SPATIOTEMPORAL ANALYSIS OF SPREADING DEPOLARIZATIONS AND SPREADING DEPRESSIONS

A Thesis

Presented to

The Faculty of the Department of Electrical and Computer Engineering

California State University, Los Angeles

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

in

Electrical Engineering

By

Axel Ochoa Aguirre

September 2022

© 2022

Axel Ochoa Aguirre

ALL RIGHTS RESERVED

The thesis of Axel Ochoa Aguirre is approved.

Deborah Won, Ph.D., Committee Chair

Marina Mondin, Ph.D., Committee Member

Charles Liu, Ph.D., Department Chair, Committee Member

California State University, Los Angeles

September, 2022

ABSTRACT

Spatiotemporal Analysis of Spreading Depolarizations and Spreading Depressions

By

Axel Ochoa Aguirre

Spreading depolarizations (SD) refer to ripples in the patient's brain's electrical activity that are associated with a depression of overall brain activity levels when the brain is damaged. One of our long-term goals is to develop a closed-loop stroke therapy which is continuously adapted and tailored based on the patient's current brain state. Such a therapy will require us to explore spatiotemporal patterns that emerge. With new microarrays developed to record simultaneously from many electrodes in awake behaving rats, our first step was to attempt to classify brain regions according to electrophysiological characteristics. We hypothesized brain regions affected by the stroke are distinguishable from healthy and compromised tissue by electrophysiological features. We first defined and measured electrophysiological characteristics on each channel. Then, we developed data analytics to cluster electrode channels by electrophysiological features, including SD peak amplitude, SD width, and baseline DC levels. K-means clustering classified channels into contiguous brain regions in 3 rats. In all three of the rats, the channels clustered into three physically contiguous brain regions, which we suspect correspond to injured, compromised, and healthy brain tissue regions. Our future work will explore where these depressions originate and how they spread across the brain as time progresses.

ACKNOWLEDGMENTS

This is the segment where I look back at the beginning of my journey and reflect on the people who have accompanied me along my journey, shaping me in their own way. I can say however without a doubt my mentor Dr. Won has been the most influential person to accompany m in this journey. She is constantly pushing her students to do more and be better step outside our comfort zones and see what more we can achieve when we give ourselves the confidence to do so. Its moments like coaching me to take leadership in engineering projects, to pushing me to attending conferences across the world that really show how, Dr. Won never fails to help bring the best out of her students. I can only say I am grateful to have had the opportunity to work with her.

In my personal life I must give thanks to my mom and sister for pushing me to do more. They are the reason I am still in school stressing out over tests and deadlines, but they know I wouldn't like to live any other way. They push me to keep going to see how far I can go, every check point I reach along the my journey makes them cheer me on more and makes me want to try harder to not let them down.

And of course, not of this would be possible without financial support for that I have to give thanks to the NSF for their financial support for this work under award No. HRD-1547723.

TABLE OF CONTENTS

Abstract iv

Acknowledgments v

List of Tables viii

List of Figures ix

Chapter

1. Measurable Indicators of Stroke 1

1.1 Stroke Prevalence, Incidence, Current Treatments / Intervention 1

1.2 Current Treatment and Intervention 2

1.3 Electrophysiological Markers of Stroke – Spreading Depolarization, Depression 4

1.4 Prospects for Future Stroke Therapies 6

2. Chronic Recordings in Awake Behaving Rats with Induced Stroke 8

2.1 Technical Advances in Micro ECoG Arrays 8

2.2 Stroke Model 11

2.3 Experimental Paradigm 12

3. Data Analytics 15

3.1 ECoG Features and SD Detection 15

3.2 SD Detection 22

3.2.1 SD Event Features 25

3.3 DC Baseline Analysis 26

4. Clustering of ECoG Features by Tissue Health / State 32

4.1 K-Means Clustering 32

4.2 Determining Number of clusters....33
4.3 Mapping Clusters onto Physical Location....34
4.4 Relating Clusters Defined by Electrophysiological Features to Tissue Health....45
5. Future Directions....49
References....51
Appendices
A. MATLAB Script Files....58
A.1 Main....58
A.2 Remove Data DC Component....68
A.3 Derive Cross Correlation and Plot....68
A.4 Plot Cross Correlation Results as Heatmaps....70
A.5 Derive Key Features Using Moving Average....71

LIST OF TABLES

Table

1.1 Immunohistochemistry result shared with CSULA by Dr. Hoffmann 13

LIST OF FIGURES

Figure

1.1 Depictions of the two types of strokes. Left depicts a blockage in an artery caused by a blood clot, the blockage prevents oxygen from reaching the brain thus causing an ischemic stroke. Right depicts a ruptured artery causing bleeding within the brain causing a hemorrhagic stroke 2

1.2 The core of a stroke with an outward expanding penumbra.....3

1.3 Comparison of healthy neuron against neuron with compromised NaCl pump.....5

2.1 Depiction of the developed micro electrode array developed by Dr. Viventi for the purposes of this study.....11

2.2 Photograph of rat subject post surgery, with recording equipment set.....12

3.1 ECoG data recording of rat 9 in file 5 with green lines depicting SD event beginning and red lines indicating event end times. The start and end times were marked by an SD expert and were used as our ground truth to develop our SD event detection algorithm.....16

3.2 Recorded ECoG data of rat 17 data file 3, SD events' start and end times are denoted by green and red lines respectively. Unlike rat 9 file 5 the SDs in this data file are not as prominent and obvious, hence the need for an algorithm to calculate correlation metrics. 16

3.3 Graphs depicting the cross correlation of data channel E6 against every other channel on the electrode array as indicated above each subplot. Correlation coefficients at each lag were obtained by shifting E6 in time in both the positive

and negative direction. Although an 8 by 8 array was used when recording data only 60 of the data channels were used the unused data channel are depicted as DNE in the set of correlation functions depicted; as shown in Fig. 3.4 19

3.4 Depiction of the placement of the electrode array in reference to the subjects' skull. Unused data channels are blacked out. 20

3.5 Correlation metrics are sampled using evenly spaced time lags, such that the reference channel is shifted in fixed increments of time. The resulting correlation metric is mapped onto a heat map to visualize how the activity is spreading across space over time. 22

3.6 Detection of spreading depolarization events entailed: A) The ECoG signal from channel G1 with the ground truth start and end times manually marked by an SD expert. The green vertical line indicates the ground truth for the start of an SD and the red vertical lines corresponds to the ground truth of the end of an SD. B) The first order difference of (A) with the start and end times also marked in the same fashion. Peaks are also marked in dark blue, and any valleys are marked in cyan. C) Second order difference which was used to determine the crossings in (B) 24

3.7 Depicting what the features extracted from each of the SD events, the red lines Denote the markers time stamps the program used to derive the length of the SD event, the black lines denote the markers the program used to derive the amplitude, of the SD event.....26

3.8 Best fit line is calculated and superimposed onto the raw data signal used to calculate said best fit line. The best fit line does capture a good image of what is going on overall but the nature of the best fit line fails to capture important key

concepts such as the steep slope seen in the beginning of the recording before the first SD event 28
3.9 Piece wise best fit lines are calculated using split data files rather than the entire recording, the steep slope seen before the first SD event is now better represented it is logged in the first best fit line's slope. 29
3.10 The blue data signal is the raw ECoG signal, the black line is the moving average resulting signal, the green and red vertical lines represent the considered window's start and stop times respectively, the magenta lines denote the derived best fit line of its respective SD event.....31
4.1 Distortion metric decreases as number of clusters *k* increases, but there are greatly diminishing returns after the "elbow" in the curve, which occurs at *k*=3.....34
4.2 Results of the clustering algorithm for rat nine file five plotted in space and color coded, unused channels are denoted with white space. The figure shows promising results as the segregated channels form contiguous regions 35
4.3 Figures depicts the result of the clustering algorithm results for rat subject 9 file 5. Left is the raw ECoG data, right is the same data channels color coded based on the clusters. 35
4.4 Results of the clustering algorithm for rat 9 file 6 plotted in space and color coded, unused channels are denoted with white space 36
4.5 Figures depicts the result of the clustering algorithm results for rat subject 9 file 6. Left is the raw ECoG data, right is the same data channels color coded based on the clusters. 37

4.6 Results of the clustering algorithm for rat 16 split file 2 - 1 plotted in space and color coded, unused channels are denoted with white space 37
4.7 Figures depicts the result of the clustering algorithm results for rat subject 16 split File 2 – 1. Left is the raw ECoG data, right is the same data channels color coded based on the clusters. 38
4.8 Results of the clustering algorithm for rat 16 split file 2 - 2 plotted in space and color coded, unused channels are denoted with white space 39
4.9 Figures depicts the result of the clustering algorithm results for rat subject 16 split File 2 – 2. Left is the raw ECoG data, right is the same data channels color coded based on the clusters. 39
4.10 Results of the clustering algorithm for rat 16 split file 2 - 2 plotted in space and color coded, unused channels are denoted with white space 40
4.11 Figures depicts the result of the clustering algorithm results for rat subject 16 split file 2 – 3. Left is the raw ECoG data, right is the same data channels color coded based on the clusters.....41
4.12 Results of the clustering algorithm for rat 17 file 3 plotted in space and color coded, unused channels are denoted with white space 41
4.13 Figures depicts the result of the clustering algorithm results for rat subject 17 file 3. Left is the raw ECoG data, right is the same data channels color coded based on the clusters 42
4.14 Results of the clustering algorithm for rat 18 file 3 plotted in space and color coded, unused channels are denoted with white space 43

4.15 Figures depicts the result of the clustering algorithm results for rat subject 18 split file 3. Left is the raw ECoG data, right is the same data channels color coded based on the clusters. 43
4.16 The immunohistochemistry results rat 9 with rat nine file 5 overlayed to show the results of the clustering algorithm. Even though there are contiguous regions that suggest successful tissue health classification these staining results are the only form of immunohistochemistry we have to the accuracy of our model can be debated. 47
4.17 The immunohistochemistry results rat 16 with rat 16 file 2 - 1 overlayed to show the results of the clustering algorithm 47
4.18 The immunohistochemistry results rat 17 with rat 17 file 3 overlayed to show the results of the clustering algorithm 48
4.19 The immunohistochemistry results rat 18 with rat 18 file 3 overlayed to show the results of the clustering algorithm 48

CHAPTER 1

Measurable indicators of stroke

1.1 Stroke Prevalence, Incidence, and Outcomes

Strokes are one of the more deadly diseases in modern American times yearly nearly 800,000 Americans suffer from some form of stroke with about 160,000 of those leading to deaths it is no surprise it holds a spot among the top leading causes of death among Americans [1]. Of these less lethal stroke patients 75% end up with some form of dysfunction, with 15% to 30% remaining with a severe disability that drastically changes their quality of life.

Although there are up to five different types of stokes observed, ischemic stroke, hemorrhage stroke, transient ischemic attack, brain stem stokes, and cryptogenic stroke. The major ones are ischemic and hemorrhagic strokes, the rest can simply be described as variations of these two strokes [2]. Hemorrhagic strokes are caused by ruptures in arteries in the brain, these ruptures are mainly can be caused by high blood pressure and aneurysms. The second type of stroke, ischemic stroke, are by far the more common and are the focus of this study, this variation accounts for about 87% of all stokes [3]. These strokes are caused by a blockage in a brain artery. These blockages can be caused by blood clots, or fatty deposits which lead to loss of oxygen and subsequently tissue damage.

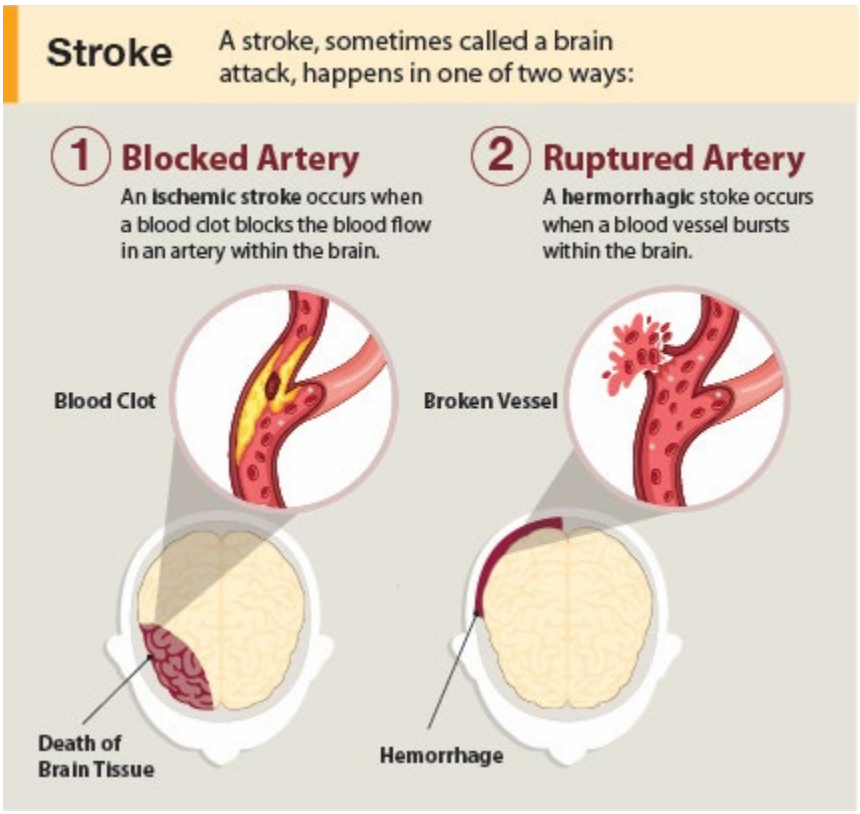


Figure 1.1: Depictions of the two types of strokes. Left depicts a blockage in an artery caused by a blood clot, the blockage prevents oxygen from reaching the brain thus causing an ischemic stroke. Right depicts a ruptured artery causing bleeding within the brain causing a hemorrhagic stroke [4]

Stroke can lead to varying degrees of debilitation. Stroke commonly leads to motor control deficit, mobility impairment and speech impairment. Stroke can be severe enough to lead to partial or complete paralysis and cognitive deficit. Of course, if action to stop the damage is not taken quickly enough, stroke can be fatal.

1.2 Current Treatments and Intervention

Current treatments for stroke depend heavily on the circumstances in which the patient finds themselves when a suffering from a stroke.

Immediate treatment begins by removing the blockage causing the stroke this can be done in a variety of ways. Ideally treatment includes using a thrombolytic drug known as tissue plasminogen activator (tPA). Thrombolytic drugs are used to break up blood clots, using drugs to treat the stroke is ideal as it eliminates the need for surgery to physically remove the clot. Ideally tPA is administered within 30 minutes and is used as treatment if it can be administered within the first 3 hours of the stroke. If this threshold is exceeded or if the patient is not responding positively to the drugs, endovascular surgery is next best option to remove the clot, a catheter is inserted up the afflicted artery where a medical equipment such as suction tubes or claws are used to physically removes the blockage [5] [6] [7] [8] [9]. This procedure is ideally done within the first 6 hours of the stroke. From these procedures it is clear that a critical factor in stroke treatment is time, the longer the patient remains without treatment the more tissue is compromised and damaged. Moments immediately after the stroke are critical, there two main areas of interest in the brain at this moment, the area of the infarct otherwise known as the core, and the tissue region immediately surrounding it known as the penumbra.

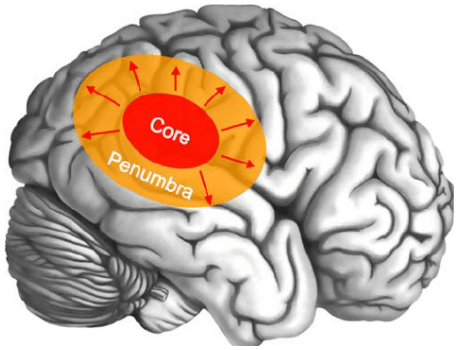


Figure 1.2: The core of a stroke with an outward expanding penumbra [10]

While the core tissue suffers from irreversible brought on by the ischemic stroke, the surrounding penumbra tissue is compromised; this tissue, with treatment, can repolarize diminishing or even eliminating long term heath issues. Live monitoring of tissue health in stroke is currently possible; but not practical as they either involve invasive surgeries or lack the resolution necessary to make sounds medical decisions. Our research seeks a less invasive, high-resolution method of classifying tissue health live, to aid patient treatment and recovery therapy.

1.3 Electrophysiological Markers of Stroke – Spreading Depolarization, Depression

To study strokes, we study spreading depolarizations (SD) in electrocorticographic (ECoG) data taken from surface electrodes implanted on subjects' brain. SDs are benign events that are observed in normal electrical brain activity, they are associated with tissue damage as they have been known to manifest when patients suffer head trauma such as stroke, migraines, and the like. Although the SD events themselves are known to be mostly benign, they can be used as a marker of tissue health and thus we intend to use them their characteristics to analyze and classify tissue health.

SDs appear as a relatively high amplitude triphasic wave form in the recorded electrical brain typically in the range 5 to 15mV in amplitude lasting for a 1 to 2 minutes. They originate in the area of the infarct in terms of strokes and tend to propagate outwards at speeds of 2 to 6 mm/min across brain grey matter [11] [12]. This propagation is what compromises the health of the local tissue that is not part of the area of the infarct to and spreads the effect strokes. SDs are cause by an imbalance of ions in the electrical potential in the dying cells. During the events of a stoke the sodium potassium pumps in the area of the infarct fail, typically they help maintain a healthy balance of K^+, Na^+, and

Ca^{2+} ions balancing the incoming current with its own outward current. During the events of a stroke the NaCl pumps are unable to keep up with the persistent inward current and thus lose ion equilibrium within the cell. To maintain homeostasis, water rushes into the cells causing intracellular swelling and a dramatic decrease in the local electrical potential and thus the SD event.

