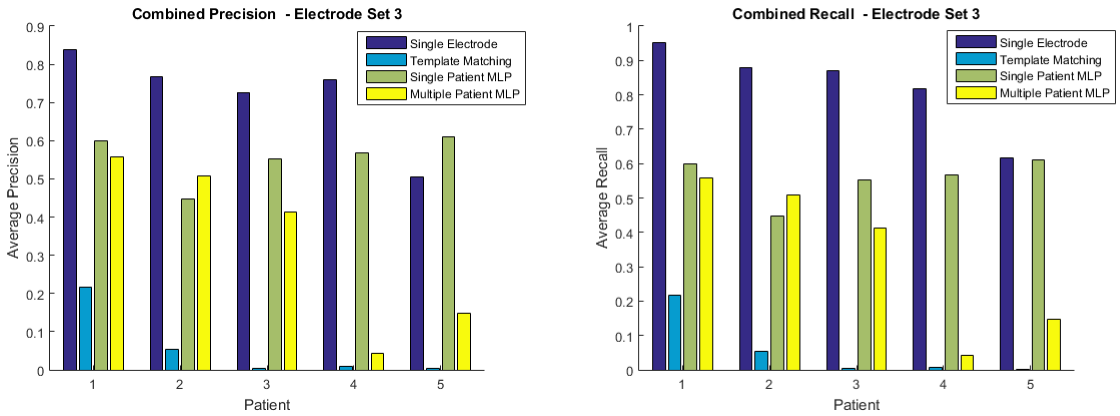


Fig. 41.: Combined time analysis for electrode set 2 across all patients.



(a) (b)

Fig. 42.: Combined reduction percentage averages (a) and F1 score averages (b) for electrode set 3 across all patients.



(a) (b)

Fig. 43.: Combined precision averages (a) and recall averages (b) for electrode set 3 across all patients.

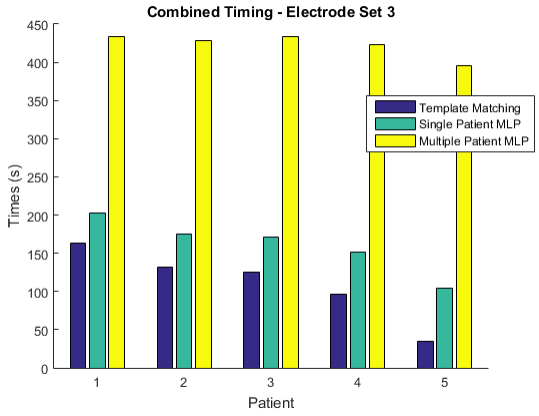


Fig. 44.: Combined time analysis for electrode set 3 across all patients.

CHAPTER 6

CONCLUSION AND FUTURE WORK

This thesis presented a framework for reducing the amount of data needing to be manually analyzed by a neurologist to identify interictal spikes within an epileptic patient's brains. It analyzed both classical artificial neural networks and the more recent convolutional neural networks. Three analyses were examined: single electrode, multiple electrode, and multiple patient training. The single electrode multilayer perceptron managed to reduce the total amount of data by a minimum of 91.98%, the multiple electrode analysis improved on this by achieving a no less than 98.25% reduction in data size. The template matching multiple electrode approach showed a 99.78% best case reduction for Patient 1, but performed significantly worse on the rest of the patients, showing 0% reduction for a few of the tests. The convolutional network achieved no noticeable data reduction in most cases. In tests where there was a measurable amount of reduction, the accuracy of spike vs. nonspike classification was low enough to make the system useless in a practical sense. The CNN results demonstrated the difficulty in generalizing multi-layer neural networks. The multiple patient analyses showed, at best, a 99.64% reduction and at worst 84.12% reduction in data.

All systems, with the exception of the convolutional neural networks and template matching in a few cases, were able to drastically reduce the size of data to be analyzed. Weighing the accuracy, data reduction, and time analysis together, the multiple electrode multilayer perceptron performed the best all around. It was able to generalize across patients, reducing data effectively, and keeping a necessary amount

of accuracy to allow the neurologist to complete the task of identifying spikes to locate the seizure epicenter within the brain.

The identification of interictal spikes presents numerous challenges for machine learning algorithms. First and foremost is the imbalanced dataset. With as little as 0.32% of the data being the positive classification value, pooled batching was used to provide more weight to the spikes during training. This comes at a potential cost of overfitting to the very few spike samples used for training. The overfitting problem was most seen during the convolutional neural networks. Obtaining more training data can help alleviate this problem in the future.

