Anticipating Postoperative Delirium During Cardiac Surgeries Involving Deep Hypothermia Circulatory Arrest

by

Owen Ma

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2020 by the
Graduate Supervisory Committee:

Daniel W. Bliss, Chair
Visar Berisha
Gene Brewer
Oliver Kosut

ARIZONA STATE UNIVERSITY

May 2020

ABSTRACT

Aortic aneurysms and dissections are life threatening conditions addressed by replacing damaged sections of the aorta. Blood circulation must be halted to facilitate repairs. Ischemia places the body, especially the brain, at risk of damage. Deep hypothermia circulatory arrest (DHCA) is employed to protect patients and provide time for surgeons to complete repairs on the basis that reducing body temperature suppresses the metabolic rate. Supplementary surgical techniques can be employed to reinforce the brain's protection and increase the duration circulation can be suspended. Even then, protection is not completely guaranteed though. A medical condition that can arise early in recovery is postoperative delirium, which is correlated with poor long term outcome. This study develops a methodology to intraoperatively monitor neurophysiology through electroencephalography (EEG) and anticipate postoperative delirium. The earliest opportunity to detect occurrences of complications through EEG is immediately following DHCA during warming. The first observable electrophysiological activity after being completely suppressed is a phenomenon known as burst suppression, which is related to the brain's metabolic state and recovery of nominal neurological function. A metric termed burst suppression duty cycle (BSDC) is developed to characterize the changing electrophysiological dynamics. Predictions of postoperative delirium incidences are made by identifying deviations in the way these dynamics evolve. Sixteen cases are examined in this study. Accurate predictions can be made, where on average 89.74% of cases are correctly classified when burst suppression concludes and 78.10% when burst suppression begins. The best case receiver operating characteristic curve has an area under its convex hull of 0.8988, whereas the worst case area under the hull is 0.7889. These results demonstrate the feasibility of monitoring BSDC to anticipate postoperative delirium during burst suppression. They also motivate a further analysis on identify-

ing footprints of causal mechanisms of neural injury within BSDC. Being able to raise warning signs of postoperative delirium early provides an opportunity to intervene and potentially avert neurological complications. Doing so would improve the success rate and quality of life after surgery.

To my friends who told me to follow my heart

ACKNOWLEDGMENTS

This research was funded by the Arizona State University-Mayo Clinic Seed Grant Program. There were no conflicts of interest. Additionally, this study is approved by the Institutional Review Board.

Firstly, I would like to thank Dr. Amy Z. Crepeau, who was the principal investigator of this study at the Mayo Clinic site, for her guidance and patience throughout this circuitous journey. This work would not be possible without Daniel Crepeau of Mayo Clinic because transferring and accessing data is actually a complex procedure in this case. I want to thank Dr. Visar Berisha, Dr. Gene Brewer, and Dr. Oliver Kosut for their feedback on this study's approach. Thank you to the members of BLISS Lab past and present. The friendship and camaraderie has made the past five years enjoyable. Thanks especially to Arindam Dutta (my partner in our small biomedical division), Christian Chapman, and Sean Bryan for their help with analysis. I would like to thank my parents. I am fortunate and privileged to have their support. Lastly, thank you to Dr. Daniel W. Bliss, who allowed me to work with him back before I even knew what a fourier transform was. His mentorship and the learning environment he created has been invaluable.

Electroencephalogram signal processing is fickle. Honestly, who would have thought that any useful information could be obtained from squiggles on a screen? Yet, here we are. Thank you all for the help.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

Chapter 1

INTRODUCTION

More recently, analysis techniques typically associated with electrical engineering and computer science have been increasingly used to inform biomedical analysis. Merging these disciplines has been particularly useful when the data is complex and difficult to process through traditional techniques. In particular, machine learning has been applied more frequently to solve open questions. These techniques are capable of classifying data, predicting events, and estimating parameters that would be far more difficult through other means. Machine learning is often applied inappropriately though. When abused, the machinery learned to accomplish classification, prediction, or estimation lack interpretability and robustness, which can be dangerous in the context of medicine. Consequently, these tools must be used more judiciously.

In situations where machine learning cannot be leveraged for raw, complex data, a comprehensive understanding of the intersecting discipline must be applied with alternative analysis techniques to accomplish the same goal. The research documented in this work is an example of such a scenario. A biomarker predictive of a neurological diagnosis must be obtained from hours of multidimensional physiological recordings. The number of cases involved in this study is small enough such that the data set is ill suited for most machine learning techniques though. Consequently, a working knowledge of the pertinent neurophysiology must inform statistical signal processing methods to extract information important for making these predictions. Presented here is a retrospective study attempting to predict incidences of postoperative delirium, a condition that can arise after a particular type of cardiac surgery, by intraoperatively monitoring patient neurophysiology. Ultimately, the goal is to antic-

ipate scenarios of heightened risk intraoperatively, facilitating the future development of interventions for minimizing and averting neurological injury.