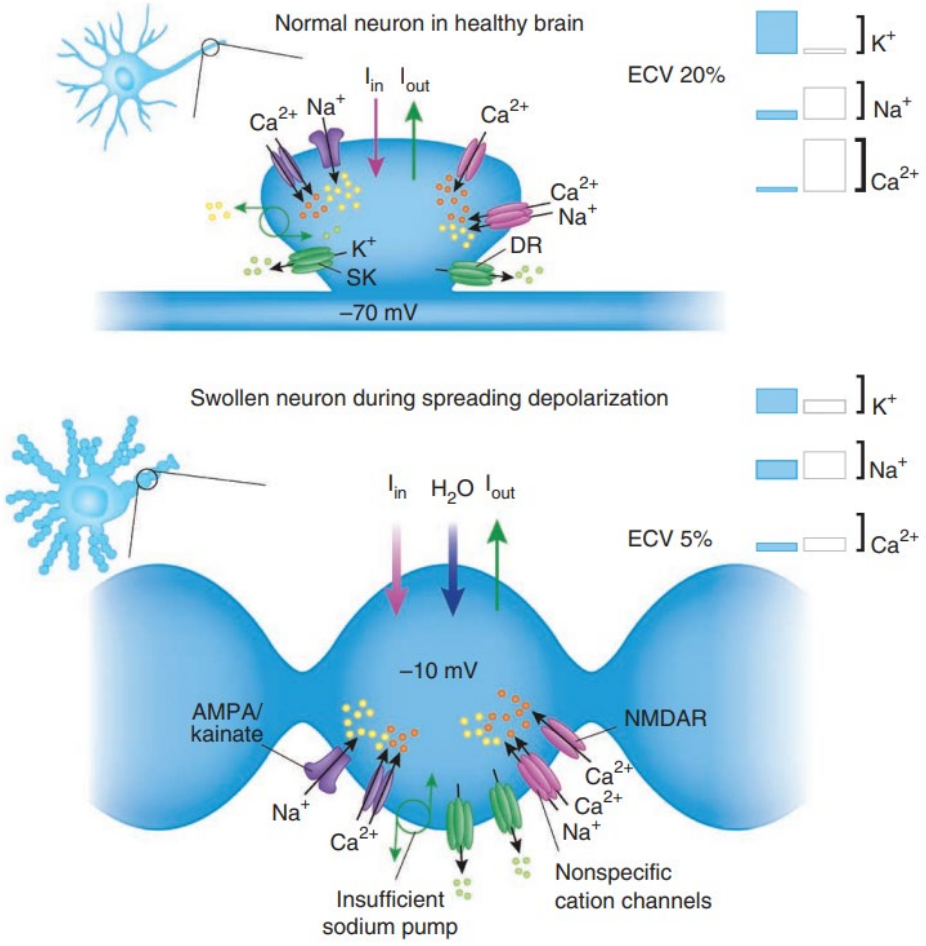


Figure 1.3 Comparison of healthy neuron against neuron with compromised NaCl pump [13]

Spreading depressions are an induced after-effect of SDs, characterized prominently by a silencing in ECoG brain activity as well as decline in DC levels and are believed to be the

main cause of permanent brain damage. In healthy tissue, after the events of the SD the local electrical potential will typically stabilize at a lower baseline and in time will return to the baseline it had before the SD event. Under damaged tissue however the baseline decrease is not only more prominent, but the baseline does fully recuperate. We know there is a threshold in the ECoG potential if this threshold is met or exceeded the cell experiences an electrochemical energy breakdown, which is what is believed to cause irreversible brain damage. With each progressive SD lowering the electrical potential of the tissue we can logically conclude, the frequency and total number of SD events is directly related to state of health of the recorded tissue.

1.4 Prospects for Future Stroke Therapies

Acute stroke intervention has been described in Section 1.2. The primary approaches of fibrinolysis via administration of anticoagulants and endovascular surgery are carried out as soon as possible in order to prevent tissue death and limit the initial infarct area [14]. However, as described, the damage can continue to spread despite the acute interventions. Furthermore, surgical intervention is risky and can itself lead to complications and death. Longer term treatment of stroke symptoms primarily involve physical or occupational therapy. With improvements in high precision monitoring technology and the ability to monitor neurological changes continuously and chronically, there is hope of advancing treatment to further limit damage and improve the clinical outcome of the originating ischemic or hemorrhagic incident. Some of these potential advancements in stroke therapy include cell transplantation and electrical stimulation [15]. Cell transplantation is a procedure that involves using stem cells to repair damaged tissue in the brain. Stem cell research fall under one of two categories, endogenous and

exogenous. Endogenous essentially refers to using stem cells that the body is already making in order to repair damaged tissue in the brain. The goal is to stimulate body into creatin more of these stem cells to help the body repair itself this is done using chemokine receptors such as stromal derived factor1 and integrin β1 or using electrical fields to direct these stem cells is also currently being investigated, research is developing but is still locked to preclinical studies. Alternatively, there are exogenous stem cells these are stem cells that are added into the body. There are three main variations to this procedure, immortalized cell lines, human derived stem cells, and bone marrow delivered cells. While these methods have been studied they have yet to be fully understood, though the conducted studies have concluded with promising results. Aside from the ethical issues regarding stem cells however there are other factor limiting this approach, such as supply limitations.

Electrical stimulation is another stroke treatment that has been developing. So far there are both invasive and non-invasive treatments using electrical stimulation transcranial magnetic stimulation and implantable epidural electrodes, both treatments show promise to be beneficial to the patient. When the brain suffers from a stoke there remains a healthy hemisphere of the brain, the contralesional hemisphere. The electrical potential of this hemisphere has been shown to alter the electrical potential of the damaged hemisphere, the effects of this is still being studied whether it improves recovery or is detrimental to recovery. Therefore one of the goals of electrical stimulation aims to maintain a balance between these two regions of the brain, this is complemented by the fact that one of the benefits of electrical stimulation can be targeted affecting only select the regions of the brain.

CHAPTER 2

Chronic Recordings in Awake Behaving Rats with Induced Stroke

The experimental methods carried out by our collaborators at Duke University in Dr. Ulrike Hoffman's laboratory will be described here in Ch. 2. Their research group developed technology and methods to study the electrophysiological behavior in the cortex *in vivo* after ischemic injury. In this chapter, the advances made in neural acquisition technology that enabled such studies and the experimental protocol used to acquire the data used in my thesis work will be described.

2.1 Technical Advances in Micro ECoG Arrays

Thus far, the ability to study how brain injury develops in stroke patients and how the electrophysiological manifestations correlate with clinical outcome have been limited. One current method of studying strokes involves studying the brain postmortem. This is done by immunohistochemically staining slices of the patient's brain to analyze the effect the stroke had on the brain. Immunohistochemistry only provides a snapshot of the brain's chemical composition once the patient is deceased. Obviously, the progression of the stroke in real time cannot be studied this way.

Different studies have used various variables some of the more common ones include intracranial pressure, cerebral perfusion pressure, local tissue partial pressure of oxygen, and scalp electroencephalography to name a few. While the analysis of these results have shown promise in giving information of the progress of brain tissue health. One variable that has yet to be fully explored includes electrocorticographic (ECoG)

brain activity. Chronic electrophysiological recordings potentially provide a method by which the effects of stroke *in vivo* in real time could be studied.

To study these electrophysiological patterns, simultaneous recordings from across the affected cortical region would be required. Such studies would require invasive intracranial surgery in order to implant an electrode array onto the surface of the brain. Thus, we chose to study the phenomenon of stroke-related spreading depolarization in animals first with the expectation that these results will inform future clinical studies. The analysis in this thesis was conducted on electrophysiological data acquired from an electrode array. This section will discuss the developments made in the technology of micro electrode arrays in order record ECoG data that differentiate our data from other currently available methods and data.

There are microelectrode arrays already developed for *in vivo* recordings among these are intracortical microelectrodes, with are mainly used in pre-clinical research. The main disadvantage to these recordings is the fact that the surgery needed to properly place these recording electrodes requires the patient to be anesthetized and therefore leads to short term non-survival trials. The fact that viable electrode sites are limited, and the patient is not awake and moving, results in potentially distorted data.

A second method of recording data in pre-clinical research are epidural recording screws, these are electrodes that are physically screwed into the subjects' skulls. The advantages to this method involve, being able to perform this procedure without needing to anesthetize the subjects. Unfortunately, this procedure to results in low resolution data recordings, too low for the purposes of accurately studying SDs and are thus less than ideal.

Finally, there are linear electrode arrays, these electrode arrays typically contain four to six electrodes typically with a 4mm diameter. This electrode strip is surgically implanted on the surface of the subject's brain. One of the drawbacks of this method is this type of electrode array requires intracranial surgery on a stroke patient, aside from adding the unnecessary risk of surgery complications to the patient the electrode array needs to be precisely implanted directly over the area of the trauma. This coupled with that fact that the electrode has a such a low electrode density, makes detection of SD events a challenge even if properly implanted.

This led to the development of custom-made electrode array designed by Dr. Jonathan Viventi and his research team at Duke University. This custom 60 channel electrode array is outfitted with high-density μECoG, platinum iridium coated gold electrodes to reduce the impedance which achieved levels of ~600 to 700 kΩ at 0.5Hz. The electrodes were DC coupled to five recording screws using a follower circuit with ~1 GΩ impedance; the screws were tapped into the subjects' skulls down to the dura, connected to one another using silver wire, the recorded signals of the recording screws are then averaged. The overall electrode array has dimensions of 3.4 mm by 3.4 mm allowing for recording of data at the area of the infarct as well as the surrounding tissue. With a thickness of less than 50 μm, this electrode array can easily be implanted beneath the subject's skull without requiring major surgery; with the added benefit of being thin enough to not disrupt local brain activity.

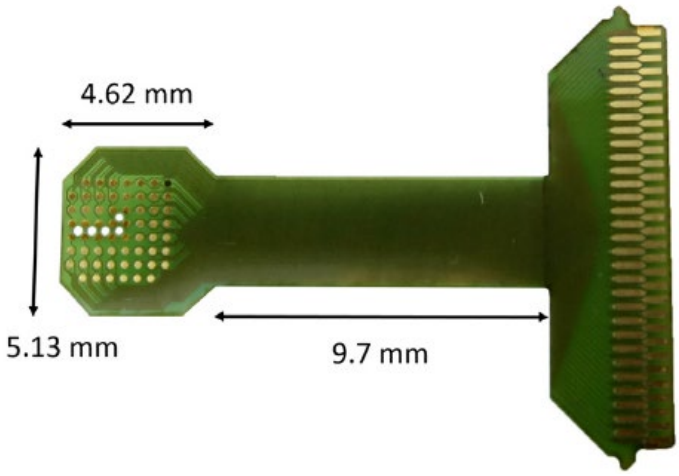


Figure 2.1: Depiction of the developed micro electrode array developed by Dr. Viventi for the purposes of this study [16]

Moving on to the data acquisition hardware the typical high pass filter used in headstage amplifier is removed, the overall gain is reduced to x3, and accommodated the input range of the ECoG data and SDs events. These augmentations allow for the recording of ultra-slow, near DC levels ECoG signal recordings. Next an 18-bit analog to digital converter is used, all tethered using silver wires to allow for low noise data recordings of *in vivo* ECoG data signals [17].

2.2 Stroke Model

Initially to test the performance of the electrode array developed by Dr. Viventi two rats were used as subjects where SDs were induced by exposing the subjects' frontal cortex to topical doses of KCl. The subjects' ECoG activity was then monitored by implanting both the developed electrode array along with glass microelectrodes implanted into the subjects' brain parenchyma.

The recorded data was taken with a focal cerebral ischemia induced into to the subjects' brain. Three days after the electrode array was implanted in 20 of the rat subjects' brain, the subjects were once again operated on undergoing middle cerebral

artery occlusion (MCAO), the external carotid artery was isolated and blocked off by inserting commercially available nylon monofilaments for 60-90 minutes. The monofilaments were then removed, and wounds were closed, to allow for reperfusion. Immediately after anesthesia wore off, medical analysis was done to assess the success of stroke induction as well as mortality, of the surgery.

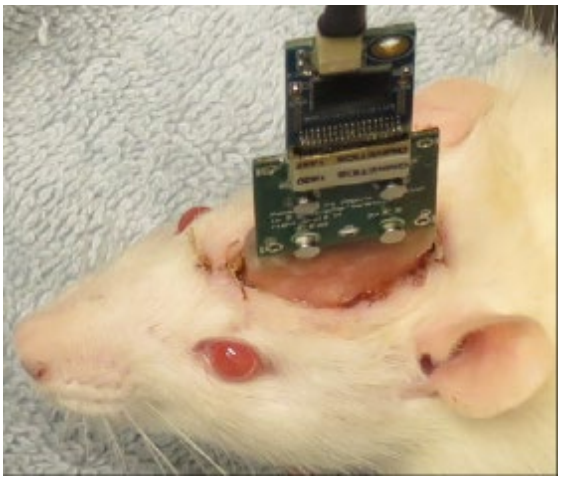
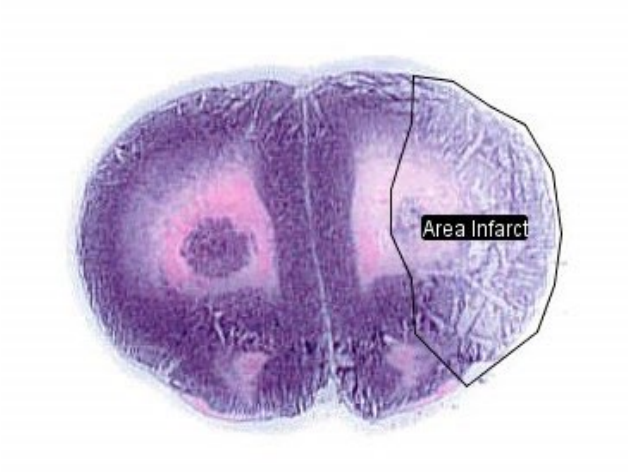


Figure 2.2: Photograph of rat subject post surgery, with recording equipment set [18]

2.3 Experimental Paradigm

Dr. Ulrike Hoffman and her team at Duke University recorded ECoG data signals in live rats by surgically implanting an epidural micro electrode array, onto the surface of a subjects' brain activity. A total of 25 rat subjects were awake and aware while their ECoG brain activity is measured, using the 60-channel electrode array custom designed and developed by Dr. Jonathan Viventi's research group at Duke University. A stroke was then induced by Dr. Hoffman's team and the various subjects' activity was recorded. The rats were awake behaving anesthetized after the stroke was induced until their death which varied from 1 to 21 days. The ECoG brain activity of a seven of the rats, described in table 1.1 were shared with us to analyze in CSULA.

Rat Number	Time Recorded (Days)	Immunohistochemistry
9	1	
10	No Information	No Information
11	8	

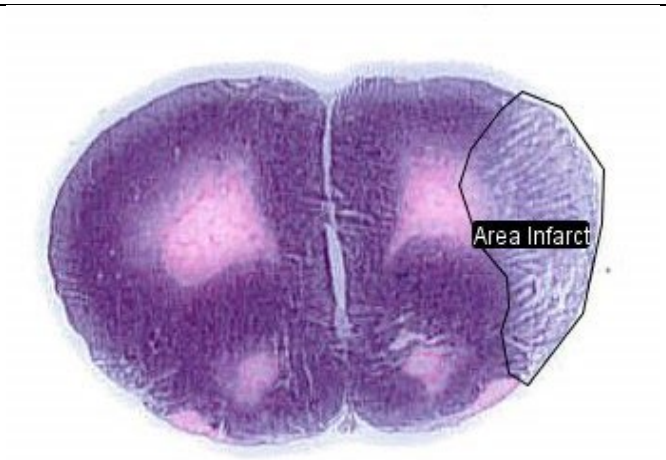
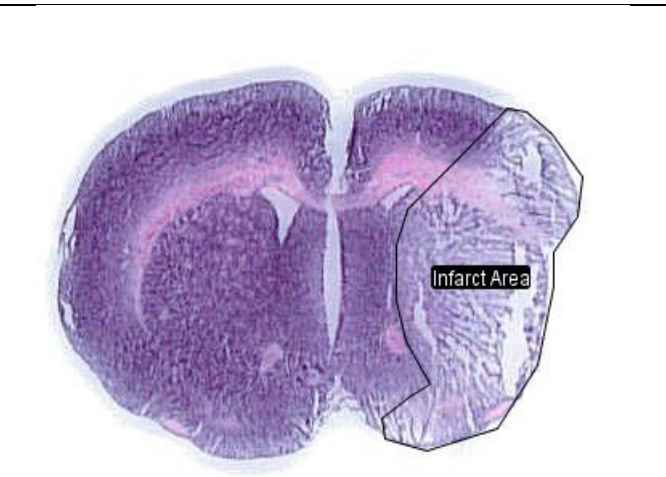
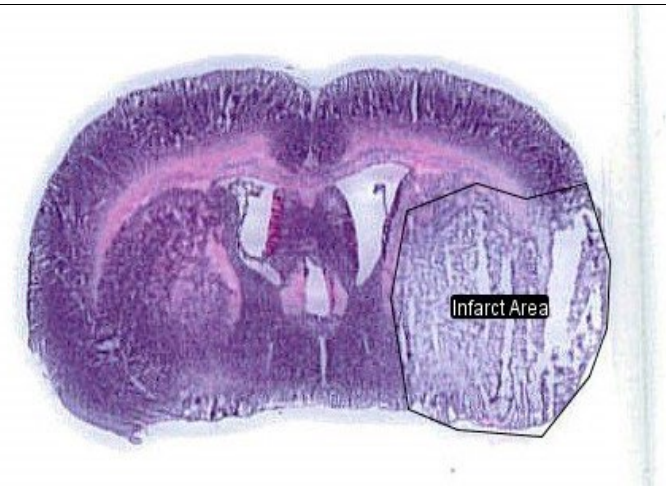
16	12	
17	11	
18	18	

Table 1.1: Immunohistochemistry result shared with CSULA by Dr. Hoffmann

CHAPTER 3

Data Analytics

Various signal processing tools and data analytics were used to analyze the ECoG data and explore spatiotemporal patterns in the population signals. The methods for detection of the SD events are first described, followed by feature extraction of these SD events. The chapter delineates the methods used to search for patterns across the electrode array using cross correlation metrics, and concludes with a description of the clustering technique applied to the population ECoG data in an attempt to classify regions of the cortex into different health states.