Another major challenge that was encountered was the quality of the annotated spikes. The doctor manually identified the spike locations for this study, and there is a chance that the region used for training nonspikes included actual spikes that the doctor missed during the analysis. With there being no guarantee that the dataset was perfectly clean, it is entirely possible that the learning algorithms were shown an actual spike during training but taught it was not a spike. This contamination of the training set has the potential to simply break the learning system. While it is unclear whether this occurred during this study or not, it is important to note the possibility.

The results from this thesis show that classical multilayer perceptrons perform better than simple template matching in most cases, and can be used to help neurologists identify interictal spikes in patients by massively reducing the amount of data needed to be manually analyzed. The multiple patient analysis shows promising potential, and should be explored more in depth going forward.

Addressing the contributions from Section 1.2:

1. Achieved reduction in amount of data for a neurologist to analyze.

2. Maintained sufficient accuracy in spike detection.

3. Identified a loss in accuracy when adding electrodes to analysis.

4. Achieved accurate data reduction while training on spikes from multiple patients, generalizing the problem.

Possible future work includes:

1. Use learning algorithms, or other methods, on the remaining data to further reduce data set.

2. Utilize a first-pass search using a spread out electrode mapping, followed by a more targeted analysis.

3. Investigate feature analysis techniques to reduce the input space to a more manageable size, allowing for more electrodes to be analyzed.

4. Further test the CNNs to find an optimal architecture.

5. Increase the amount of data for analysis. Both in terms of more patients, and more data from each patient.

6. Expand the testing to include entire EEG datasets.

7. Further clean data used for training.

Appendix A

ABBREVIATIONS

EEG	Electroencephalogram or Electroencephalography
ANN	Artificial Neural Network
CNN	Convolutional Neural Network
MLP	Multilayer Perceptron
ReLU	Rectified Linear Unit
VCU	Virginia Commonwealth University
MCV	VCU Health (Formerly Medical College of Virginia)

Appendix B

FIGURES AND TABLES

Table 22.: Confusion matrix averages for single electrode analysis in section 3.2

Patient	True Pos.	False Pos.	True Neg.	False Neg.
1	2.86	0.54	4988.27	0.14
2	3.52	1.07	4969.62	0.48
3	3.48	1.31	4970.58	0.52
4	5.73	1.81	4975.59	1.27
5	3.09	3.03	4961.73	1.91

.

Table 23.: Confusion matrix for template matching multi-electrode analysis of electrode set 1 in section 3.3.2

Patient	True Pos.	False Pos.	True Neg.	False Neg.
1	4.00	0.00	221710.00	2.00
2	6.00	85.00	179219.00	1.00
3	6.00	1315.00	170484.00	2.00
4	8.00	486.00	134441.00	6.00
5	6.00	501.00	48781.00	3.00

.

Table 24.: Confusion matrix for template matching multi-electrode analysis of electrode set 2 in section 3.3.2

Patient	True Pos.	False Pos.	True Neg.	False Neg.
1	4.00	1.00	221709.00	2.00
2	5.00	146.00	17918.00	2.00
3	6.00	976.00	170823.00	2.00
4	12.00	1098.00	133829.00	2.00
5	3.00	465.00	48817.00	6.00

.

Table 25.: Confusion matrix for template matching multi-electrode analysis of electrode set 3 in section 3.3.2

Patient	True Pos.	False Pos.	True Neg.	False Neg.
1	5.00	18.00	221692.00	1.00
2	7.00	124.00	179180.00	0.00
3	5.00	1174.00	170625.00	3.00
4	10.00	1221.00	133706.00	4.00
5	1.00	353.00	48929.00	8.00

.

Table 26.: Confusion matrix averages for MLP multi-electrode analysis of electrode set 1 in section 3.3.3

Patient	True Pos.	False Pos.	True Neg.	False Neg.
1	3.0	1.2	44140.6	0.0
2	3.0	5.8	35574.8	0.0
3	4.0	3.6	34067.4	0.0
4	6.6	4.8	26544.4	0.4
5	3.2	2.4	9587.8	0.8

.

Table 27.: Confusion matrix averages for MLP multi-electrode analysis of electrode set 2 in section 3.3.3

Patient	True Pos.	False Pos.	True Neg.	False Neg.
1	3.0	1.6	44144.2	030
2	3.0	4.0	35596.8	0.0
3	3.8	2.6	34070.8	0.2
4	6.4	3.4	26558.8	0.6
5	1.6	3.4	9578	2.4

.