Aortic aneurysms and dissections are life threatening conditions requiring immediate medical attention. A compromised aorta is especially concerning because it is the main artery that supplies oxygenated blood from the heart to the entire body. Aneurysms arise from stressing the aorta's walls or disease. If left untreated, the aorta can rupture. In 2018, 9923 deaths in the United States (3 per 100000 people) were the result of an aortic aneurysm of some variety [1]. Conditions contributing to the development of aneurysms include age, being male, tobacco use, high blood pressure, atherosclerosis, Marfan syndrome, bicuspid aortic valve, and family history [2]. When the aorta ruptures, there is a high risk of mortality, so aneurysms must be treated promptly. In some cases, a dissection can be repaired in time. The aorta is repaired by replacing the damaged sections with a prostheses. These surgeries are complex and risky though. Circulation must be halted to repair the aorta, which places the brain in danger. Neuroprotective techniques, the primary of which is deep hypothermia circulatory arrest (DHCA), are required to facilitate the surgery.

DHCA is effective in preventing neural injury for a majority of these surgeries. Cooling patients down to deep hypothermic temperatures ($\sim 18^{\circ}$C) staves off neuronal death, buying the surgeon time to complete repairs. Gega et al. studied the outcomes of cardiac surgeries solely using DHCA, reporting that it is typically effective with a low rate of mortality (6.3%) and stroke (4.8%) based on a 394 patient sample population [3]. Risk of neural injury is not completely eliminated though. DHCA certainly cannot protect patients indefinitely. Optimality of parameters and methods for providing neuroprotection is often debated between surgeons and institutes. Furthermore, patients can respond differently to one set of such parameters. Insufficient protection can lead to compromised neurological function.

Temporary neurological dysfunction (TND) is not uncommon following cardiac surgeries using DHCA [4]. Patients may experience postoperative delirium, a correlate of poor prognosis, as they recover in the intensive care unit [5]. Additionally, patients are at risk of developing deficiencies with motor control and memory in the long term following the operation, which would certainly affect their quality of life. Long DHCA duration, older age, severity of disease, and extent of repairs contribute to motor memory recovery difficulties [6]. The required durations for circulatory arrest vary according to surgery complexity. Several studies have correlated the duration of DHCA and postoperative prognosis to establish how long DHCA can protect the patient. Ziganshin et al. propose a 40 minute time limit DHCA [7]. Other studies argue that the limit is 10 to 15 minutes shorter, which restricts the scale of repairs that can be completed [8]. Supplementary surgical techniques have been developed in an attempt to bolster the protection.

Two techniques, antegrade cerebral perfusion (ACP) and retrograde cerebral perfusion (RCP), have been utilized to compensate for any shortcomings of DHCA. Both provide blood flow specifically throughout the brain. As a result, supplementary perfusion techniques are able to extend the duration for which circulatory arrest is safe [7]. This feature is especially important for complex and extensive procedures. Still, the risk of neurological injury is not completely eliminated. The primary issue in employing these techniques is the additional procedure complexity when implementing the appropriate perfusion circuit. Moreover, the parameters of these techniques must be carefully managed to be effective. Varying degrees of efficacy for both techniques have been reported in [9, 10, 11, 12, 13]. Some of these studies attempt to argue that either of these supplemental techniques is better than the other. Each combination of the mentioned techniques has benefits and drawbacks [7]. Certain cases may require both techniques though.

Optimally coordinating all possible surgical parameters is fickle and a general solution likely does not exist. One approach that has been researched more recently to improve the surgical outcome is the development of physiological monitoring techniques. Monitoring provides a way to tailor neuroprotection and extracorporeal support based on specific needs of each case. Ideally, we would be able to anticipate injury and identify situations of heightened risk. Anticipating such scenarios creates opportunities to intervene in hopes of averting future complications. Initial work using cues obtained from physiological monitoring to adjust patient support has been explored with some success [14]. Fortunately, different courses of action can be taken to address specific concerns depending on the signs [15, 14]. The future of being able to correct course is not simply a dream; though, these methods must be studied more specifically in context of cases utilizing DHCA. This study strictly focuses solely on providing a method for anticipating complications.

Taking advantage of physiological monitoring and metrics may provide better insight on postoperative neurological health and patient outlook. This study discusses how to anticipate postoperative delirium from intraoperative electroencephalography (EEG) data. We develop an EEG based metric that characterizes a particular neurophysiological response. Predictions about incidences of postoperative delirium are made by tracking the status of this metric. The developed algorithm, in principle, can function in real time. Congenital heart development complications may also necessitate surgeries involving DHCA. Neonatal cases present a unique set of complications though. This work focuses only on adult cases regarding repairs of the aorta, but perhaps ideas presented here may prove useful for these other delicate cases as well. Ultimately, the objective of this study is to provide a larger window of opportunity to take action and hopefully provide a pathway to improve the expected quality of life after surgery.