3.1 Correlation Metrics

Upon initial inspection of the raw data as seen in Fig 3.1 below, a stereotyped pattern emerges, with features which distinguish some channels from each other. Some of the data channels appear stay at a relatively stable level, the amplitude of the signal remains overall constant throughout the recording, while other signals' average value consistently decreases as time progresses. We hypothesize, the data channels with the greatest decline in electrical activity correspond to the electrodes recording the electrical activity over the area of the stroke.

The start and stop of manually detected ground truth SD events are demarcated in Fig. 3.1 by the green and red markers, respectively. The marked SD events in all data channels in this sample recording exhibit a similar stereotyped triphasic waveform. A

steady decline in the overall electrical potential is also visible on a number of the data signals.

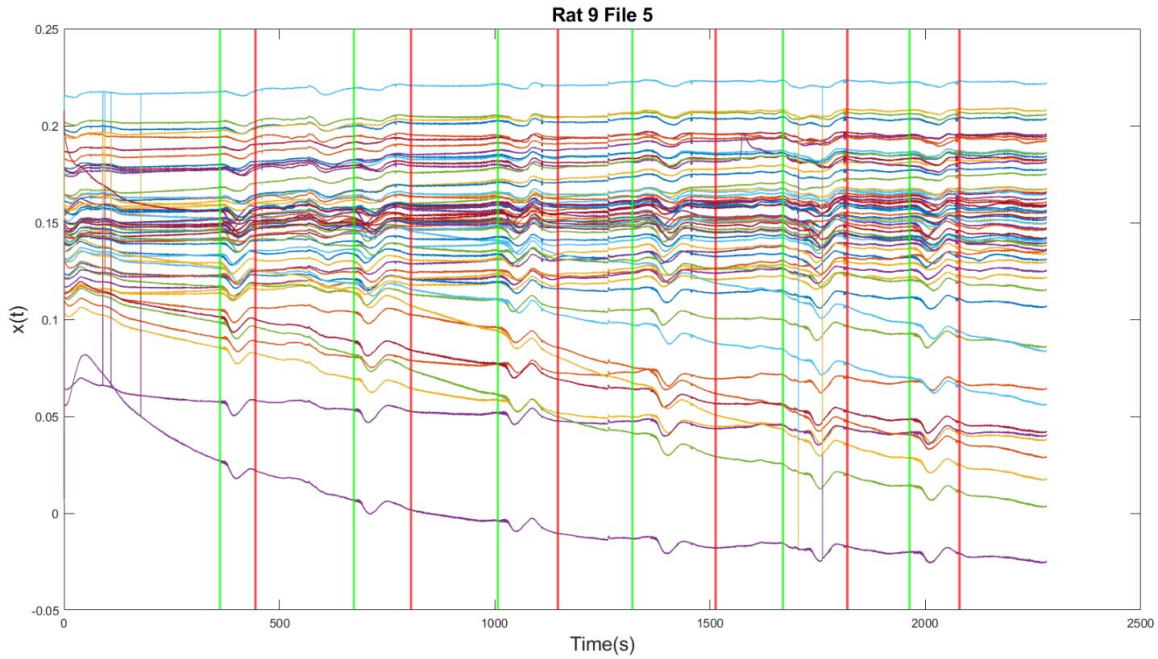


Figure 3.1: ECoG data recording of rat 9 in file 5 with green lines depicting SD event beginning times and red lines indicating event end times. The start and end times were marked by an SD expert and were used as our ground truth to develop our SD event detection algorithm.

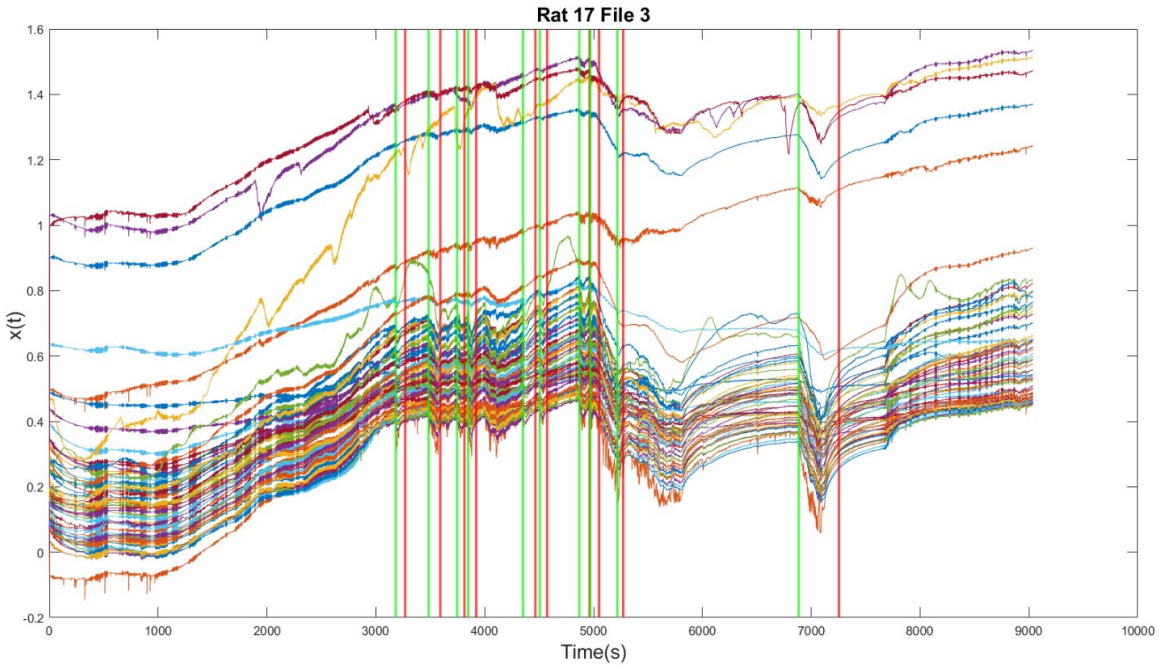


Figure 3.2: Recorded ECoG data of rat 17 data file 3, SD events' start and end times are denoted by green and red lines respectively. Unlike rat 9 file 5 the SDs in this data file are not as prominent and obvious, hence the need for a math algorithm to calculate correlation metrics.

This is not the case for all data sets; however, we suspected that there was correlated activity that was propagating across the array because the neurons that are in the region covered by the array are potentially in an interconnected neural network. The Pearson correlation coefficient between each pair of channels was calculated according to Eqn. 1, using the MATLAB's built-in function `corrcoef` to see how strong linear relationships were, if any, existing that cannot be seen with the naked eye.

$$r = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum(x_i - \overline{x})^2 \sum(y_i - \overline{y})^2}}$$

Equation 1

where *r*, corresponds to the correlation metric; x_i and y_i correspond to the i^{th} sample of signals *x* and *y*, respectively; and $\overline{x}$ and $\bar{y}$ correspond to the mean value of the respective data signal.

We know the electrochemical imbalance during a spreading depolarization propagate across the network and the spread and propagation velocity can be determined by using information derived when cross correlating the signals. Thus, even if the correlation coefficients were low, there could be a relationship and connections between the channels but due to network propagation delay could have phase delay. Therefore, cross-correlation between pairwise channels was additionally computed using MATLAB's `xcorr` function. The function calculates the correlation between two signals *x[n]* and *y[n]* at time lags from *–N* to *N*, according to Eqn. 2.

$$\hat{R}_{xy}(m) = \begin{cases} \sum_{n=0}^{N-m-1} x_{n+m} y_n^*, & m \geq 0 \\ \hat{R}_{yx}^*(-m), & m < 0 \end{cases}$$

Equation 2

where $\hat{R}_{xy}$(m) is the resulting correlation coeeficient between signal *y* (or *y[n]*) and signal *x* delayed by *m* lags (or *x[n+m]*)

The `xcorr` function does not normalize the data, therefore the signals were normalized manually before they were inputted into the `xcorr` function. The DC baseline for each channel is computed by taking the mean $\overline{x_i}$ of the respective channel's ECoG across time, according to Eqn. 3.

$$\overline{x_i} = \frac{1}{N} \sum_{n=1}^{N} x_i(n)$$

Equation 3

where $x_i(n)$ corresponds to the n^{th} sample of the i^{th} channel from 1 to 60, and N corresponds to the total number of samples in the current data file.

The DC baseline values were then stored for feature analysis described in Sec. 3.1 and removed from the raw ECoG according to Eqn. 4.

$$X_i = x_i - \overline{x}_i$$

Equation 4

The raw data signal x_i was then adjusted by subtracting the mean. The cross correlation was then applied to the resulting normalized data signals.

For each channel i, cross correlation function $Rij[l]$ with each channel j was computed across lags l according to Eqn. 2. Fig. 3.3 shows the cross-correlation function for each channel with the arbitrarily selected, channel E6. We can clearly see a pattern in the correlations plots; where the north west side of the electrode array displays a high correlation metric in the negative lags, as well as the positive lags, but remains negatively correlated in the center zero lag. Conversely in the south east side of the electrode array, there is high correlation in the center zero lag while the negative and positive lags remain negatively correlated. This indicates a propagation of activity across the array from the northwest generally toward the southeast end.

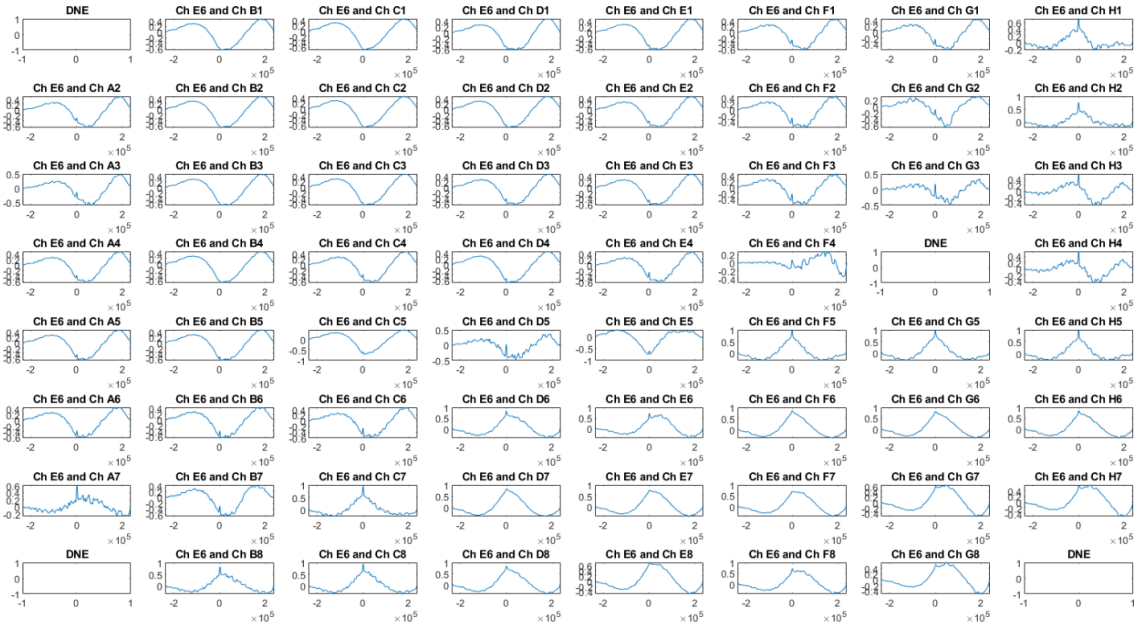


Figure 3.3: Graphs depicting the cross correlation of data channel E6 against every other channel on the electrode array as indicated above each subplot. Correlation coefficients at each lag were obtained by shifting E6 in time in both the positive and negative direction.

Although an 8 by 8 array was used when recording data only 60 of the data channels were used the unused data channel are depicted as DNE in the set of correlation functions depicted; as shown in Fig. 3.4

In order to further analyze patterns in these cross-correlation functions we stored a set of 8 by 8 matrices, populated with correlation coefficients at specified lags. In order to get a full picture of what is going on throughout the functions we take evenly spaced samples across the entire function. To do this we begin with a sample of the center zero lag and take samples every 192 seconds in both the positive and negative directions of the lags. These correlation metrics are then stored, with the elements corresponding to the unused channels, according to Fig. 3.4 set to "`NaN`".

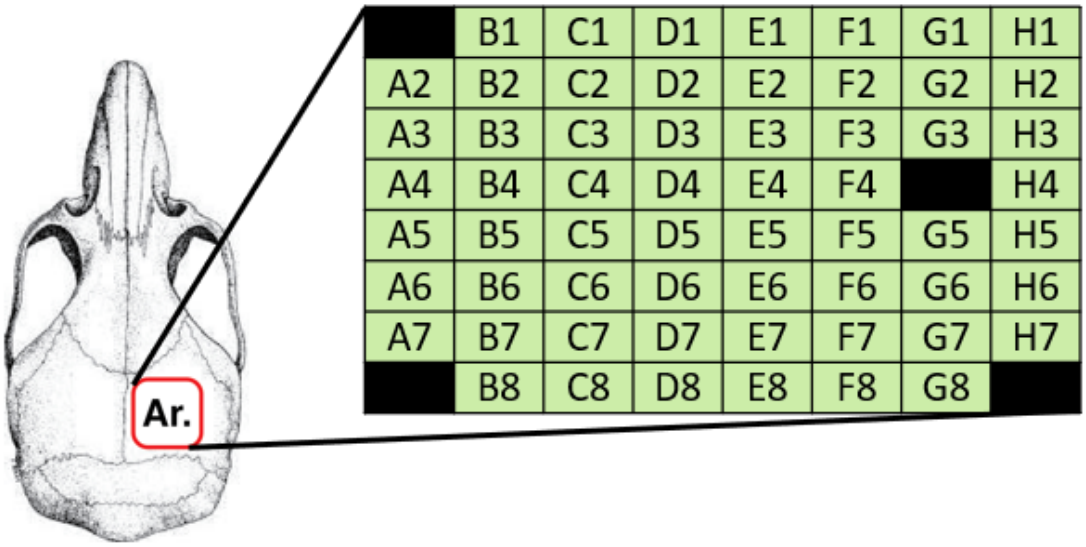


Figure 3.4: Depiction of the placement of the electrode array in reference to the subjects' skull. Unused data channels are blacked out.

These matrices are then mapped onto heatmaps in order to visualize the correlation across the array and changes over time shifts. A higher correlation value will be indicated by a "hotter" color (closer to the red wavelengths) and lower values will be indicated by "cooler colors (closer to blue wavelengths). Using these heatmaps we more easily analyze the data looking for any form of spatiotemporal shifts.

Ideally we hope to see a gradual spread of some kind across the electrode array. We hypothesized that because the stroke would originate in a specific section of the electrode array, as the stroke progresses we expect to see the afflicted area spread radially across the electrode array. Fig. 3.5 shows an example of a spatial map of the correlation values with a given channel across the array at various lags across the range of possible lags. Starting at the center, the time lag equals zero, seven lags are considered, shifting the selected reference channel's signal using intervals of 192 seconds, in the positive (moving right then down from the center in Fig. 3.5) and negative time direction (moving left then

up from the center). We can see in the heatmaps in Fig. 3.5 at zero time shift we have split in the data where the north west side of the electrode array is negatively correlated denoted by the blue coloring contrasting the south east side of the array which is positively correlated denoted by the yellow. As we increase the negative time lag of the reference channel E6, we see the north west side of the array shift towards positive correlation; this indicates that a common pattern of activity reaches electrode E6 later than the channels on the northwest end of the array. Conversely if we apply a positive time shift to the reference channel, the signals becomes more negatively correlated towards the south east side of the array. These results indicate that there exists some pattern of activity propagates from the northwest to the southeast side of the electrode array over several minutes (~25 minutes). We believe this is the result of how SD events propagate through the subject's brain.