Table 28.: Confusion matrix averages for MLP multi-electrode analysis of electrode set 3 in section 3.3.3

Patient	True Pos.	False Pos.	True Neg.	False Neg.
1	3.0	2.0	44127.0	0.0
2	2.6	3.2	35615.0	0.4
3	3.2	2.6	34073.0	0.8
4	4.2	3.2	26568.8	2.8
5	2.2	1.4	9597.4	1.8

.

Table 29.: Confusion matrix averages for CNN multi-electrode analysis of electrode set 1 in section 3.3.4

Patient	True Pos.	False Pos.	True Neg.	False Neg.
1	0.2	276.6	39838	2.8
2	1.2	3291.8	25789.2	1.8
3	1.6	4453.8	25945.8	2.4
4	2.0	3579.4	19949.6	5
5	1.6	937.6	7409.8	2.4

.

Table 30.: Confusion matrix averages for CNN multi-electrode analysis of electrode set 2 in section 3.3.4

Patient	True Pos.	False Pos.	True Neg.	False Neg.
1	1.0	366.4	30550.8	2.0
2	0.8	4598.8	22752.8	2.2
3	1.8	5767.6	23066.6	2.2
4	0.6	2579.4	22022.2	6.4
5	1.6	650.2	8501.8	2.4

.

Table 31.: Confusion matrix averages for CNN multi-electrode analysis of electrode set 3 in section 3.3.4

Patient	True Pos.	False Pos.	True Neg.	False Neg.
1	0.2	190.6	41697.2	2.8
2	1.0	5099.6	26412.6	2.0
3	2.8	5018.4	22654.4	1.2
4	0.4	1468.0	23847.4	6.6
5	1.0	376.8	9058.2	3.0

.

Table 32.: Confusion matrix averages for MLP multi-patient analysis of electrode set 1 in section 4.2

Patient	True Pos.	False Pos.	True Neg.	False Neg.
1	6	0.600	220860.60	0.00
2	6.60	3.80	178290.00	0.40
3	5.60	6.40	170612.20	2.40
4	7	90.80	132037.40	7.00
5	2.80	6.60	47973.40	6.20

.

Table 33.: Confusion matrix averages for MLP multi-patient analysis of electrode set 2 in section 4.2

Patient	True Pos.	False Pos.	True Neg.	False Neg.
1	6.00	2	220850.20	0.00
2	6.00	4.60	178285.60	1.00
3	4.80	3.60	170666.80	3.20
4	4.80	61.40	132501.20	9.20
5	5.20	19.60	47891.00	3.80

.

Table 34.: Confusion matrix averages for MLP multi-patient analysis of electrode set 3 in section 4.2

Patient	True Pos.	False Pos.	True Neg.	False Neg.
1	5.80	4.60	220843.20	0.20
2	6.20	6.00	178264.80	0.80
3	6.20	8.80	170603.00	1.80
4	7.80	170.80	131268.20	6.20
5	4.60	26.40	47855.80	4.40

.