1.1 Background

1.1.1 Neuroprotective Techniques for Cardiac Surgeries

Circulatory arrest is necessary to repair a patient's aorta because the heart supplies oxygenated blood to the rest of the body through it. Of course, stopping blood flow deprives all organs of necessary metabolic resources to continue proper function. The brain is especially at risk because it has a much higher metabolism than other organs. Restricting oxygen reduces the production of adenosine triphosphate (ATP) and prevents cells from functioning [7]. Anaerobic glycolysis is the only mechanism that produces ATP otherwise, but intracellular waste would accumulate as a byproduct. Hypoxia also causes a sustained imbalance in calcium ion concentration levels by causing neurotransmitters to open channels through which calcium ions can easily enter, triggering intracellular dysfunction [7]. The combination of these conditions causes neuronal decay. DHCA circumvents ischemic and hypoxic injury by reducing the brain's metabolic demand.

Inducing hypothermic or deep hypothermic body temperatures (below 20° C [17]) reduces metabolic demand [8] and staves off neuronal death, buying the surgeon time to make the required repairs while circulation has ceased. Cooling is accomplished while the patient is placed on cardiopulmonary bypass (CPB). CPB is implemented by diverting deoxygenated blood away from the right atrium, pumping it out of the body, oxygenating it extracorporeally, and delivering it back through the aorta or a different artery. The patient's blood is cooled extracorporeally through a heat exchanger. Body temperature drops as a result of perfusing this cooled blood. Generally lower temperatures are more beneficial for successful recovery, but a point of diminishing returns exists past deep hypothermia levels [8]. James et al. determined that cooling patients until they exhibit electrocerebral silence (ECS), a condition in which

neural electrophysiological activity is no longer observable, minimizes metabolic demand [18]. A global standard for executing this technique does not exist, but there are institutional standards that may not all agree with each other. Patients also respond differently to cooling, so confirming that DHCA is working as intended can be complicated.

In the event that DHCA is insufficient, cerebral perfusion techniques are employed in conjunction to provide metabolic resources in case they are needed. Selective antegrade cerebral perfusion (SACP) and RCP both use the pump to circulate blood specifically through the brain. The circuit used for cooling must be altered to accomplish this while bypassing the damaged section of the aorta though. RCP is accomplished by rearranging the connections between the superior and inferior vena cava to the pump. Instead of pumping deoxygenated blood from the superior vena cava out of the body, oxygenated blood is delivered into it from the pump [19, 20]. This flow supplies blood to the brain in a retrograde fashion. Blood from the inferior vena cava is still diverted into the pump. Blood flow back toward the aorta is redirected toward the oxygenator as well so that repairs can be made. Several variations on implementing SACP exist to accommodate specific repairs [21, 22, 23]. The primary idea is to perfuse blood through the arteries stemming from the aortic arch. There are unilateral and bilateral variants of SACP. The former perfuses blood strictly through the brachiocephalic artery, and the former does so through both the brachiocephalic and left carotid arteries. Surgeons tap the right carotid artery directly while clamping the brachiocephalic artery instead if the branching point also needs to be repaired. Both sets of arteries are the antegrade pathways supplying blood to the brain. Deoxygenated blood from right atrium is funneled into a reservoir, which feeds to the oxygenator. The oxygenator then feeds blood through the branching arteries of the aorta. Modifying the circuit in this way bypasses flow through the aorta while

still supplying blood toward the brain.

Both techniques permit a significant extension in the duration of halted circulation, which is crucial for more involved repairs. Making such modifications complicates the procedure and incurs risk though. Furthermore, more parameters must now be delicately managed to ensure the correct pressure and oxygen saturation levels. Still, the most concerning sources of injury during the surgery are either metabolic or embolic. All of these neuroprotective techniques would benefit from physiological monitoring to assess patients' needs and address them by adjusting parameters of extracorporeal support. To anticipate complications stemming from these surgeries, indicators of ischemic, hypoxic, or microembolic events need to be constructed and detected.