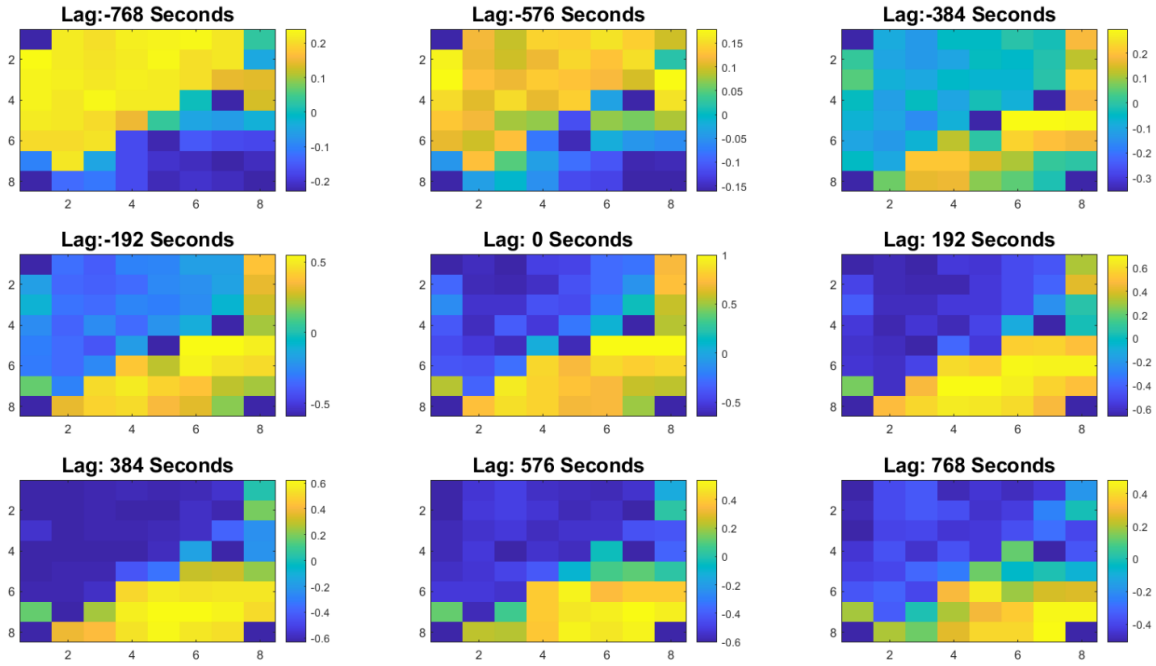


Figure 3.5: Correlation metrics are sampled using evenly spaced time lags, such that the reference channel is shifted in fixed increments of time. The resulting correlation metric is mapped onto a heat map to visualize how the activity is spreading across space over time.

3.2 SD Detection

Our goal is to classify the data, to do this first we need to identify and compute significant features contained in the data. The initial features attempt to identify are peak to peak amplitude and duration of the SD event. To do this the team initially developed a graphical user interface where an expert in SDs at Duke University, can easily load and explore select data sets in order to manually identify where these SDs are occurring with start and stop markers, an example is depicted in Fig 3.1. Using this information as our ground truth we worked to develop a MATLAB algorithm that can automatically detect SD events.

We know an SD event is primarily characterized by a basic triphasic wave form, therefore we develop an algorithm that detects what we define as 'peaks' and 'valleys'. A peak is defined as a point in which the data reaches a local maxima, while a valley is defined as when the data reaches a local minima. We approximate a derivative of the raw ECoG data by calculating its Euler difference as shown in Eqn. 5, we label this output data prime ($\dot{D}$).

$$f'(x) = f(x + \Delta) - f(x)$$

Equation 5

where we define delta as a small value relative to the data set. We chose to use the value of 20 seconds as this is how long peaks in SD events last. Using this data set we know the zeros of $\dot{D}$ can be used to determine where in the raw data these peaks and valleys exist. This process gives us markers to the samples where a local extrema exists in the raw data,

but it does not tell us whether they are peaks or valleys. Therefore, a second derivative is calculated using the aforementioned process. This new data set which we label data double prime ($\ddot{D}$), by looking at where $\ddot{D}$ crosses the x axis we can use the second derivative can be used to determine the concavity of the extremas calculated by $\dot{D}$, thus labeling them as either peaks or valleys. [19]

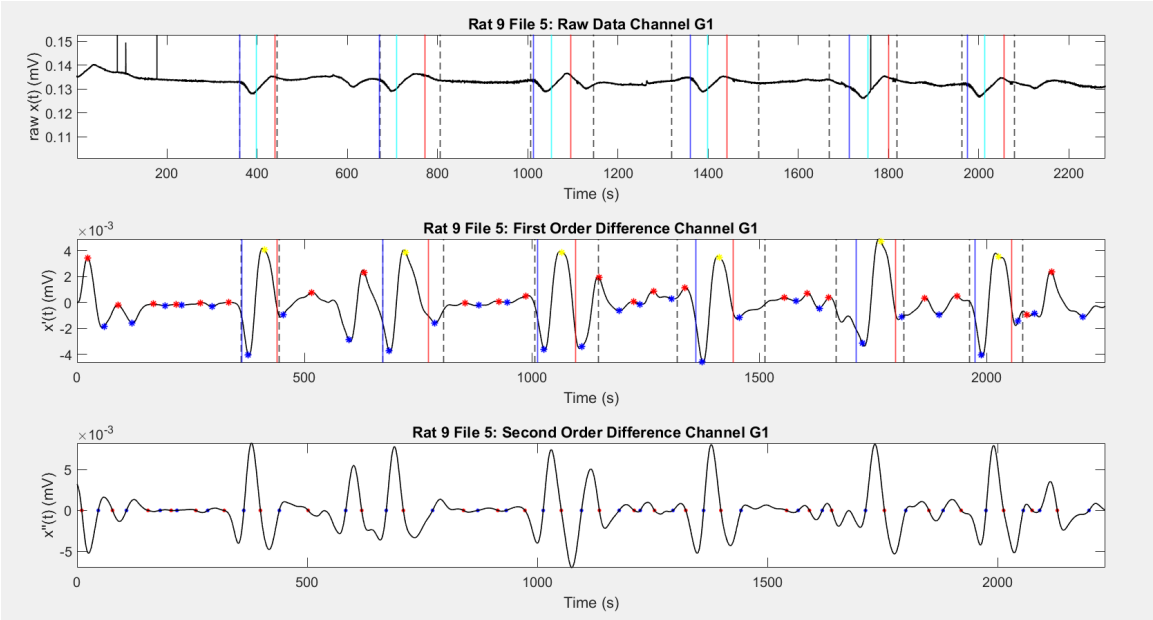


Figure 3.6: Detection of spreading depolarization events entailed: A) The ECoG signal from channel G1 with the ground truth start and end times manually marked by an SD expert. The green vertical line indicates the ground truth for the start of an SD and the red vertical lines corresponds to the ground truth of the end of an SD. B) The first order difference of (A) with the start and end times also marked in the same fashion. Peaks are also marked in dark blue, and any valleys are marked in cyan. C) Second order difference which was used to determine the 0-crossings in (B).

Knowing the nature of SD events, we seek triphasic a pattern of valley-peak-valley within an acceptable time frame. The algorithm begins by taking the derived peaks and searching for a valley both before and after the observed peak, these paired valleys must be within a time window of 50 seconds, if these conditions are met the algorithm denotes the event within elapsed time frame as an SD event, otherwise the peak is discarded and labeled as a false positive.

Problems arise with this method however, normal background ECoG brain activity that is not filtered out as noise can fit this description and be falsely label as a SD event. To filter these false positives we set a second condition taking advantage of another defining characteristic of an SD event, we take the largest SD event in the bank of derived events and set an arbitrary threshold of 65%. We then comb through the SD events, if they have an amplitude less than 65% of the largest recorded SD event, then it is discarded as a false positive caused by background brain activity otherwise it is labeled as a true positive.

3.2.1 SD Event Features

SD events were quantitatively characterized by identifying and measuring two key features: peak-to-peak amplitude and duration of the SD event. The SD event detection and feature extraction described in this subsection was developed and carried out by Andrea Abelian, an undergraduate research assistant in Dr. Won's lab. Once the times of the SD events are detected, the maximum and minimum amplitude values within the SD event window are determined. The duration of the SD event was defined as the time from the inflection point before the first valley in the first derivative until the inflection point after the second valley.

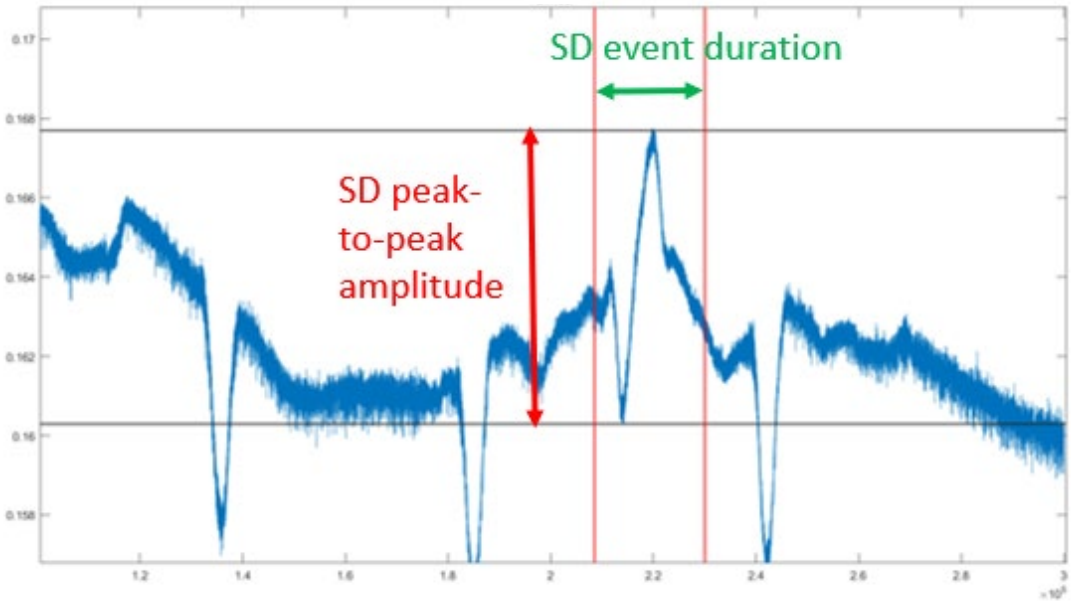


Figure 3.7: Depicting what the features extracted from each of the SD events, the red lines denote the markers time stamps the program used to derive the length of the SD event, the black lines denote the markers the program used to derive the amplitude, of the SD event.

These features were extracted for each detected SD in each data channel. Once all data channels in the data file have been logged the program looks through each data channel's recorded SD events and looks to match it to recorded SD events in other data channel's; if no matching SD event was recorded a zero-padding event is created and inserted with the appropriate time signature. This process is done in order to keep the total number of SD events per data channel consistent throughout each data file, and because an absence of an SD event that is present in other data channels is also a significant feature. The final table of receded SD events is logged into an Excel spreadsheet file.

3.3 DC Baseline Analysis

The next feature we measured about is the DC baseline of the raw data. The literature points to evidence that if the electrical potential depolarizes past some threshold, that neural tissue

reaches a point of no return, and the tissue dies [20]. In order to quantify this, we first record the average data channel's DC baseline. Individual channels are loaded in and fed into the built in MATLAB function `polyfit` in order to calculate the best fit first-degree polynomial, or line, for the ECoG signal of each channel. The coefficients of this simple linear regression function return the overall slope and DC offset of the recorded data.

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} \begin{bmatrix} x_1^2 & x_1^1 \\ x_2^2 & x_2^1 \end{bmatrix}$$

Equation 6

`Polyfit` uses the matrix algebra of Eqn. 6 where the p_m matrix is the weights needed to take the inputs, the x_m^n matrix, to the output values, the y_m.

One problem we have with this process however, is the change in the DC baseline level is not always linear. Some data sets seem to have a steady decrease in potential energy while other a more exponential decent that later seems to stabilize. In order to improve the accuracy of this feature we opted to split the data in two; the data is split at different intervals and best fit lines are derived. While alternative methods can be used such as potentially fitting the data onto an exponential function, some advantages to using this piece wise linear regression is we have more potential features to extract as well as getting an idea of local average offset. This contrasts with the exponential fit; it would output a single offset value that depending on the data channel observed may not yield an acute representation of how the average offset changes over time. Local offset is an important feature to monitor, we know there is a point of no return in the ECoG we need to look out for.

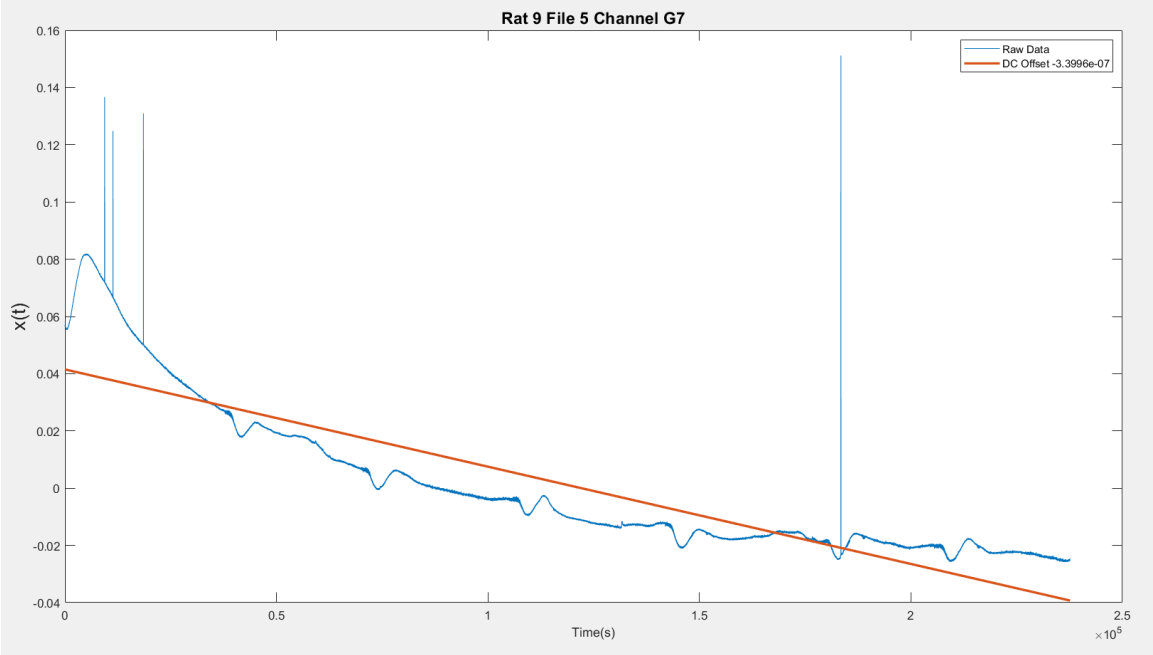


Figure 3.8: Best fit line is calculated and superimposed onto the raw data signal used to calculate said best fit line. The best fit line does capture a good image of what is going on overall but the nature of the best fit line fails to capture important key concepts such as the steep slope seen in the beginning of the recording before the first SD event

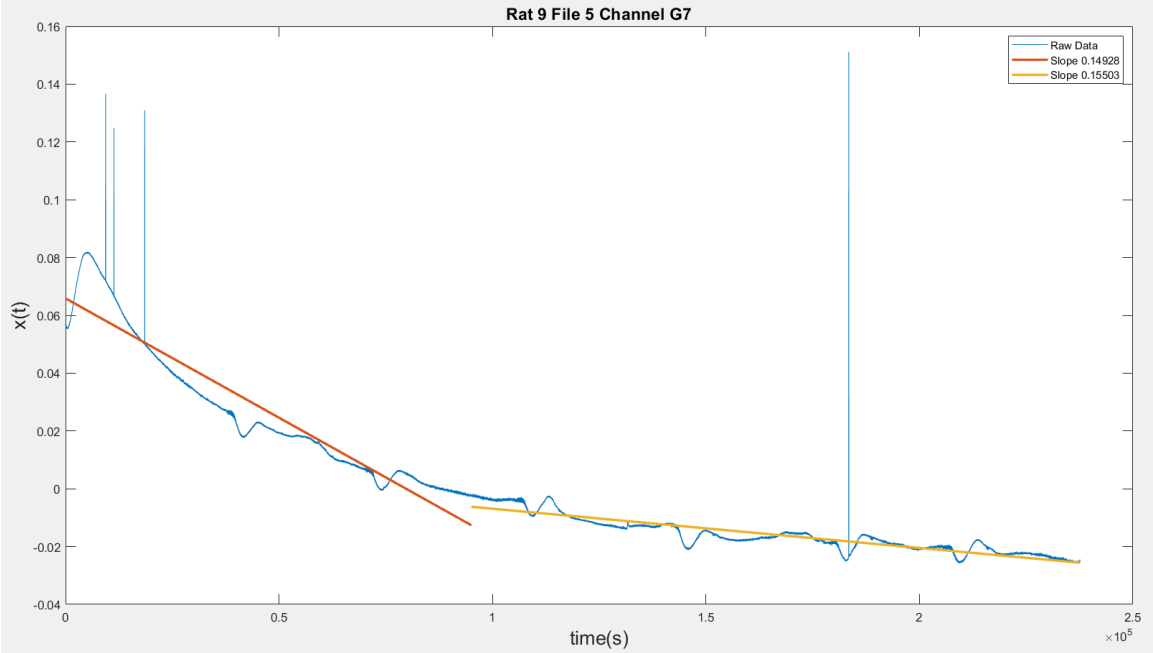


Figure 3.9: Piece wise best fit lines are calculated using split data files rather than the entire recording, the steep slope seen before the first SD event is now better represented it is logged in the first best fit line's slope.