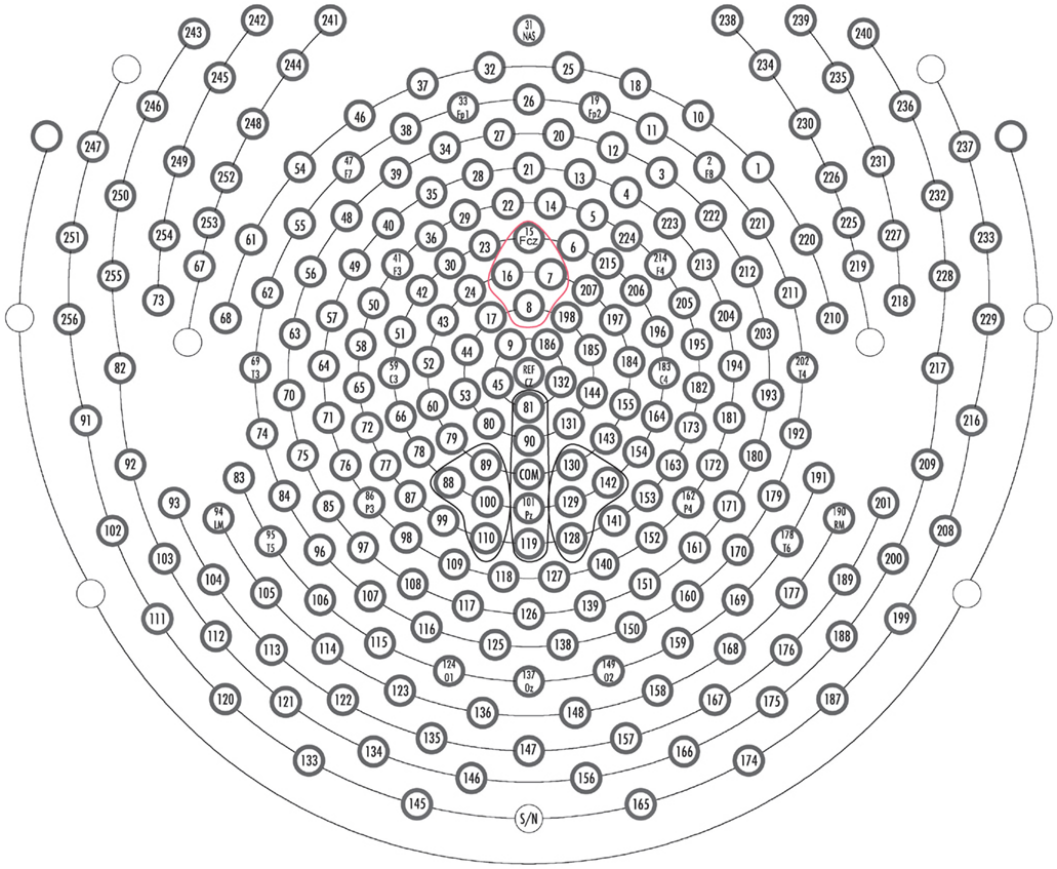


Fig. 45.: Electrode numbering for the EGI Dense Array.

Table 35.: Electrodes tested in multi-electrode analysis.

Test Number	Electrode Numbers
1	1,15,33,19,47,41,214,59,183,69,202,86,101,162,95,178,94, 190,124,137,149
2	1,15,33,19,47,41,214,59,183,69,202,86,101,162,95,178,94, 190,124,137,149,250,61,21,220,232,44,8,185,80,131,103, 107,169,200,121,175,145,165
3	1,15,33,19,47,41,214,59,183,69,202,86,101,162,95,178,94, 190,124,137,149,250,61,21,220,232,44,8,185,80,131,103, 107,169,200,121,175,145,165,245,31,235,252,35,4,226,73, 56,212,218,9,186,71,81,181,91,216,88,142,111,117,139,199, 147

REFERENCES

[1] Nurettin Acir, Ibrahim Oztura, Mehmet Kuntalp, Baris Baklan, and Cuneyt Guzelis. "Automatic detection of epileptiform events in EEG by a three-stage procedure based on artificial neural networks". In: *IEEE Transactions on Biomedical Engineering* 52.1 (2005), pp. 30–40.

[2] Sharanreddy Mallikarjun Akareddy and PK Kulkarni. "EEG signal classification for epilepsy seizure detection using improved approximate entropy". In: *International Journal of Public Health Science (IJPHS)* 2.1 (2013), pp. 23–32.

[3] Roozbeh Aliabadi, Farshid Keynia, and Mehran Abdali. "Epilepsy Seizure Diagnosis in EEG by Artificial Neural Networks". In: *Majlesi Journal of Multimedia Processing* 2.2 (2013).

[4] KS Anusha, Mathew T Mathews, and Subha D Puthankattil. "Classification of Normal and Epileptic EEG Signal Using Time & Frequency Domain Features through Artificial Neural Network". In: *Advances in Computing and Communications (ICACC), 2012 International Conference on*. IEEE. 2012, pp. 98–101.

[5] Apple. *Performing Convolution Operations*. 2016. URL: `https://developer.apple.com/library/content/documentation/Performance/Conceptual/vImage/ConvolutionOperations/ConvolutionOperations.html` (visited on 10/02/2016).

[6] Eishi Asano, Otto Muzik, Aashit Shah, Csaba Juhász, Diane C Chugani, Sandeep Sood, James Janisse, Eser Lay Ergun, Judy Ahn-Ewing, Chenggang

Shen, et al. “Quantitative interictal subdural EEG analyses in children with neocortical epilepsy”. In: *Epilepsia* 44.3 (2003), pp. 425–434.