1.1.2 Hemodynamical Monitoring

One approach to anticipating embolic and ischemic events is to monitor hemodynamics of the brain, which measures the delivery and oxygenation of blood. Additionally, these techniques are useful for executing other phases of the surgery including cannulation, cooling, rewarming, and resuming autonomous circulation [24]. Several techniques exist including regional cerebral oximetry, diffuse correlation spectroscopy, and transcranial doppler. Regional cerebral oximetry is measured through near infrared spectroscopy. An infrared source illuminates the patient's head, and a detector records the scattered response. Oxygen saturation can be estimated based on the response. Two concerns identified by Busch et al. with this technique are the changes in scattering characteristics of tissues during hypothermia and inter-device variation [24]. Diffuse correlation spectroscopy uses infrared light intensity fluctuations to monitor microvascular blood flow [24]. Direct monitoring of red blood cell flow is possible, whereas oximetry needs to extrapolate this information. Transcranial

doppler is a method that measures the speed of blood flow through sonography. Opposite of diffuse correlation spectroscopy, transcranial doppler better measures flow through large vessels. These monitoring techniques have recently been consulted to identify cues of heightened risk and intervention opportunities.

Fischer et al. measured cerebral tissue oxygen saturation through near infrared spectroscopy during DHCA surgeries with and without SACP [25]. They retrospectively examined the relationship between the duration and amount of desaturation associated with incidences of postoperative complications. The types of problems observed include delirium, sepsis, failure of different organs, stroke, and even death. They analyze the effects of different degrees of cerebral tissue oxygen desaturation on these postoperative outcomes. An association was found between the duration and amount of desaturation with incidences of complications. No interventions were employed in this study, so the efficacy of intervening based on these cues remains to be determined. Other studies have demonstrated some potential success for intervening based on related indicators though.

Zanatta et al. monitored patient physiology through several modalities simultaneously and attempted to compensate for concerns whenever observed. The monitoring modalities include electrophysiological measurements as well as transcranial doppler and brain oximetry [14]. Using the hemodynamical monitoring techniques, they are able to detect and react to microembolic events and changes in blood flow velocity. A number of specific actions can be taken to address different types of events. The control group had a 4.06% postoperative delirium incidence rate, but there were no cases of postoperative neurological dysfunction in the intervention group. Only a small fraction of the study population underwent DHCA though. The number of corrected cases that underwent DHCA needs clarification. Still, understanding that interventions can help in related procedures further motivates proceeding with this

research.

1.1.3 Electrophysiological Monitoring

Electrophysiological monitoring has also been utilized to better understand the neurophysiology during aorta surgeries. The brain coordinates activity through complex electrical activations of neurons. This activity induces a mix of fields propagating in various directions through layers of tissue and bone. Placing electrodes onto the scalp allows us to observe this collective activity. By measuring this activity, we hope to be able to directly understand how the brain responds to the surgery. Electrophysiological measurement methods include EEGs and somatosensory evoked potentials (SEPs). The difference between EEG and SEPs is that the former passively observes the activity, whereas the latter observes responses evoked near the somatosensory cortex after applying an electrical stimulus elsewhere on the body. Examining the electrophysiology during the surgery may allow us to observe developing complications.

Early work in analyzing patient electrophysiology during procedures involving DHCA focused on characterization. In 1959, Pearcy et al. observed that decreasing body temperature for hypothermia circulatory arrest reduces intraoperative EEG "waveform complexity" and signal amplitude [26]. Additionally, they noticed waveforms did not resemble those prior to arrest occasionally, and hypothesized that this characteristic signifies a poor prognosis. In 1984, Levy investigated the relationship between temperature changes during DHCA and spectrograms [27]. A correlation between temperature, power, and high frequency band peak power frequency was observed. Later in 2003, Levy and colleagues examined the relationship between the approximate entropy of intraoperative EEG signals and body temperature over the course of DHCA [28]. This metric, based on Kolmogorov-Sinai entropy [29], is in-

tended to measure the predictability of the following sample given a subsequence. They found that temperature was related with approximate entropy through a sigmoid function.

Stecker et al. characterized changes in SEPs and EEG activity throughout DHCA. In the first of a pair of studies, they analyzed changes in both measurements as function of cooling [30]. They identified nasopharyngeal temperature ranges at which certain common phenomena arise, including ECS, burst suppression, and periodic complexes. Only 60% of patients exhibited ECS at 18° C, which is concerning if cooling to ECS in fact ensures a better outcome by optimally suppressing metabolism. Supposing the promises of ECS, this finding suggests blindly cooling to a target temperature may not be advisable. Consequently a need for electrophysiological monitoring to guide the cooling process is prompted. Additionally, a relationship between body temperature and the disappearance of evoked responses was found. As temperature drops, the N20-P22 complex first stops, followed by the N13 wave which occurs when approaching average deep hypothermic temperatures [30].