In order to further improve the base line metric; we calculated best fit lines over individual SD events to do this we began by taking the data of record SD events. The duration of all record SD events in the data file are compared isolating the longest lasting event in the data file. A window is then created, using half the length of the longest event as a buffer to be used both before the detected event start marker, and the after the event end marker. This method creates a unique window to be used for each SD event. A simple linear best fit line will not work for this application as in this time scale the nature of the SD events, to spike in both positive and negative directions, will skew the best fit line. To fix this problem a moving average is calculated, using Eqn. 7, using a large

window in order to “flatten” the SD event while keeping the amplitude decrease brought by the depression event.

$$\frac{y(n)}{l} = \sum_{k=-w}^{w} x(k)h(n-k)$$

Equation 7

A basic convolution between x, the raw data signal, and h, an array of ones, outputs y; and by dividing l, the length of h, we get the moving average value for the specific point. The length of h can be adjusted by increasing the window we use more surrounding points to weight down the amplitude spikes caused by the SD events; a too large window however will result in a skewed average. A best fit line is then calculated using the moving average data points in the derived window of the SD event. This results in a trendline that gives us the change in amplitude caused by the SD event and associated depression as well as the current amplitude of the data signal at the time of the recorded SD event.

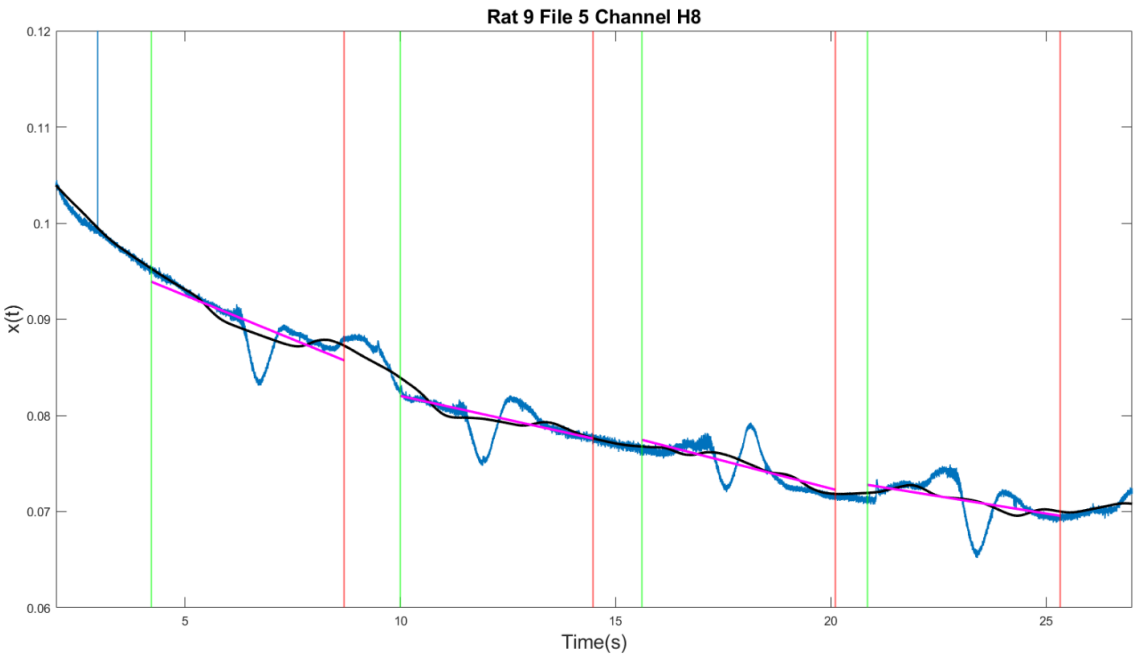


Figure 3.10: The blue data signal is the raw ECoG signal, the black line is the moving average resulting signal, the green and red vertical lines represent the considered window's start and stop times respectively, the magenta lines denote the derived best fit line of its respective SD event

CHAPTER 4

Clustering of ECoG Features by Tissue Health / State

After key electrocorticographical features were defined, the question remained whether these features have any significance regarding the brain tissue health. These features were derived, logged, and then fed into a clustering algorithm in order to segregate data channels in this new feature space. The methods used for these clustering algorithms as well as the results and what this means for our hypothesis is explored in this chapter.

4.1 K-Means Clustering

In the previous chapter we discussed key features taken from the raw ECoG data, average DC baseline, average slope, duration of the SD event, as well as the peak-to-peak amplitudes of the SD events, the average DC baseline of each SD event as well as the average slope of each SD event. We took these features and ran them through a k-means clustering algorithm. A k-means clustering is an unsupervised algorithm which transforms the ECoG waveforms into a set of features in a new multi-dimensional space. We can visualize the ECoG data by reducing the dimensionality to 2 or 3 using principal component analysis and plotting each channel at its feature coordinates. The k-means algorithm initialized the clusters with k randomly selected centroids. Each data point was then assigned to the nearest centroid; this is done by calculating the Euclidean distance between the current data point and each of the centroids, using Eqn. 6, the data point is assigned to the centroid that returns the smallest distance.

$$C = \sqrt{(A_2 - A_1)^2 + (B_2 - B_1)^2}$$

Equation 4.1

Once all points have been assigned, the center of the clustered points is calculated, the centroid is moved to this derived center and the algorithm is run again with the new centroid locations reassigning data points to better fitting centroids if needed. The algorithm is repeated a preset number of times, for our application we used ten total replicates to ensure the algorithm properly finishes clustering the data points.

4.2 Determining Number of Clusters

We need to determine the best value of k to use for this k-means algorithm. To do this we run the algorithm multiple times using different values of k. Each time the algorithm is run we calculate the distortion metric we do this by calculating a simple Euclidean distance metric, as shown in Eqn. 4.1, from the centroid to each of the clustered channels in pc space. The value of this distortion metric is taken and averaged out with the rest of the clusters in each iteration of the k-means algorithm; the final distortion metric value is then stored. We ran this loop for values of k from one to ten. The average distortion metric values are then plot onto graph to more easily visualize the data and we see an elbow develop around the values of three and four. Thus we use k equal to three for our algorithm.

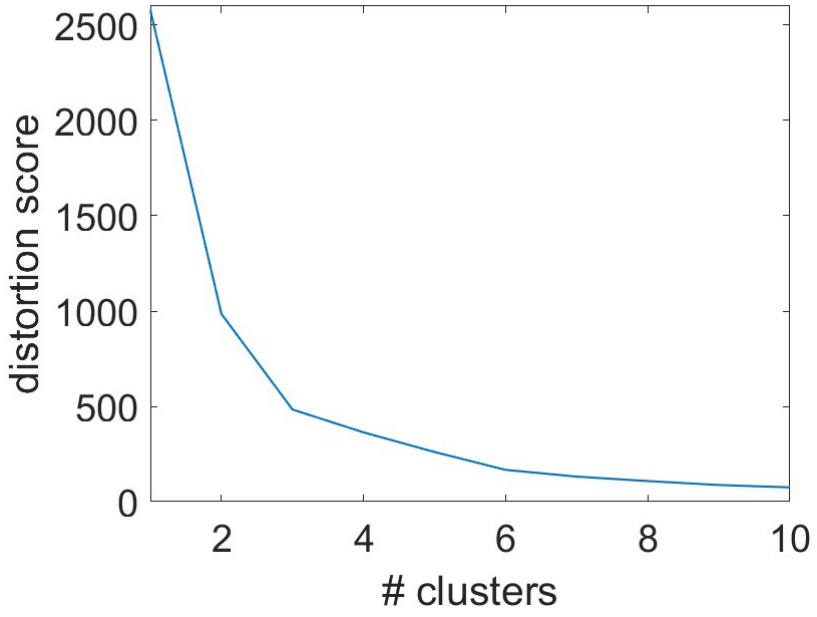


Figure 4.1: Distortion metric decreases as number of clusters *k* increases, but there are greatly diminishing returns after the "elbow" in the curve, which occurs at *k*=3.

4.3 Mapping clusters onto physical location

Using a value of 3 for our clustering algorithm took the output and mapped it onto a color coded 8 by 8 matrix in order to better visualize what data channels are being grouped together. The results are shown in the figures below. Additionally, the raw data was then also color coded and again in order to better visualize how the data was being grouped together.

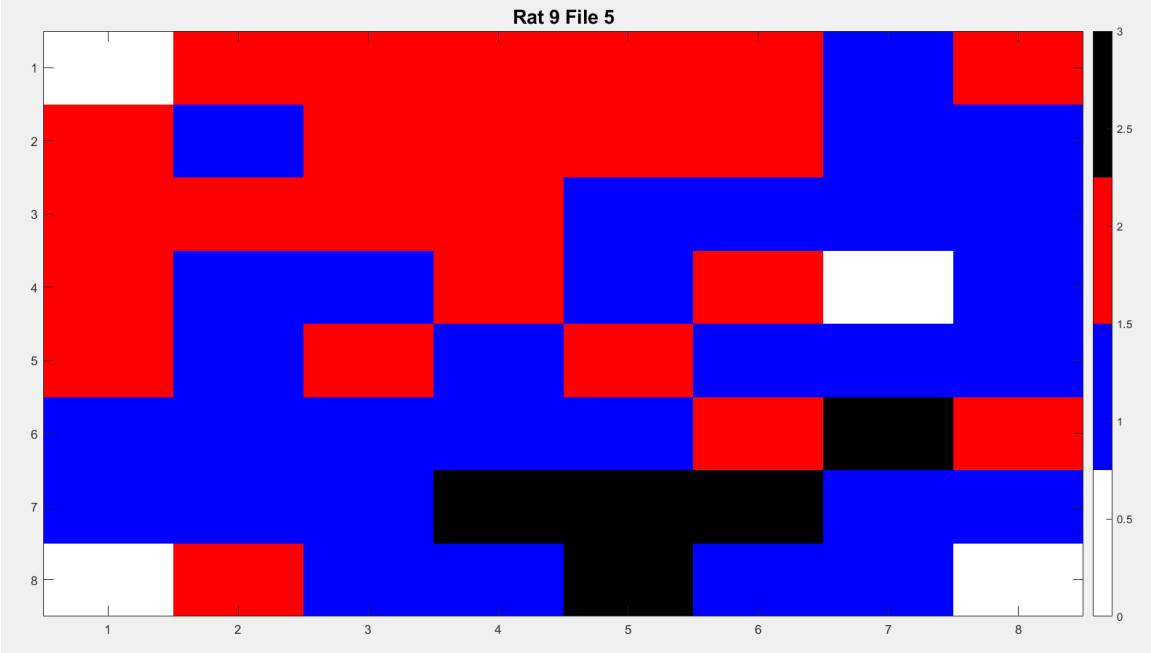


Figure 4.2: Results of the clustering algorithm for rat 9 file 5 plotted in space and color coded, unused channels are denoted with white space. The figure shows promising results as the segregated channels form contiguous regions.

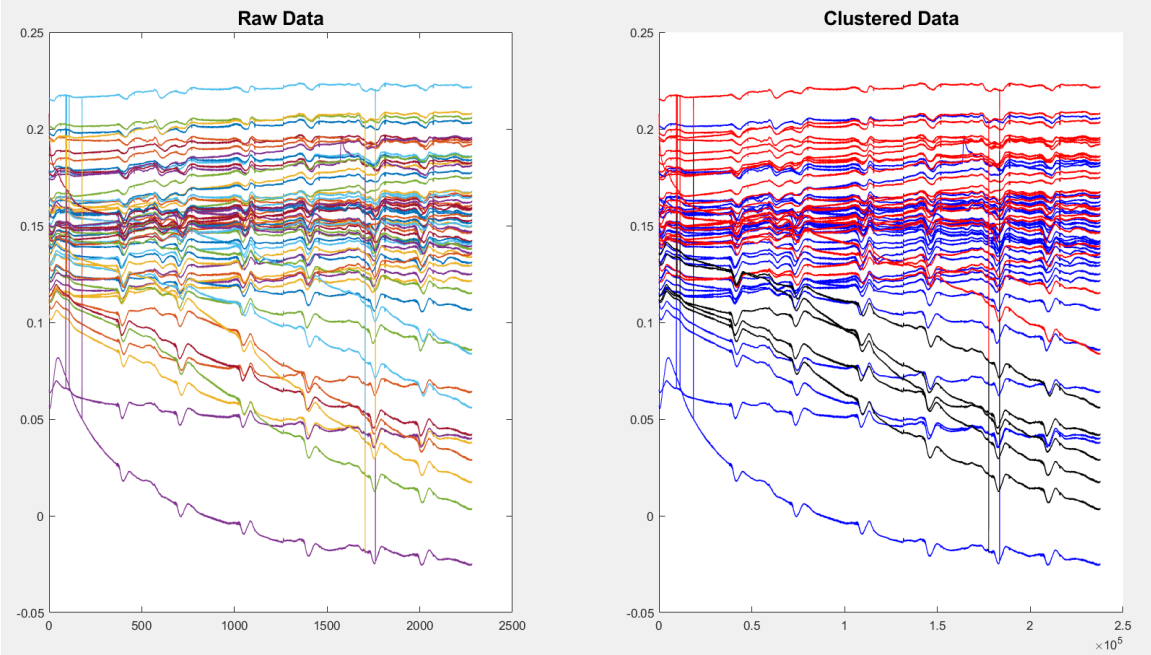


Figure 4.3: Figures depicts the result of the clustering algorithm results for rat subject 9 file 5. Left is the raw ECoG data, right is the same data channels color coded based on the clusters.

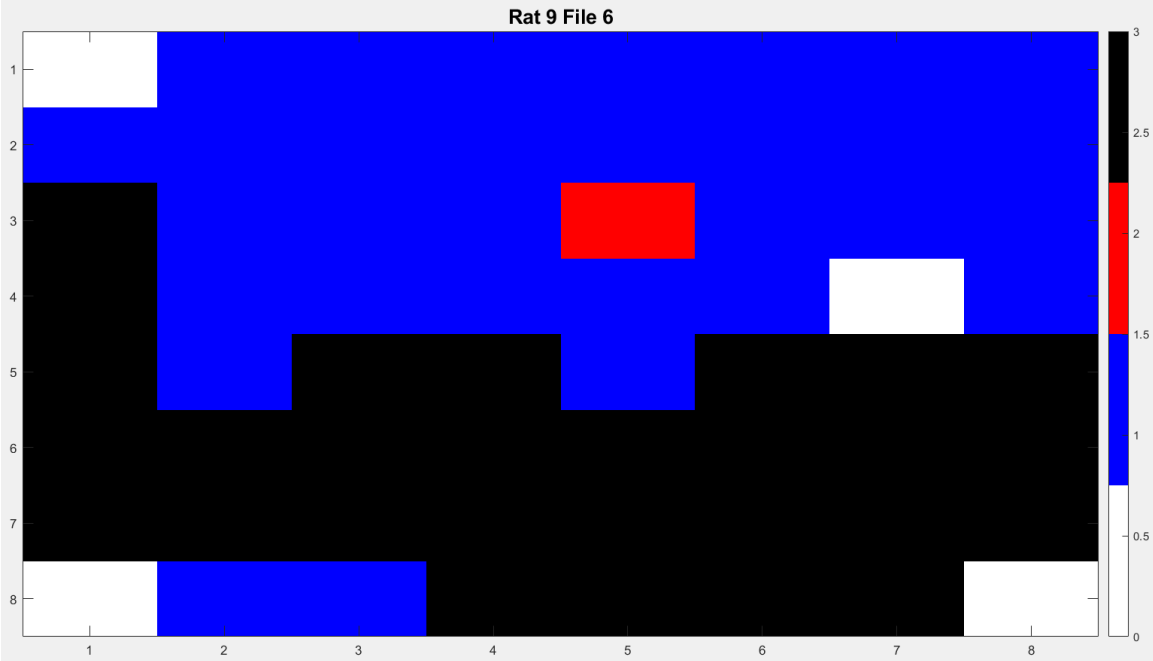


Figure 4.4: Results of the clustering algorithm for rat 9 file 6 plotted in space and color coded, unused channels are denoted with white space

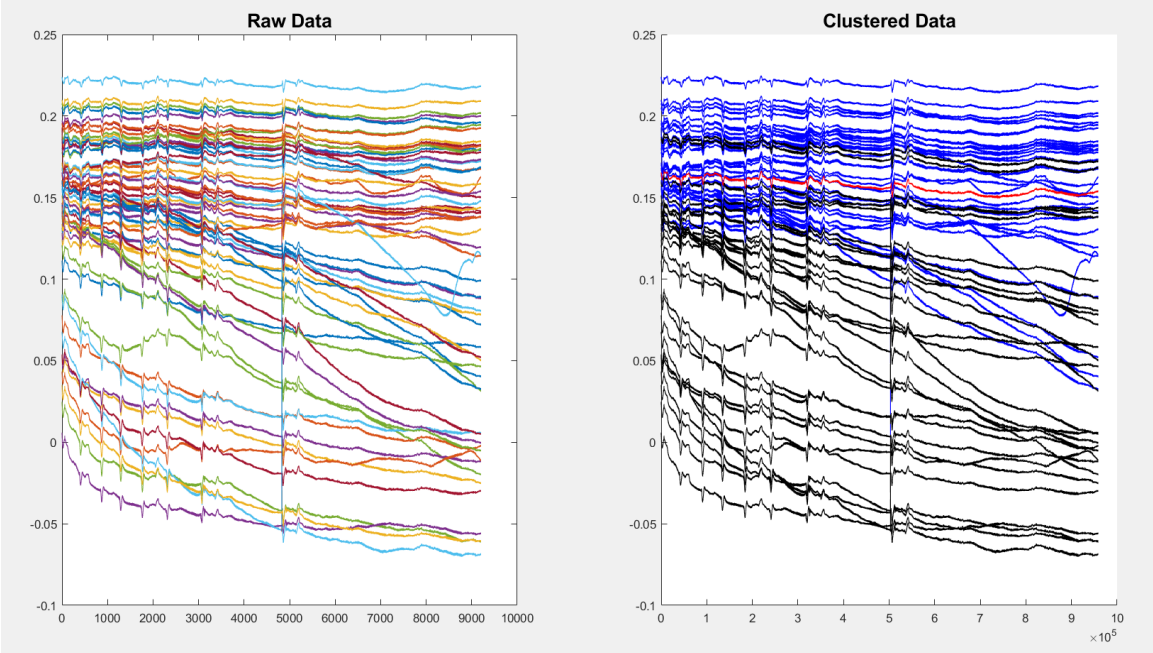


Figure 4.5: Figures depicts the result of the clustering algorithm results for rat subject 9 file 6. Left is the raw ECoG data, right is the same data channels color coded based on the clusters.