[7] Daniel T Barkmeier, Aashit K Shah, Danny Flanagan, Marie D Atkinson, Rajeev Agarwal, Darren R Fuerst, Kourosh Jafari-Khouzani, and Jeffrey A Loeb. “High inter-reviewer variability of spike detection on intracranial EEG addressed by an automated multi-channel algorithm”. In: *Clinical Neurophysiology* 123.6 (2012), pp. 1088–1095.

[8] Howard J Carey, Milos Manic, and Paul Arsenovic. “Epileptic Spike Detection with EEG using artificial Neural Networks”. In: *Human System Interactions (HSI), 2016 9th International Conference on*. IEEE. 2016, pp. 89–95.

[9] Sahbi Chaibi, Tarek Lajnef, Abdelbacet Ghrob, Mounir Samet, and Abdennaceur Kachouri. “A robustness comparison of two algorithms used for EEG spike detection”. In: *The open biomedical engineering journal* 9 (2015), p. 151.

[10] Clarifai. *How does Clarifai’s visual recognition API work?* 2016. URL: `https://www.clarifai.com/technology` (visited on 09/30/2016).

[11] EGI. *Dense Array EEG Neuroimaging*. 2016. URL: `https://www.egi.com/clinical-division/clinical-division-care-center/clinical-division-dense-array-neuroimaging` (visited on 08/10/2016).

[12] Kunihiko Fukushima. “Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position”. In: *Biological cybernetics* 36.4 (1980), pp. 193–202.

[13] AJ Gabor, RR Leach, and FU Dowla. “Automated seizure detection using a self-organizing neural network”. In: *Electroencephalography and clinical Neurophysiology* 99.3 (1996), pp. 257–266.

[14] Harish Kumar Garg and Amit Kumar Kohli. "EEG spike detection technique using output correlation method: A Kalman filtering approach". In: *Circuits, Systems, and Signal Processing* 34.8 (2015), pp. 2643–2665.

[15] Google. *TensorFlow*. 2016. URL: `https://www.tensorflow.org/` (visited on 09/10/2016).

[16] Jonathan J Halford. "Computerized epileptiform transient detection in the scalp electroencephalogram: Obstacles to progress and the example of computerized ECG interpretation". In: *Clinical Neurophysiology* 120.11 (2009), pp. 1909–1915.

[17] John E Hall. *Guyton and Hall textbook of medical physiology*. Elsevier Health Sciences, 2015.

[18] David H Hubel and Torsten N Wiesel. "Receptive fields and functional architecture of monkey striate cortex". In: *The Journal of physiology* 195.1 (1968), pp. 215–243.

[19] Christopher J James, Richard D Jones, Philip J Bones, and Grant J Carroll. "Detection of epileptiform discharges in the EEG by a hybrid system comprising mimetic, self-organized artificial neural network, and fuzzy logic stages". In: *Clinical Neurophysiology* 110.12 (1999), pp. 2049–2063.

[20] Tulga Kalayci and Ozcan Ozdamar. "Wavelet preprocessing for automated neural network detection of EEG spikes". In: *IEEE engineering in medicine and biology magazine* 14.2 (1995), pp. 160–166.

[21] Reza Khosrowabadi, Chai Quek, Kai Keng Ang, and Abdul Wahab. "ERNN: A biologically inspired feedforward neural network to discriminate emotion from

EEG signal". In: *IEEE transactions on neural networks and learning systems* 25.3 (2014), pp. 609–620.

[22] Cheng-Wen Ko, Yue-Der Lin, Hsiao-Wen Chung, and Gwo-Jen Jan. "An EEG spike detection algorithm using artificial neural network with multi-channel correlation". In: 4 (1998), pp. 2070–2073.

[23] R Kottaimalai, M Pallikonda Rajasekaran, V Selvam, and B Kannapiran. "EEG signal classification using principal component analysis with neural network in brain computer interface applications". In: *Emerging Trends in Computing, Communication and Nanotechnology (ICE-CCN), 2013 International Conference on*. IEEE. 2013, pp. 227–231.

[24] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems*. 2012, pp. 1097–1105.

[25] Yatindra Kumar, ML Dewal, and RS Anand. "Epileptic seizures detection in EEG using DWT-based ApEn and artificial neural network". In: *Signal, Image and Video Processing* 8.7 (2014), pp. 1323–1334.

[26] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324.

[27] Yung-Chun Liu, Chou-Ching K Lin, Jing-Jane Tsai, and Yung-Nien Sun. "Model-based spike detection of epileptic EEG data". In: *Sensors* 13.9 (2013), pp. 12536–12547.