One important electrophysiological phenomenon observed during cooling and warming phases of DHCA is burst suppression. This behavior is characterized by alternating periods of nominal amplitude activity and suppressed activity [31]. In the context of DHCA for cardiac surgeries, burst suppression is triggered when patient body temperatures are low enough. The temporal characteristics of this activity is atypical. These quasi-periodic cycles are observed on a temporal scale ranging from tens of seconds to milliseconds depending on temperature. Lowering the temperature sufficiently will continually suppresses EEG activity, or in other words ECS will be induced. Westover et al. developed a thorough characterization of burst suppression and temperature through frequency and time domain features [32]. They demonstrate that the probability the brain is in a suppression state and correspondingly the

duration of suppression phases are modulated by body temperature. Decreasing the temperature increases the likelihood of EEG activity being suppressed. Burst suppression is commonly characterized using burst suppression ratio (BSR), which measures the fraction of an epoch spent in a suppression phase as defined in [33]. Naturally, a negative correlation exists between temperature and BSR [32, 34]. Lastly, they observed that the spectral shape of bursts, after normalizing by its energy, is maintained despite changing the temperature.

Westover et al. did not specifically examine how pathologies affect their observations, but this work in characterizing burst suppression is important for recognizing typical behavior under such strange circumstances. Understanding what is typical aids us in identifying abnormal cases, which presumably lead to postoperative delirium. Previous studies have hypothesized that cues of neural injury may be expressed in characteristics of burst suppression activity [32]. This hypothesis is motivated by a proposed link with brain metabolism [35] informed by a mathematical model for the electrodynamical behavior of a network of neurons. The model construction accounts for possible avenues through which the surgery affects metabolism. The electrodynamics are reproduced by adjusting parameters of the model that the surgery may cause. The occurrence of burst suppression may not immediately indicate poor prognosis, but perhaps certain abnormalities in that activity would.

More recent works in this subject are directed toward anticipating compromised neurological integrity through electrophysiological monitoring. Bispectral index (BIS) has been utilized as a method for tracking the recovery of electrocerebral activity during surgeries requiring circulatory arrest. The EEG signals from which this metric is derived are recorded typically on the frontal lobe. Typically BIS is monitored to track depth of anesthesia, but a correspondence between temperature and BIS has been determined [34]. BIS is a weighted sum of disparate measurements derived

from bispectra and power spectra of EEG recordings [33]. One of the features comprising the BIS is SynchFastSlow, defined as the ratio between the volumes of two regions within the bispectrum of an EEG signal $x(t)$. The bispectrum is evaluated as $X(f_1)X(f_2)X^*(f_1 + f_2)$, where $X(f)$ is the fourier transform of the recording. Another measurement comprising BIS is the beta ratio, defined as the log ratio of the two beta band energies. BSR and QUAZI suppression (similar to the former, but considers time spent nearly suppressed) are also incorporated into the BIS. Santarpino et al. observed that a reduction in BIS greater than 20% from baseline for at least 15 minutes was correlated with postoperative neurological complications [36]. Determining the occurrence of these situations only comes after attempting to maintain patient BIS above some threshold. Intervention was either pharmacological or through cooling. If a decrease was identified despite their attempts intervene within a window of time, then they predict the occurrence of postoperative neurological complications. The authors only mention that these episodes came "approximately 40-60 minutes during maintenance". One discrepancy between the current study and the one by Santarpino and colleagues is that their study used strictly hypothermia circulatory arrest, which is distinct from deep hypothermia and elicits a different set of circumstances.

In the second of their pair of studies, Stecker et al. examined electrophysiological responses to the warming phase [37]. Additionally, they determined the effect of incidences of intraoperative strokes or postoperative confusion on these responses. Patients experiencing intraoperative strokes or postoperative confusion tended to take longer to recover continuous EEG activity once full body perfusion resumed. The queried events were annotated by the electroencephalographer. The distinction between the pathological group and non-pathological group is less clear based on changes in SEP responses throughout warming. Similarly, Soehle et al. were able to corre-

late average intraoperative BSR and burst suppression duration measurements with postoperative delirium diagnoses [38]. They found that the median BSR was statistically significantly higher for the delirium positive group than the delirium negative group. Surgeries studied in this study, however, were restricted to cardiac surgeries that involved CPB, but did not use DHCA.

Ma et al. were able to accurately infer postoperative delirium diagnoses based on spatiotemporal covariance structure of EEG signals during the warming phase following circulatory arrest [39]. The constructed detector requires post-DHCA eigenvalue distribution structure to sufficiently resemble baseline activity (i.e. characteristics prior to cooling). This study assumes that no complications occurred prior to stopping circulation. Although predictions are made from intraoperative data, there is no analysis on how much foresight can be achieved with this. The prediction procedure currently does not provide much opportunity to be proactive.