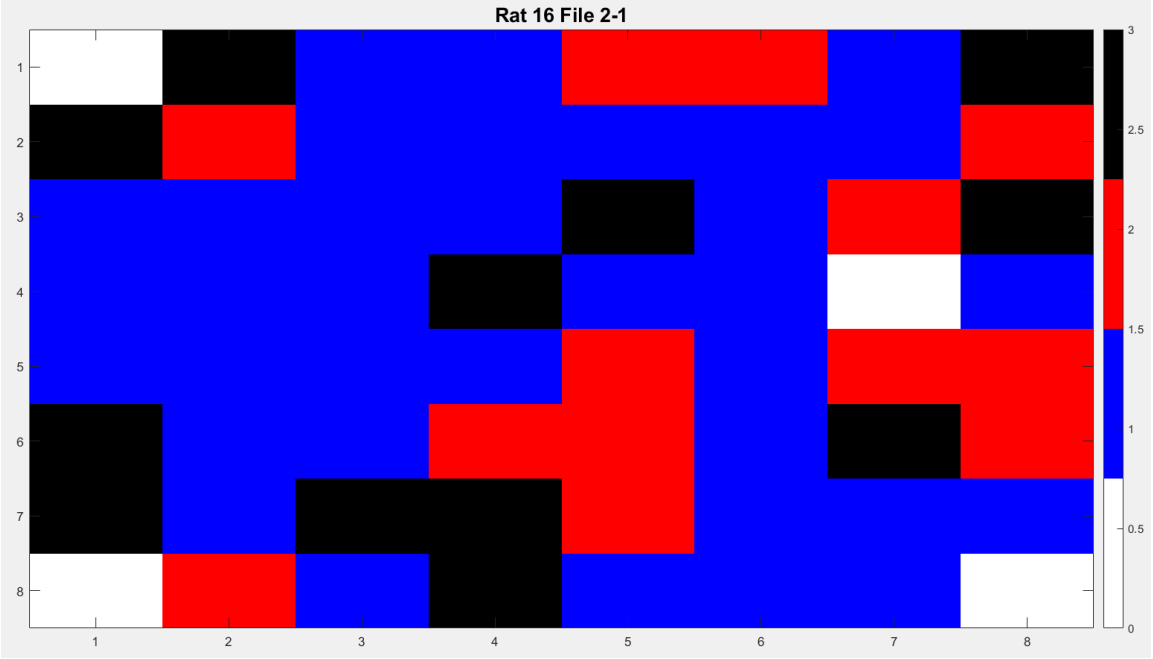


Figure 4.6: Results of the clustering algorithm for rat 16 split file 2 - 1 plotted in space and color coded, unused channels are denoted with white space

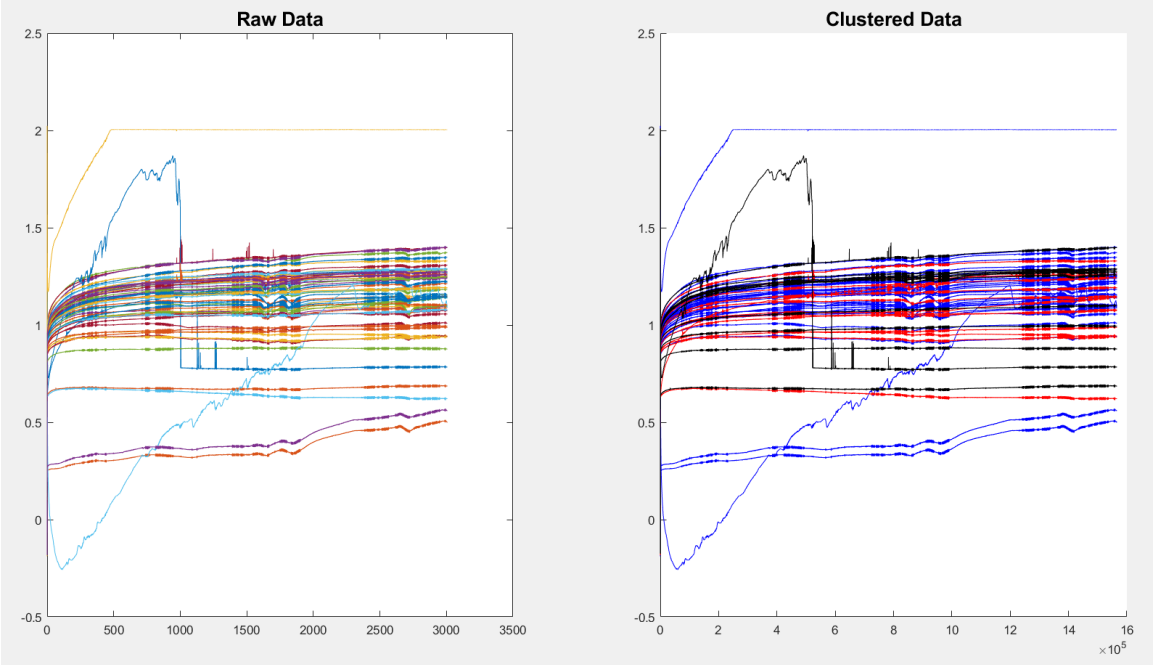


Figure 4.7: Figures depicts the result of the clustering algorithm results for rat subject 16 split file 2 - 1. Left is the raw ECoG data, right is the same data channels color coded based on the clusters.

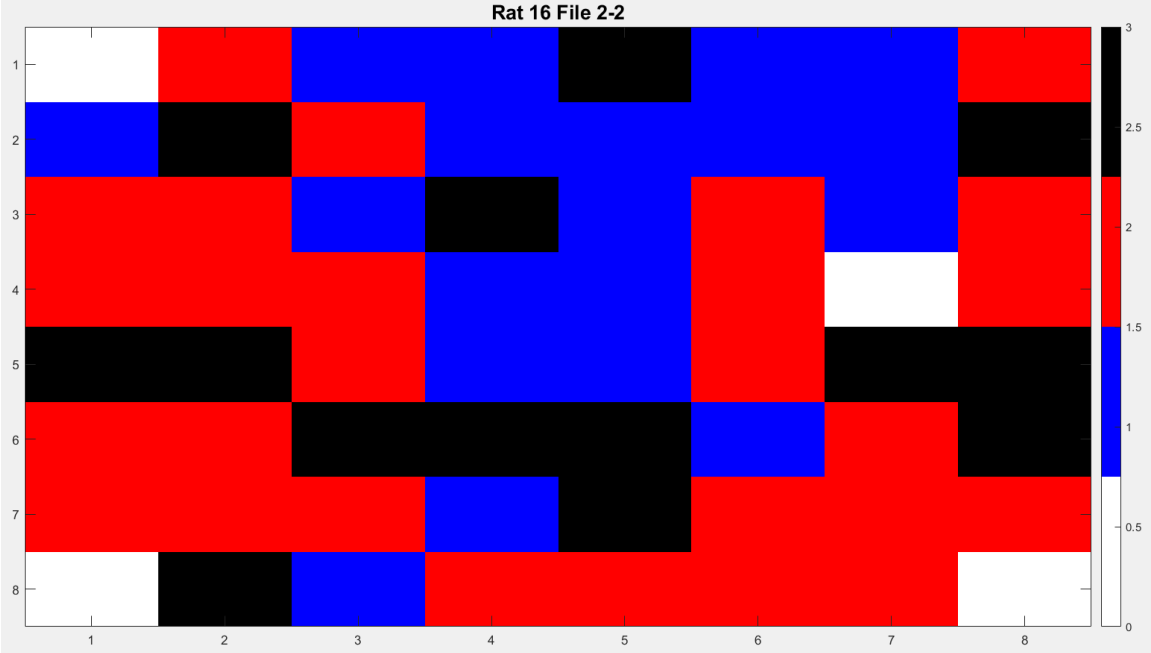


Figure 4.8: Results of the clustering algorithm for rat 16 split file 2 - 2 plotted in space and color coded, unused channels are denoted with white space

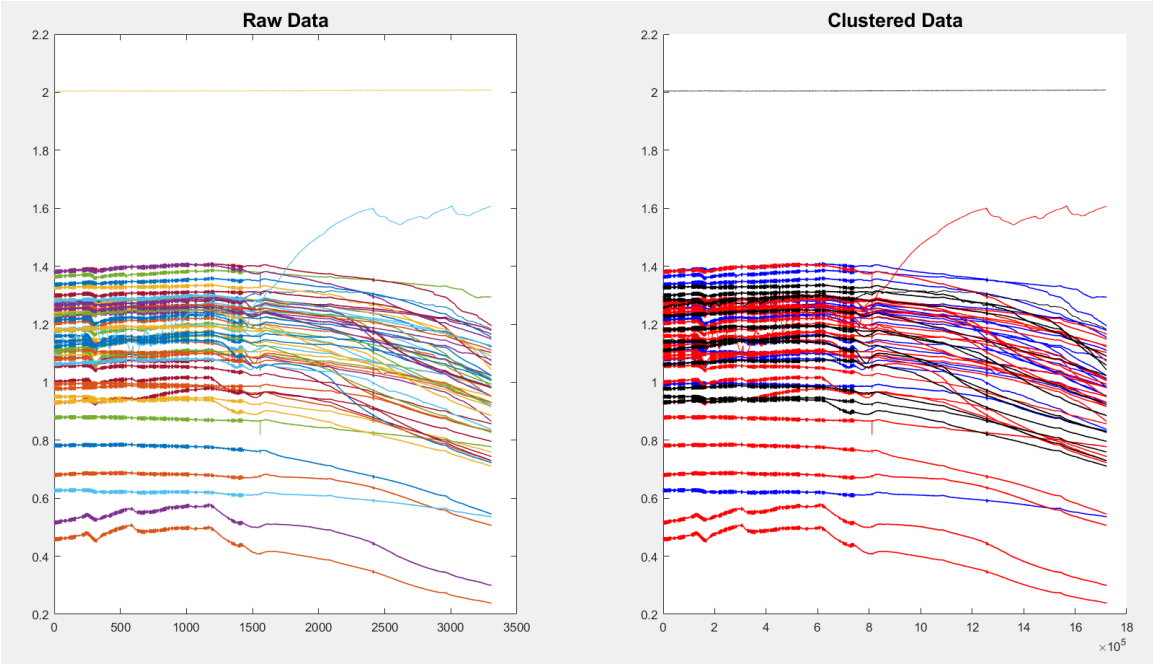


Figure 4.9: Figures depicts the result of the clustering algorithm results for rat subject 16 split file 2 - 2. Left is the raw ECoG data, right is the same data channels color coded based on the clusters.

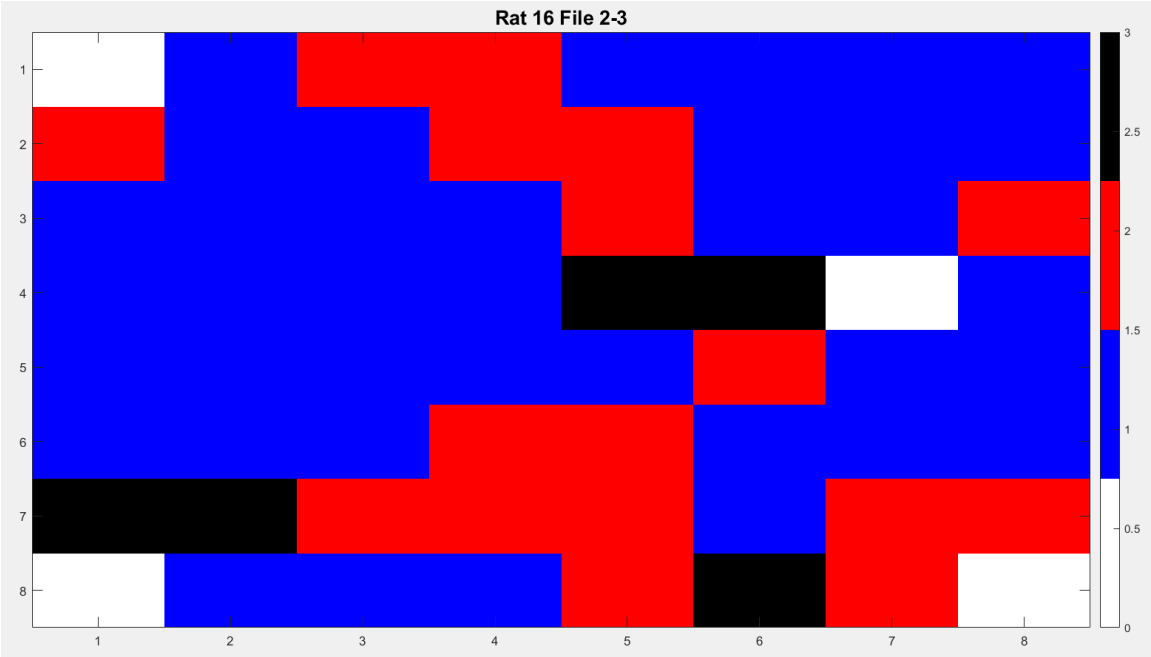


Figure 4.10: Results of the clustering algorithm for rat 16 split file 2 - 2 plotted in space and color coded, unused channels are denoted with white space

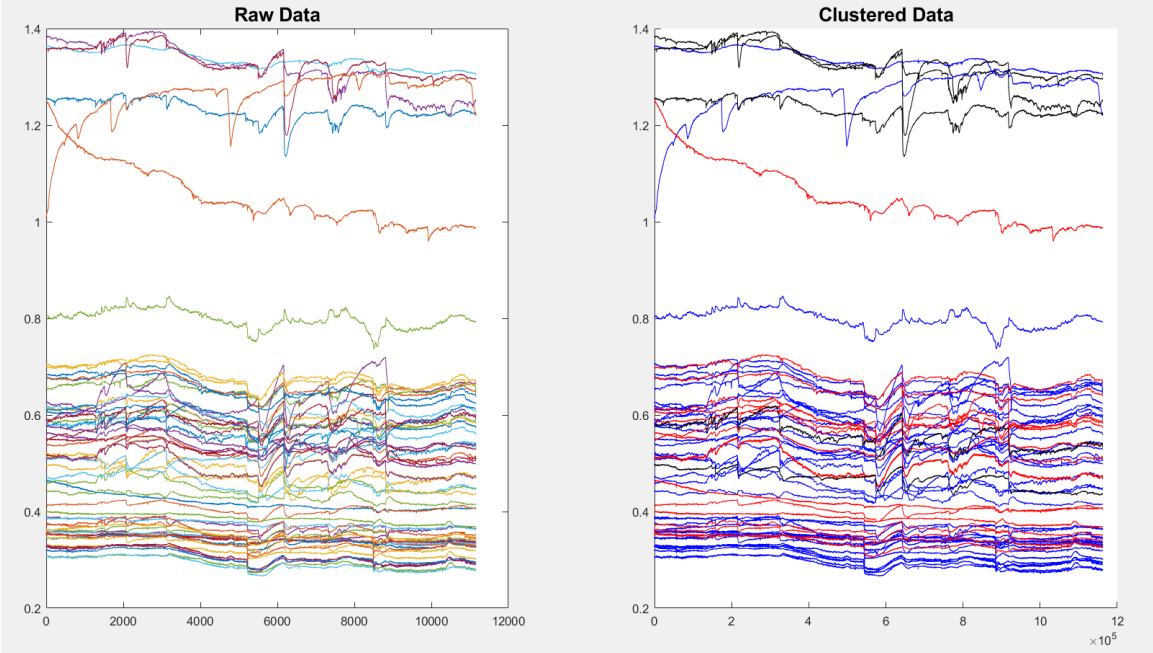


Figure 4.11: Figures depicts the result of the clustering algorithm results for rat subject 16 split file 2 - 3. Left is the raw ECoG data, right is the same data channels color coded based on the clusters.