[28] Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. "Rectifier nonlinearities improve neural network acoustic models". In: *Proc. ICML*. Vol. 30. 1. 2013.

[29] Eric D Marsh, Bradley Peltzer, Merritt W Brown III, Courtney Wusthoff, Phillip B Storm Jr, Brian Litt, and Brenda E Porter. "Interictal EEG spikes identify the region of electrographic seizure onset in some, but not all, pediatric epilepsy patients". In: *Epilepsia* 51.4 (2010), pp. 592–601.

[30] Ernst Niedermeyer and FH Lopes da Silva. *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 2005.

[31] Vivek Prakash Nigam and Daniel Graupe. "A neural-network-based detection of epilepsy". In: *Neurological Research* (2013).

[32] World Health Organization. *Epilepsy*. 2016. URL: `http://www.who.int/mediacentre/factsheets/fs999/en/` (visited on 10/02/2016).

[33] O Ozdamar and T Kalayci. "Detection of spikes with artificial neural networks using raw EEG". In: *Computers and Biomedical Research* 31.2 (1998), pp. 122–142.

[34] R Quian Quiroga, Zoltan Nadasdy, and Yoram Ben-Shaul. "Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering". In: *Neural computation* 16.8 (2004), pp. 1661–1687.

[35] Yuanfang Ren and Yan Wu. "Convolutional deep belief networks for feature extraction of EEG signal". In: *2014 International Joint Conference on Neural Networks (IJCNN)*. IEEE. 2014, pp. 2850–2853.

[36] Karen Simonyan and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition". In: *arXiv preprint arXiv:1409.1556* (2014).

[37] Abdulhamit Subasi and Ergun Ercelebi. "Classification of EEG signals using neural network and logistic regression". In: *Computer methods and programs in biomedicine* 78.2 (2005), pp. 87–99.

[38] Michal Teplan. "Fundamentals of EEG measurement". In: *Measurement science review* 2.2 (2002), pp. 1–11.

[39] Inc. The MathWorks. *MATLAB and Neural Network Toolbox*. Version 2015a. Sept. 2013. URL: `https://www.mathworks.com/`.

[40] Alexandros T Tzallas, Dimitrios G Tsalikakis, Evaggelos C Karvounis, Loukas Astrakas, Margaret Tzaphlidou, Markos G Tsipouras, and Spiros Konitsiotis. *Automated epileptic seizure detection methods: a review study*. Citeseer, 2012.

[41] Stanford University. *CS231n Convolutional Neural Networks for Visual Recognition*. 2016. URL: `http://cs231n.github.io/convolutional-networks/` (visited on 10/03/2016).

[42] WRS Webber, Ronald P Lesser, Russell T Richardson, and Kerry Wilson. "An approach to seizure detection using an artificial neural network (ANN)". In: *Electroencephalography and clinical Neurophysiology* 98.4 (1996), pp. 250–272.

[43] Scott B Wilson and Ronald Emerson. "Spike detection: a review and comparison of algorithms". In: *Clinical Neurophysiology* 113.12 (2002), pp. 1873–1881.

[44] Mohammad Hossein Zarifia, Negar Karimi Ghalehjogh, and Mehdi Baradarannia. "A new evolutionary approach for neural spike detection based on genetic algorithm". In: *Expert Systems with Applications* 42.1 (2015), pp. 462–467.

[45] Matthew D Zeiler, M Ranzato, Rajat Monga, Min Mao, Kun Yang, Quoc Viet Le, Patrick Nguyen, Alan Senior, Vincent Vanhoucke, Jeffrey Dean, et al. "On rectified linear units for speech processing". In: *2013 IEEE International*

Conference on Acoustics, Speech and Signal Processing. IEEE. 2013, pp. 3517–3521.

[46] ETH Zurich. *Neural Control of Movement Lab*. 2016. URL: `http://www.ncm.hest.ethz.ch/the-group/our-labs.html` (visited on 11/15/2016).

VITA

Howard Carey received his BSc degree in Physics from James Madison University, Harrisonburg, VA, USA, in 2011. He is currently working toward his MSc. in Computer Science at Virginia Commonwealth University, Richmond, VA, USA. His research focus is in the field of machine learning and artificial intelligent algorithms, with a more recent interest in deep learning and convolutional neural networks.

Papers:

1 **Carey H**, Arsenovic P, Manic M, Epileptic Spike Detection with EEG using Artificial Neural Networks, IEEE Human Systems Interaction, Portsmouth, UK, 2016