Again, some work has explored the possibility of using electrophysiological cues to intervene and avert risky situations. In addition to the hemodynamical monitoring, Zanatta and colleagues relied on both EEG and SEPs to intervene [14]. The EEG signals measured were the following differential pairs F3-C3 and F4-C4. A frequency domain feature, they term spectral edge frequency, was extracted from the EEG recordings and monitored over time. Spectral edge frequency is the frequency below which 95% of the total energy is contained. When this frequency drops below 50% of baseline, then brain oxygen delivery is increased. The measured SEPs include N20-P25, N30, P14-N18, N13, and N9. Brain oxygen delivery was also increased when the amplitudes of the evoked potentials drop below 50% of baseline. The presumed cause of the drops in edge frequency and evoked potential amplitudes were either microembolic events or low blood pressure. Zanatta et al. note that 3% of the monitored group had reversible SEP abnormalities. Again, which of the intervention

group were patients that underwent DHCA was not clarified. Only a small fraction of these surgeries utilized DHCA. Further study on intervention methods in DHCA surgeries is needed.

Fortunately, these previous studies demonstrate the worth of using additional physiological monitoring to guide these surgeries, anticipate injuries, and address specific concerns in extracorporeal support. The goal of this work is to improve our ability to anticipate the development of neurological complications in the context of surgeries requiring DHCA. Specifically, this study pursues the hypothesis that anomalous burst suppression characteristics may be correlated with postoperative delirium.

1.2 Contributions

So far in Chapter 1 we introduced and motivated the problem this dissertation addresses. We reviewed the basis of neural injury in these surgeries and previous research regarding the extraction of intraoperative biomarkers predicting postoperative neurological complications. Here, we outline the contributions we made toward this field of research, the challenges of this specific study, and the methods involved to overcome them. This dissertation expounds on a journal article, titled "Anticipating Postoperative Delirium During Burst Suppression Using Electroencephalography", we recently published in *IEEE Transactions on Biomedical Engineering* [40]. We improve the methods outlined in [40], achieving a slight improvement in the results while revising the procedure such that it is more robust. This work also contextualizes the current results in terms of our initial analysis documented in [39] .

This retroactive study researches how to predict postoperative delirium from EEG data recorded during cardiac surgeries involving DHCA and supplementary perfusion techniques. More specifically, this research aims to make predictions early enough so

that an earlier opportunity for intervention is provided to avert neural injury. Toward that end, this study examines the burst suppression phenomenon during patient warming because it is the first observable electrophysiological activity following circulatory arrest. We hypothesize that abnormalities in the burst suppression sequence are correlated with delirium positive cases. Burst suppression duty cycle (BSDC) is formulated to specifically monitor changes in this phenomenon over time. BSDC is based on signal energy, so an artifact detection algorithm and a spatiotemporal averaging preprocessing step are also developed to properly extract meaningful information. The algorithm is designed such that all necessary processing and detection parameters are learned prior to when BSDC needs to be monitored. This methodology can function in real time granted a priori knowledge of characteristics describing typical delirium negative and delirium positive burst suppression sequences. Details of these first steps are discussed in Chapter 2. In Chapter 3, the predictive power of this methodology is assessed by a Monte Carlo supervised learning simulation. We demonstrate through the results of the constructed prediction methodology that delirium diagnoses are correlated with how BSDC, a burst suppression characteristic, evolves over time. Supporting analyses to validate these results are presented in Chapter 4. These additional forms of analyses include additional classification problems regarding other parameters of the surgeries and evaluating descriptive statistics to identify any confounding factors. The results are reconciled with the surgical outcomes in Chapter 5. Additionally, the significance of this study's findings are contextualized within the previous works we reviewed. Lastly, we summarize in Chapter 6 what this study achieves and the implications it has for future research. Much work remains, so specific paths forward are proposed as well.

Chapter 2

DATA COLLECTION AND FIRST STEPS

2.1 Surgery and Physiological Measurements

This study is a collaboration between Arizona State University (ASU) and the Mayo Clinic. All documented analysis is retrospective. Operations were performed at a Mayo Clinic facility. Intraoperative physiological recordings, postoperative neurological evaluations, and general data regarding the types of cases each patient underwent were transferred to ASU after information possibly leading to the identification of each patient was removed. Sixteen patients underwent aorta repair surgery with DHCA supplemented by cerebral perfusion techniques. ASU possesses data for patients numbered 1 through 19, except 16. Two patients, labeled 7 and 15, did not undergo DHCA, and therefore are exempt from analysis. Identification numbers were not adjusted.