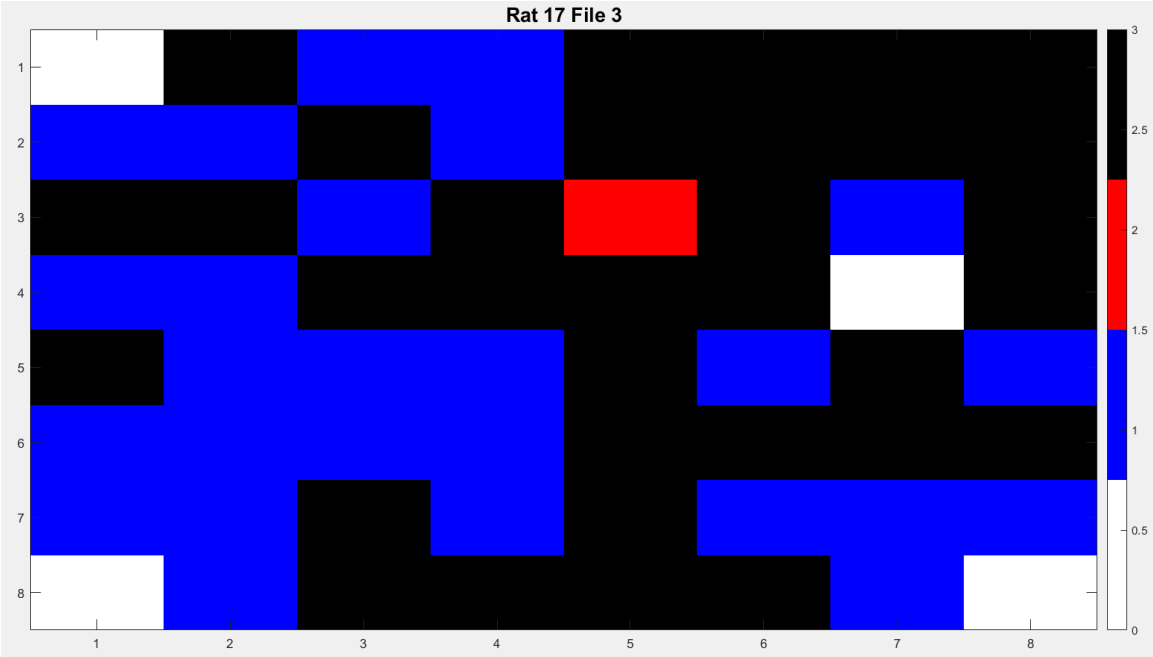


Figure 4.12: Results of the clustering algorithm for rat 17 file 3 plotted in space and color coded, unused channels are denoted with white space

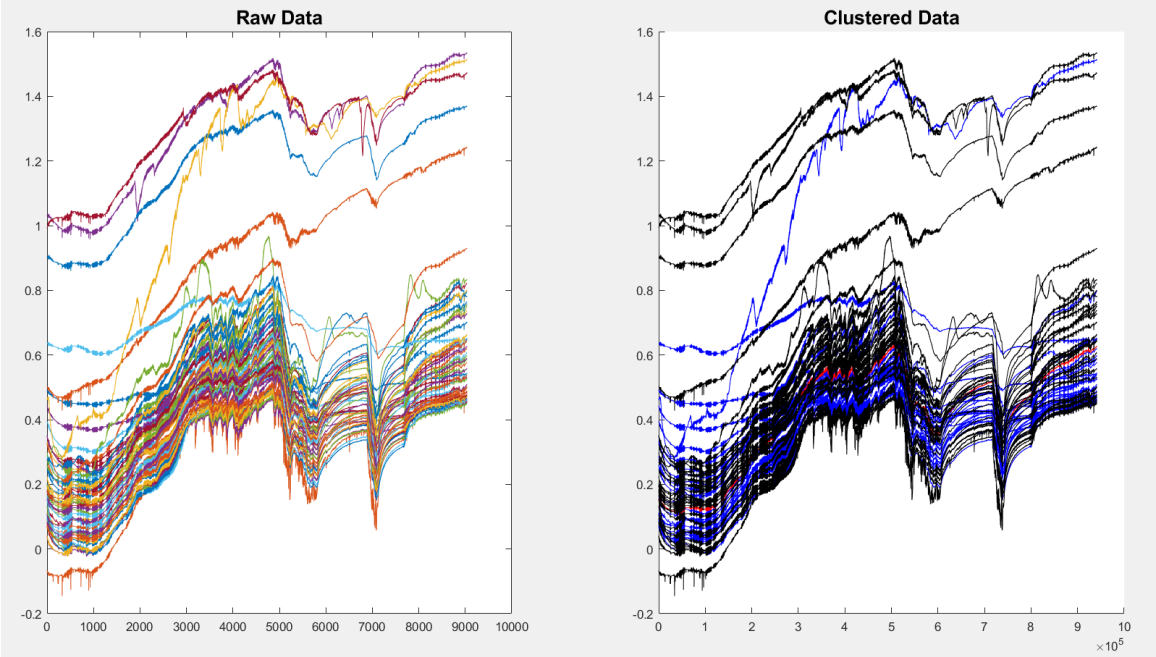


Figure 4.13: The result of the clustering algorithm results for rat subject 17 file 3. Left is the raw ECoG data, right is the same data channels color coded based on the clusters.

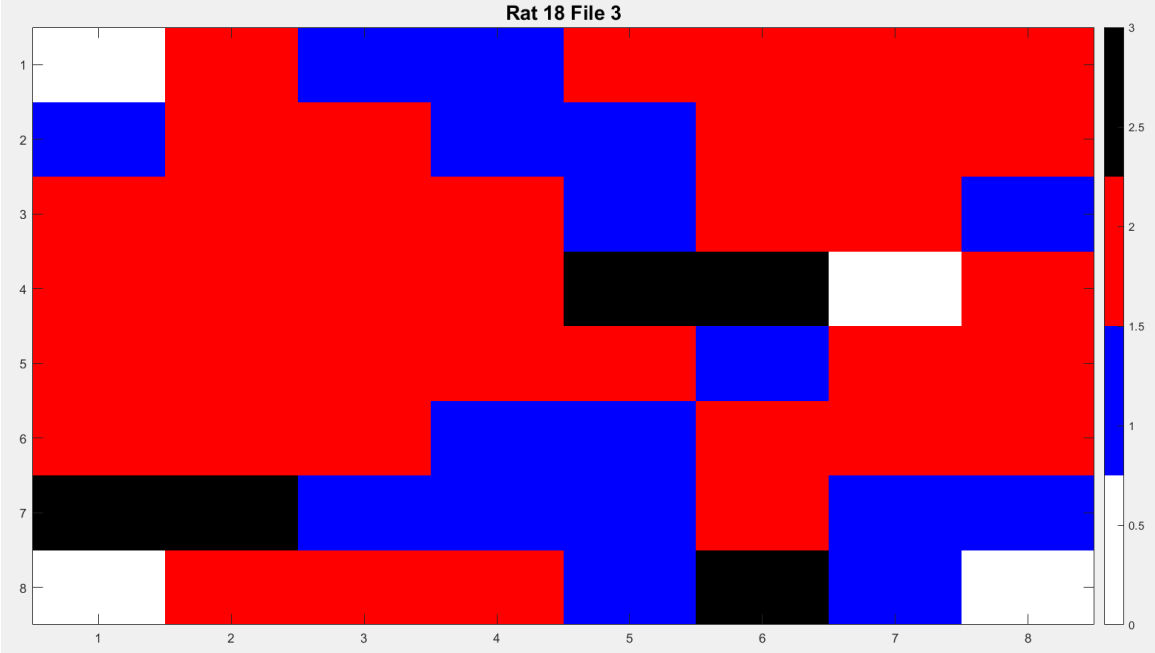


Figure 4.14: Results of the clustering algorithm for rat 18 file 3 plotted in space and color coded, unused channels are denoted with white space

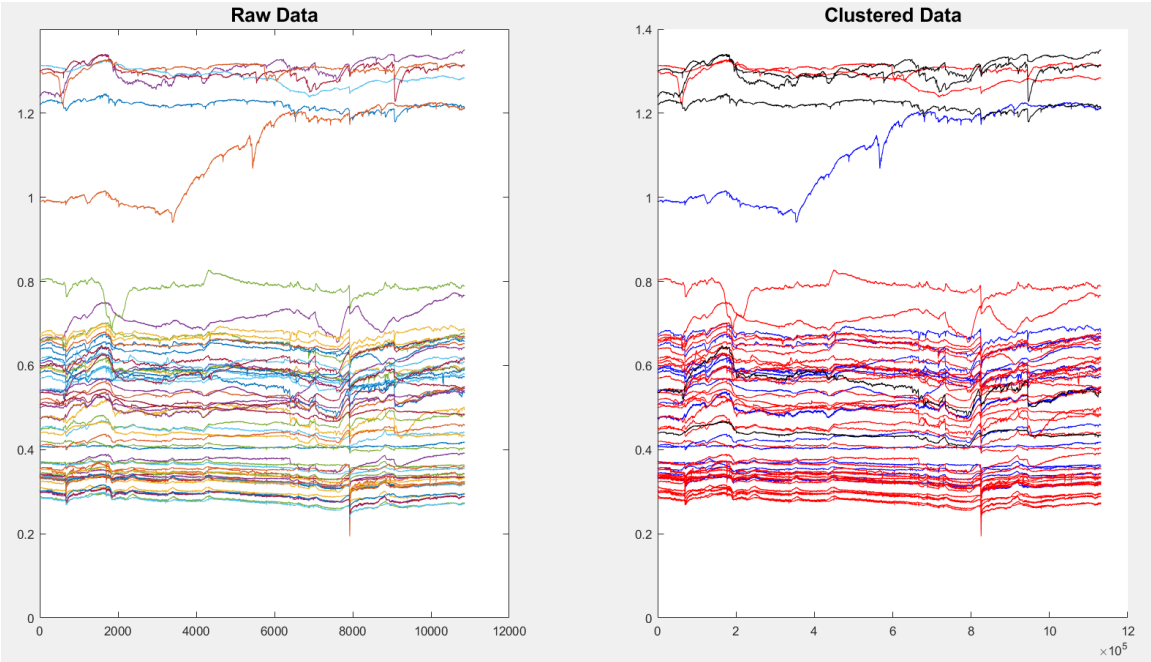


Figure 4.15: The result of the clustering algorithm results for rat subject 18 split file 3. Left is the raw ECoG data, right is the same data channels color coded based on the clusters.

The results are clustering algorithm depicted on Figures 4.2 – 4.15 show mixed results. We have results such as rat 9 where the ECoG data shows very prominent features that make is easy to differentiate tissue both extract and differentiate from one another. This outputs promising results we can see the color coded clustered special map, even when the algorithm had no information as to where the channels lie in space the algorithm segregated the channels into contiguous regions. These results lead us to believe the algorithm with some degree of error can successfully identify tissue health using only the derived ECoG features.

In comparison, data sets like rat 16 do not show as clearly defined borders nor do the channels that were classified together according to the ECoG features cluster in as contiguous a region as with Rat 9. Furthermore, even when looking at the color coded raw ECoG data, there do not appear to be distinguishable features within clusters that set them apart from the other two clusters. The results might indicate that tissue damage does not necessarily spread in radially concentric contiguous regions and we cannot determine tissue health from the visual characteristics of the ECoG features. On the other hand it is possible that we simply are not able determine the status of tissue health using the features we selected and extracted from the ECoG data.

The results from rat 9 lead us to believe it is possible to classify tissue health with the given information. The results for rat 16 can be attributed to the data files themselves being less than ideal data. In rat 16 data file 2 -1 for example, apart from the few data channels that seem to be bad recordings, the data channels seem to follow the same overall shape they remain at a similar offset level with little variance between them, no

group of data display collective change in amplitude. Additionally, we know there is a silencing component to SD events that we have not yet fully explored. From what we have observed silencing events do not always have a prominent SD event preceding so in data sets like rat 16 where amplitude across the data channels remains consistent and SD events are not as prominent, using silencing as a metric to classify the data may improve our results. This can be said about the rest of the data files for rats 17 and 18 the algorithm determined there are no outstanding collection of data channels from the rest.

One improvement to be made to the detection algorithm would be to incorporate silencing as a key feature. While our algorithm shows promise for the feasibility of using ECoG features from the epidural ECoG arrays to determine which areas of the penumbra have the greatest likelihood for reversal of damages and which are most at risk for impending necrosis, a major limitation of our current methods include the fact that the algorithm does not define which cluster is are classified as healthy, compromised, and damaged. Thus, we must rely to some extent on expert opinion from immunohistochemical analysis along with other clinical data to use as ground truth in validating our models.

4.4 Relating clusters defined by electrophysiological features to tissue health

With the results of the clustering algorithm, depicted in the images below we next need to validate our model. This can be done in several different ways, unfortunately the only results of the subjects' immunohistochemistry are the displayed tissue stains analysis postmortem. The brain is sampled coronal slices and staining said slices to depict the area of the infarct. To show how our results compares to the immunohistochemistry we have

available to us the resulting heatmap representing what the algorithm classifies as healthy, compromised, and damaged tissue; superimposed into the stained tissue sample.

The results of rat nine are mostly resemble what we were hoping to see as the outcome the clustering algorithm. The algorithm has no information as to where the data channels lie in space, regardless, when the segregated data channel are plotted in space the three clusters form contiguous regions. What is even more interesting is the fact that the results of the clustering algorithm seem to closely resemble the final immunohistochemistry where the northeast side of the array reflects the damaged tissue, blue, and southwest side of the array depicts the compromised tissue, black.

We expected rat nine to perform the best as its recorded ECoG data contains very prominent features. Aside from very prominent features rat nine also had total life span of only one day, which may have aided in a more accurate diagnosis of tissue health as the postmortem tissue health would be more in line with any data file as the recordings are within 24 hours of one another.

Conversely the rats which lived longer the final tissue diagnosis is significantly different from what the algorithm determines had much less consistent ECoG features and more difficult-to-define SD waveforms. The discrepancy could be caused by the time difference between the time of recording and the time taken until postmortem when the tissue was analyzed.

In addition to a complete immunohistochemistry, we are currently working on perfecting the detecting algorithm which would allow us to study more data sets. Currently we are relying on data sets with marked data as ground truth to where these SD

events occur start and stop and what should be discounted as noise. Once we have the necessary data and algorithm we expect to create a more detailed history, a form of time line depicting what the algorithm determines is happening to tissue health over time across files within data sets.

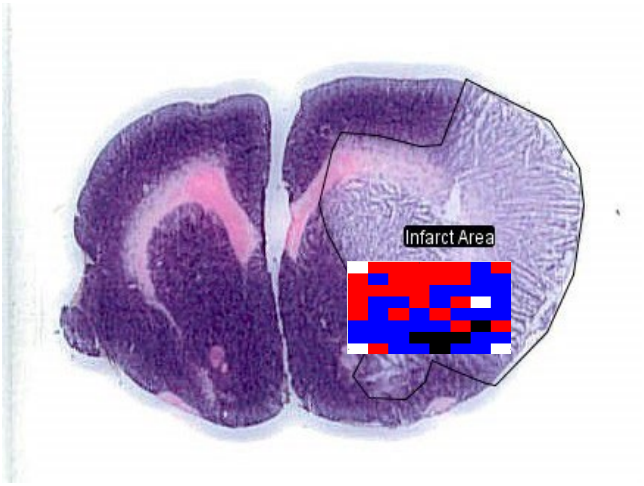


Figure 4.16: The immunohistochemistry results rat 9 with rat nine file 5 overlayed to show the results of the clustering algorithm. Even though there are contiguous regions that suggest successful tissue health classification these staining results are the only form of immunohistochemistry we have to the accuracy of our model can be debated.



Figure 4.17: The immunohistochemistry results rat 16 with rat 16 file 2 - 1 overlayed to show the results of the clustering algorithm.

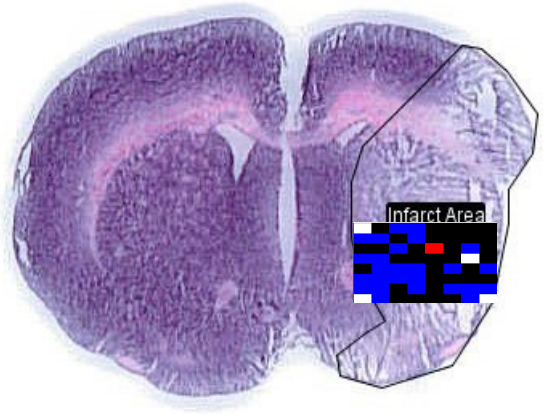


Figure 4.18: The immunohistochemistry results rat 17 with rat 17 file 3 overlayed to show the results of the clustering algorithm.

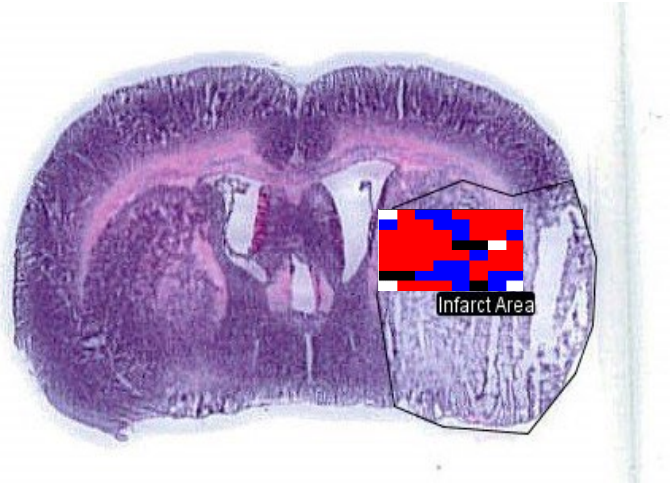


Figure 4.19: The immunohistochemistry results rat 18 with rat 18 file 3 overlayed to show the results of the clustering algorithm.

CHAPTER 5

Future Direction

While much research over the past decades has been devoted to the pathogenesis of stroke and acute care to increase the probability of survival, we now have the technology to monitor the electrophysiological changes that occur across the brain over time and use the new understanding we have gained to target the appropriate areas of the brain with the most likelihood for reversal of ischemic injury and closed-loop control of novel stroke treatment such as neural stimulation. The analysis we conducted in this thesis work using data acquired from chronic micro-recording technology serves as a pioneering contribution toward development of such closed-loop stroke therapy.