These surgeries progress generally as follows. After intubation, the patient is first anesthetized. Afterward, surgeons expose the heart and place the patient on cardiopulmonary bypass (CPB). Deep hypothermia is induced by cooling the blood as it passes through the pump. The heart's beating is suspended by infusing a cardioplegic solution into the patient's blood [16]. There are two methods, cold blood and cold crystalloid cardioplegia, which differ in the composition of the solution. Procedures in this study underwent cold blood cardioplegia. Patients have been sufficiently cooled when they exhibit ECS. Circulatory arrest is induced once ECS is observed. RCP is applied during circulatory arrest. CPB resumes after surgeons complete the required maintenance. In cases that needed SACP, patients were switched between

the perfusion strategies as needed to make the necessary repairs. Body temperature is raised by warming the patient's blood as it passes through the pump. Autonomous circulation resumes when the patient is able to maintain a stable heart rate. Cardiac pacing or defibrillation may be necessary to force this.

EEG data was recorded at a sampling frequency of 500 Hz from intubation until shortly after autonomous circulation is resumed. Twenty-three electrodes were placed at locations specified by the 10-20 system [41]. The timeline was annotated by the electroencephalographer to indicate when major surgical events occurred. These include but are not limited to anesthesia, CPB, cooling, DHCA, warming, resuming autonomous circulation, cardiac pacing, and defibrillation. Body temperature was measured using an esophageal probe and recorded intermittently throughout the surgery .

Postoperative delirium was diagnosed by nurses in the intensive care unit using a two-step procedure comprising of first the Richmond Agitation-Sedation Scale (RASS) and followed by the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU) [42]. A RASS score, which spans a range of -5 to 4 in integer increments, describes patient responsiveness to external stimuli. The lowest value indicates the patient is so sedated that they are not yet responsive to touch or voice. Zero represents a normal state of being. The highest score indicates that the patient is easily roused and combative. The CAM-ICU can be administered once a RASS score of -3 has been met, meaning the patient is at least reacting to the nurse's voice. Patients then undergo a short cognitive battery. If patients do not successfully answer the questions, then they are diagnosed with postoperative delirium.

The prescribed repairs and delirium diagnosis for each patient studied are documented in Table 2.1. Most major sections of the aorta are represented in this dataset. Aortic root repairs included a hemiarch replacement and an open distal anastomosis,

necessitating DHCA. Delirium positive is designated by a P, and delirium negative is represented by an N. Ultimately, the goal of this study is to extract information from the EEG recordings to predict the CAM-ICU result.

Table 2.1: Patient Procedures and CAM-ICU Results

Patient	Procedure	Perfusion	Delirium
1	Aortic root replacement	RCP	N
2	Mitral valve/aortic root replacement	RCP	P
3	Aortic arch replacement	RCP	N
4	Aortic aneurysm/dissection repair	SA/RCP	P
5	Aortic root aneurysm repair	RCP	P
6	Aortic arch replacement	SA/RCP	N
8	Aortic root aneurysm/dissection repair	SA/RCP	N
9	Aortic root/ascending aortic arch replacement	RCP	N
10	Aortic valve resuspension, aortic/ascending arch replacement	SA/RCP	P
11	Aortic root/ascending aortic arch replacement	RCP	N
12	Aortic root/modified total arch replacement	SA/RCP	N
13	Thoracoabdominal aortic aneurysm repair/reconstruction	RCP	P
14	Thoracoabdominal aortic aneurysm repair/reconstruction	RCP	N
17	Distal arch/proximal descending aorta replacement	RCP	N
18	Aortic root/arch replacement	RCP	N
19	Aortic root/arch replacement	RCP	N

2.2 Preprocessing

EEG data must be preprocessed before information relevant to the problem can be extracted. All recordings are first band-pass filtered from 1 to 40 Hz, which removes DC drift and filters out power mains. Afterward, the data is down sampled to 100 Hz. In doing so, we assume that information above 40 Hz is negligible. High-pass filtering at 1 Hz removes some structure from the bursts during burst suppression, but maintaining that content elsewhere is perhaps more trouble than it is worth. An example voltage time series is shown in Figure 2.1. This data is recorded from location Cz during patient 9's procedure.

Examining spectrograms allows us to better understand the significance of the time domain signal. The spectrogram associated with Figure 2.1 is shown in Figure 2.2. A 1000 sample long hamming window with 90% overlap and 2048 frequency bins were used to create this figure. Several features in this spectrogram common between