Dr. Jonathan Viventi and his research team at Duke University sought to develop a state of the art custom electrode array. Its small size means it can be implanted onto the patient's brain surface with less invasive surgery, while its high electrode density allows it to record high resolution surface brain ECoG electrical activity. Dr. Ulrike Hoffman and her research team then used these electrode arrays to record ECoG data of alive unanesthetized rats while suffering from an induced stroke. They team at Duke allowed us access to the raw ECoG data with the goal of developing an algorithm capable of analyzing key features in the ECoG data and ultimately diagnose the state of tissue health in real time. The results of this study shows promise unfortunately, we were unable to perfect the detection algorithm and are missing key components of the subjects' immunohistochemistry to completely validate our model.

Adjusting the detection algorithm should take priority in the terms of short term future direction. Once the detection algorithm is perfected more sets of data are available to be used in feature analysis algorithm. Perfecting the detection algorithm includes establishing a clear definition of when the start and end times of these SD events are. Additionally, we know silencing is a major factor in long term brain damage that has yet to be explored in this study. Silencing might prove to be a challenge to detect, since unlike depolarization that are induced by SDs, silencing can occur in ECoG brain activity without the need for an SD event to precede it. Ideally with these newly derived features our model should be able to more accurately predict tissue health, as well be able to analyze data sets that are not currently a viable option. Running the clustering algorithm with this new data we should be able to see the progression of tissue health in the recorded brain tissue.

In the long term we hope our research leads to a more customized patient-specific treatment through the use of *in vivo* tissue health diagnosis. Furthermore, the use of ECoG data from the newly developed epidural electrode arrays will reduce the use of invasive surgery to monitor brain tissue health *in vivo*. Closed-loop treatment therapies can be developed without the need for highly invasive surgery and minimize the brain damage and associated morbidity otherwise resulting from stroke.

Bibliography

[1] S. L. Murphy, B.S., K. D. Kochanek, J. Xu, M.D. and E. Arias, Ph.D., "Mortality in the United State 2021," U.S. Department of Health and Human Services, 2021.

[2] B. Nazario, MD, "Types of Stroke," WebMD, 13 8 2020. [Online]. Available: https://www.webmd.com/stroke/guide/types-stroke. [Accessed 12 8 2022].

[3] "How many people are affected by/at risk for stroke?," https://www.nichd.nih.gov/, 2022. [Online]. Available: https://www.nichd.nih.gov/health/topics/stroke/conditioninfo/risk. [Accessed 12 8 2022].

[4] "About Stroke," National Center for Chronic Disease Prevention and Health Promotion, 4 May 2022. [Online]. Available: https://www.cdc.gov/stroke/about.htm?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fstroke%2Ftypes_of_stroke.htm. [Accessed 12 8 2022].

[5] "Ischemic Stroke Treatment," American Stroke Association, 5 12 2018. [Online]. Available: https://www.stroke.org/en/about-stroke/types-of-stroke/ischemic-stroke-clots/ischemic-stroke-treatment. [Accessed 12 8 2022].

[6] "Stroke Rehabilitation: What to Expect as you Recover," Mayo Clinic, 18 5 2022. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/stroke/in-depth/stroke-rehabilitation/art-20045172. [Accessed 12 8 2022].

[7] T. N. Jilani and A. H. Siddiqqui, Tissue Plasminogen Activator, Treasure Island (Fl): StatPearls Publishing, 2022.

[8] T. S. Metkus, MD and D. Zieve MD, "Throbolitic Therapy," U.S. National Library of Medicine, 7 7 2020. [Online]. Available: https://medlineplus.gov/ency/article/007089.htm. [Accessed 12 8 2022].

[9] M. M. S. Prabhakaran, M. I. Ruff and M. P. R. A. Bernstein, "Acute Stroke Intervention," Department of Neurology, Northwestern University Feinberg School of Medicine, Chicago, Illinois, 2015.

[10] S. P. Kloska, "Acute stroke magnetic resonance imaging: Current status and future perspective," 2009.

[11] Y. C. Tas, I. Solaroğlu and Y. Gürsoy-Özdemir, "Spreading Depolarization Waves in Neurological Diseases: A Short Review about its Pathophysiology and Clinical Relevance," Curr Neuropharmacol, 2019.

[12] F. M. A. C. e. a. Dreier JP, "Recording, analysis, and interpretation of spreading depolarizations in neurointensive care: Review and recommendations of the COSBID research group," J Cereb Blood Flow Metab. , 2017.

[13] J. P. Drier, "The role of spreading depression, spreading depolarization and spreading ischemia in neurological disease," *Nature Medicine,* vol. 17, no. 4, pp. 439-447, 2011.

[14] B. K. I. S. P. W. Tomas Peisker, "Acute stroke therapy: A review,," *Trends in Cardiovascular Medicine,* vol. 27, no. 1, pp. 59-66, 2017.

[15] G. S. Paul M. George, "Novel Stroke Therapeutics: Unraveling Stroke Pathophysiology and Its Impact on Clinical Treatments," *Neuron,* vol. 87, no. 2, pp. 297-309, 2015.

[16] V. W. e. al, "A low-cost, 61-channel µECoG array for use in rodents," *7th International IEEE/EMBS Conference on Neural Engineering,* pp. 573-576, 2015 .

[17] K. M. T. e. a. Palopoli-Trojani, "High-density cortical µECoG arrays can concurrently track peri-infarct depolarizations and long-term evolution of stroke in awake rats.," *Communications Biology Submitted.,* 2021.

[18] T. M. B. B. P.-T. K. C. C. W. C. Y. C. I. M. F. R. V. J. Woods V, "Long-term recording reliability of liquid crystal polymer µECoG arrays," *Neural Eng. ,* 2018 .

[19] A. Abelian, "Automated detection of Spreading Depolarizatipns for Analysis of Stroke Progression," 2022.

[20] M. e. a. Gabor C. Petzold, "Nitric Oxide Modulates Spreading Depolarization Threshold," *Stroke,* p. 1292–1299, 2008.

APPENDIX

MATLAB Script Files

A.1 Main

```
close all
clc
clear

load('C:\Users\axelo\OneDrive\Desktop\SD Project\Raw Data\Rat
17_10012019_Original\Rat 18_10092019_003_Original.mat');

LUT = [61  7 36  6 35  3  1 62 ...  %LUT refrences where specific
channels
      39 10 40  9 38  4 31  2 ...  %are in space on the electrode
array
      12 42 13 41 11 34 33 32 ...
      44 15 45 14 43  8 37  5 ...
      59 30 60 29 58 24 52 49 ...
      27 57 28 56 26 20 19 18 ...
      54 25 55 63 53 48 17 46 ...
      23 51 22 50 21 47 16 64];
lutmx = reshape(LUT,8,8)            %reshap LUT into 8x8 matrix
```

```
data(:,1)=[];
t = (0:length(data)-1)/Fs;

figure(100)
subplot(1,2,1)
plot(t,data);
title('Raw Data','FontSize',16);

%% Remove DC
iCh = 1:60;
NormData = remove_DC(iCh, data, Fs);
figure(99);

%% 8x8 Array xcorr
ref_ch = 26;
center = 0;
step = 20000;
RxyAmpLag = corrplots(ref_ch, NormData, lutmx, center, step);
figure(98);

%% Create Heatmaps
%if you want individual figures for each heatmap set figs to 1
```

```
figs=0;
%how many heat maps total do you want
numheat=9;
%number of points to consider for the high clusters
numpoints=8;
[RxyAmpConv] = corrheatmap(ref_ch, RxyAmpLag, ...
    center, step, figs, numheat, numpoints);
figure(97)

%% Moving Average
%if you want so see data channels set figs to 1
Sel_Chan = 1:60;
figs = 0;
[Weights, xTimes, yTimes] = Moving_Average(Sel_Chan, Fs, t, figs,
data);

%% Calculate Overall Trendline
split_data=0.40;
[a, N]=size(data);
trainset=round(N*split_data);
testset=N-trainset;

X=[];
```

```
W=[];
Z=[];
MSE=[];

time = 1:N;
first_data=data(:,1:trainset);
f_t=1:trainset;
second_data=data(:,(trainset+1):N);
s_t=(trainset+1):N;

for idx = 1:60
%     figure
%     plot(data(idx,:));

    x = polyfit(time,data(idx,:),1);
    X = [X; x];                          % Weights will be used as
input parametes for the K-Means clustering
    adj_data = [time; ones(1, N)];
    yhat = x*adj_data;
%     hold on
%     plot(time,yhat,'linewidth',2);

    w = polyfit(f_t,first_data(idx,:),1);
```

```
    W = [W; w];                          % Weights will be used as
input parametes for the K-Means clustering
    adj_data = [f_t; ones(1, trainset)];
    yhat = w*adj_data;
%     hold on
%     plot(f_t,yhat,'linewidth',2);

    z = polyfit(s_t,second_data(idx,:),1);
    Z = [Z; z];                          % Weights will be used as
input parametes for the K-Means clustering
    adj_data = [s_t; ones(1,length(s_t))];
    yhat = z*adj_data;
%     hold on
%     plot(s_t,yhat,'linewidth',2);
%     legend('Raw Data', ['Slope ' num2str(w(2))], ['Slope '
num2str(z)]);
end

%% Input Andrea's Data
Andreadata = [xTimes yTimes];
PCA_Data = [X(:,1) W(:,1) Z(:,1) X(:,2) W(:,2) Z(:,2)]';
idx=1;
for channel=1:60
```

```
    first_event = find(yTimes==channel)+1;
    last_event = find(yTimes==channel+1)-1;
    events=last_event-first_event;
    last1=7+events;
    first1=last1+1;
    last2=first1+events;
    PCA_Data(7:last1,idx)=Andreadata(first_event:last_event,1);
%Add duration of SDs

PCA_Data(first1:last2,idx)=Andreadata(first_event:last_event,2);
%Add amplitude of SDs
    idx=idx+1;
end

%% Run Clustering Given the Used Parameters
figure;
[coeff, score, latent, tsquared] = pca(PCA_Data');
plot(score(:,1), score(:,2), 'c.'); hold on;
Distor_Meas = [];
for nclust = 3
    Distortion = [];
    [IDX, C] = kmeans(PCA_Data', nclust, 'Replicates', 10);
```

```
    Dist = [];
    M = [];
    Index = [];
    figure;
    pt_markers = ['rx'; 'bx'; 'kx'; 'gx'; 'mx'; 'bo'; 'ko'; 'g.';
'r.'; 'b.'; 'k.'; 'g.';];

    for iclust = 1:nclust
        hold on;
        plot(score(find(IDX==iclust), 1),
score(find(IDX==iclust), 2), pt_markers(iclust,:), 'LineWidth',
1);
        title([num2str(nclust) ' Clusters Used']);
        curr_score_1 = score(find(IDX==iclust), 1);
        curr_score_2 = score(find(IDX==iclust), 2);
        Index = [Index length(curr_score_1)];
        n = size(score, 1)
        d = (sum((score - C(iclust, :)).^2, 2))
        Distortion = [Distortion mean(d)]
    end
    %Calculate average distances and take the lowest, this is the
"center point"
     Distortion = mean(Distortion)
```

```
     Distor_Meas = [Distor_Meas Distortion];
    %Overall distortion measure is the average of the cluster's
"packing"
    Distor_Meas = [Distor_Meas mean(Distortion)];

%% Plot Clustering Results
adj_LUT = LUT;

figure(100)
% subplot(1,2,1);
% plot(data);
% title('Raw Data');
hold on
subplot(1,2,2);
title('Clustered Data','Fontsize', 16);

Clust_1 = W(find(IDX==1),:);
[a, b] = size(Clust_1)
for adx = 1:a
    ich = find(W == Clust_1(adx));
    hold on;
    plot(data(ich,:),'b');
    idx = find(LUT == ich);
```

```
    adj_LUT(idx) = 1;
end
%title('IDX = 1');

%figure
Clust_2 = W(find(IDX==2),:);
[a, b] = size(Clust_2)
for adx = 1:a;
    ich = find(W == Clust_2(adx));
    hold on;
    plot(data(ich,:),'r');
    idx = find(LUT == ich);
    adj_LUT(idx) = 2;
end
% title('IDX = 2');

%figure
Clust_3 = W(find(IDX==3),:);
[a, b] = size(Clust_3)
for adx = 1:a;
    ich = find(W == Clust_3(adx));
    hold on;
    plot(data(ich,:),'k');
```

```
    idx = find(LUT == ich);
    adj_LUT(idx) = 3;
end
%title('IDX = 3');

for i=61:64
    idx = find(LUT == i);
    adj_LUT(idx) = NaN;
end

adjLUT_nonan = adj_LUT;
adjLUT_nonan(find(isnan(adj_LUT))) = 0;
PCA_img = reshape(adjLUT_nonan, 8, 8);
%PCA_img = reshape(adj_LUT, 8, 8);
hfig = figure;
imagesc(PCA_img)
cmap = [1 1 1; 0 0 1; 1 0 0; 0 0 0];
colormap(hfig, cmap);
colorbar
end

figure
plot(Distor_Meas)
```

```
xlabel('Number of Considerd Clusters');

ylabel('Obsered Distortion Measure');
```

A.2 Remove Data DC Component

```
function NormData = remove_DC(iCh, data, Fs)


t = (0:length(data)-1)/Fs;


dataSel = data(iCh, :);

M = mean(dataSel, 2);

NormData = dataSel - repmat(M, 1, size(dataSel, 2));

figure(99);

plot(t, NormData);

title('Normalized Raw Data');
```

A.3 Derive Cross Correlation and Plot

```
function RxyAmpLag = corrplots(desch, data, lutmx, center, step)


figure(98);

z = 0;

mxdim = 8;


RxyAmpLag = zeros(15,64);
```

```
for i = 1:mxdim
    for j = 1:mxdim
        z=z+1;
        subplot(mxdim,mxdim,z);
        lutmxch = lutmx(i,j);
        if ((lutmxch <= 60))
            [Rxy, lags] = xcorr(data((desch),:),
data((lutmxch),:), 'coeff');
            plot(lags, Rxy);
            title(['ch ' num2str(desch) ' and ch '
num2str(lutmxch)]);

           for a=1:15
               RxyAmpLag(a,z) = Rxy(lags == (a-5)*step + center);
           end
 else
            plot(0, 0);
            title('DNE');

        end
    end
end
```

A.4 Plot Cross Correlation Results as Heatmaps

```
function [RxyAmpConv] = corrheatmap(desch, RxyAmpLag, ...
    center, step, figs, numheat, numpoints)


MaxX=zeros(numheat,numpoints);
MaxY=zeros(numheat,numpoints);


for a=1:numheat
    for idx=1:64
        if RxyAmpLag(a,idx) == 0
            RxyAmpLag(a,idx) = 0;
        end
    end
end
figure(97);


for a=1:numheat
    %figure
    RxyAmp = reshape(RxyAmpLag(a,:),8,8);
    %RxyAmp = RxyAmp.';
    RxyAmpConv(:,:,a) = RxyAmp;
```

```
    subplot(3,3,a)
    imagesc(RxyAmp);
    colormap(parula(100));
    colorbar;
    title(['Lag: ' num2str((a-5)*step + center)]);
end
```

A.5 Derive Key Features Using Moving Average

```
function [Weights, xTimes, yTimes] = Moving_Average(Sel_Chan, Fs,
t, figs, data)

Weights=[];
t = t/60;

EventTime = xlsread('C:\Users\axelo\OneDrive\Desktop\SD
Project\Raw Data\Rat
18_10092019_Original\eventMinMaxRat18_003.xlsx','I:I');
xTimes = round(abs(EventTime)); %Idealize data ask Andrea to set
in samples, and negative times?
LongestSD = max(EventTime);
Buffer = round(LongestSD*0.70);
```

```
TimeStamps = xlsread('C:\Users\axelo\OneDrive\Desktop\SD
Project\Raw Data\Rat
18_10092019_Original\eventMinMaxRat18_003.xlsx','J:J');
TimeStamps = round(abs(TimeStamps)); %Idealize data ask Andrea to
set in samples, and negative times?

yTimes = xlsread('C:\Users\axelo\OneDrive\Desktop\SD Project\Raw
Data\Rat 18_10092019_Original\eventMinMaxRat18_003.xlsx','K:K');

for channel = Sel_Chan
    Weights=[Weights; channel, channel];
    sel_data = data(channel,:);
    window = ones(1,15000)/15000;
    mov_avg = conv(sel_data, window, 'same');

    if figs == 1
        figure(channel)
        plot(t, data(channel,:), 'linewidth', 1);
        hold on;
        plot(t, mov_avg,'k', 'linewidth', 2);
    end

    first_event = find(TimeStamps==channel)+1;
```

```
last_event = find(TimeStamps==channel+1)-1;


for event = first_event:last_event
    if (TimeStamps(event)-Buffer)>0
        StartMarker = TimeStamps(event)-Buffer;
    else
        StartMarker = 1;
    end
    EndMarker = TimeStamps(event)+Buffer;
    Test_Data = mov_avg(1,StartMarker:EndMarker);
    TimeVec = StartMarker:EndMarker;
    w = polyfit(TimeVec,Test_Data,1);
    Weights=[Weights; w];
    adj_data = [TimeVec; ones(1, (EndMarker-StartMarker+1))];
    yhat = w*adj_data;


    if figs == 1
        %xline(TimeStamps(event),'k','linewidth',2);
        xline(StartMarker/(Fs*60),'g','linewidth',1);
        xline(EndMarker/(Fs*60),'r','linewidth',1);
        %plot(TimeVec/(Fs*60), Test_Data);
        plot(TimeVec/(Fs*60),yhat,'m','linewidth',2);
    end
```

```
    end
end
```