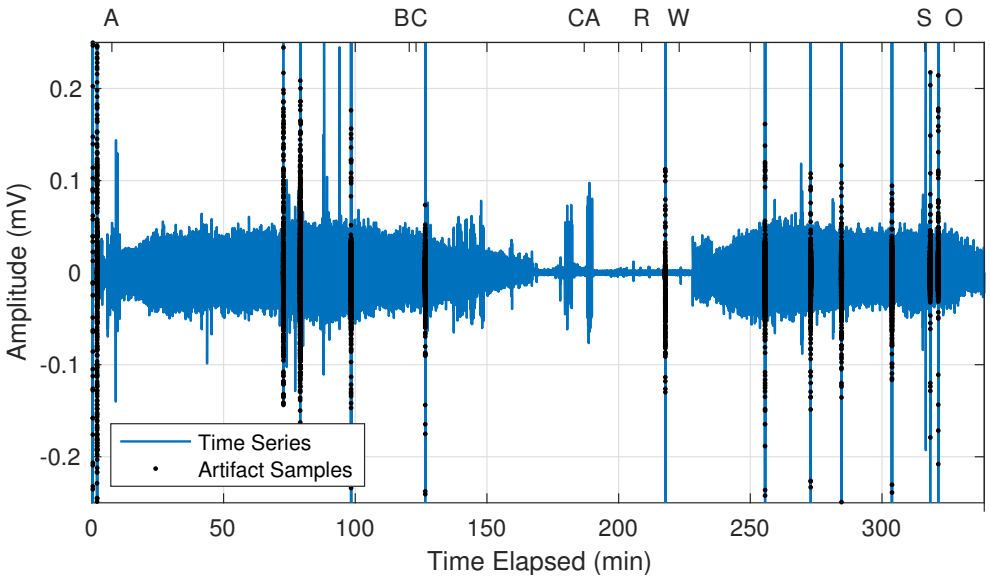

Figure 2.1: Time series voltage, recorded at EEG location Cz from patient 9. The labels annotating the timeline include: A-Anesthetization, B-CPB on, C-cooling, CA-circulatory arrest, R-resume CPB, W-warming, S-cardiac shock, and O-CPB off.

all channels and patients are identified. As anesthesia takes effect, a time-frequency ridge appears. The precise frequency of this ridge fluctuates over time and typically resides within or near alpha wave frequencies. This frequency falls during cooling. Burst suppression begins shortly afterward. Eventually, cooling the patient enough causes ECS. All of these events happen in reverse order once rewarming begins. The anesthesia time-frequency ridge reconstitutes itself, signifying that seemingly typical anesthesia activity has resumed. Sometimes this ridge progresses along a slightly different course in frequency than where it was located prior to cooling. When comparing against pre-cooling power spectra, we notice the general shape post-warming is similar, but the power distribution can be different.

Additional context within the overall surgery is provided through the electrocardiogram's spectrogram shown in Figure 2.3. Heart activity can be recognized through the definitive harmonic structure. Despite bypassing the heart, it continues to beat

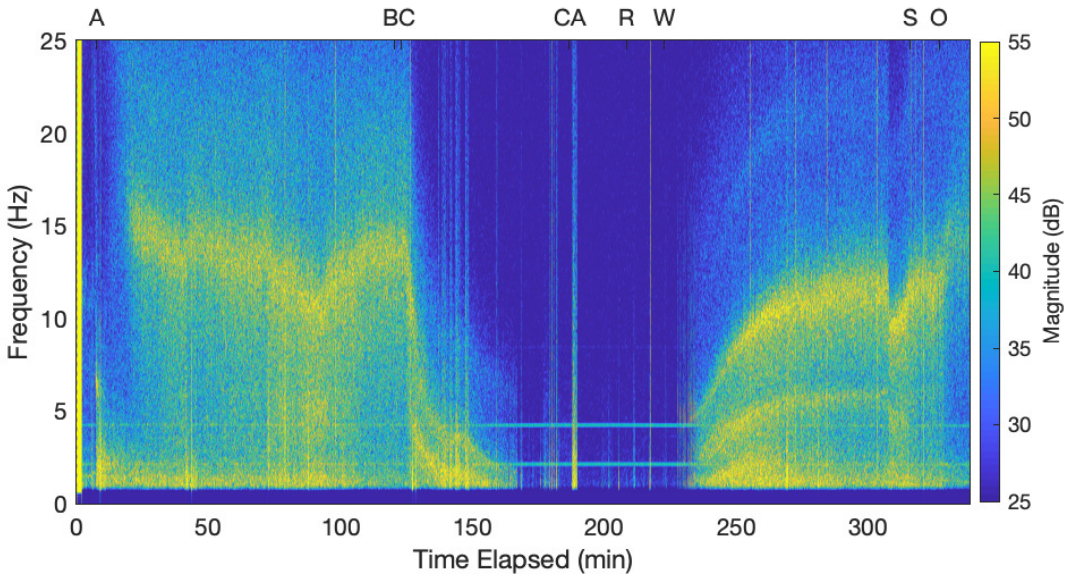

Figure 2.2: Spectrogram of patient 9's Cz recording. Interpreting the time series in Figure 2.1 is easier when examining its spectrogram. The labels annotating the timeline include: A-Anesthetization, B-CPB on, C-cooling, CA-circulatory arrest, R-resume CPB, W-warming, S-cardiac shock, and O-CPB off.

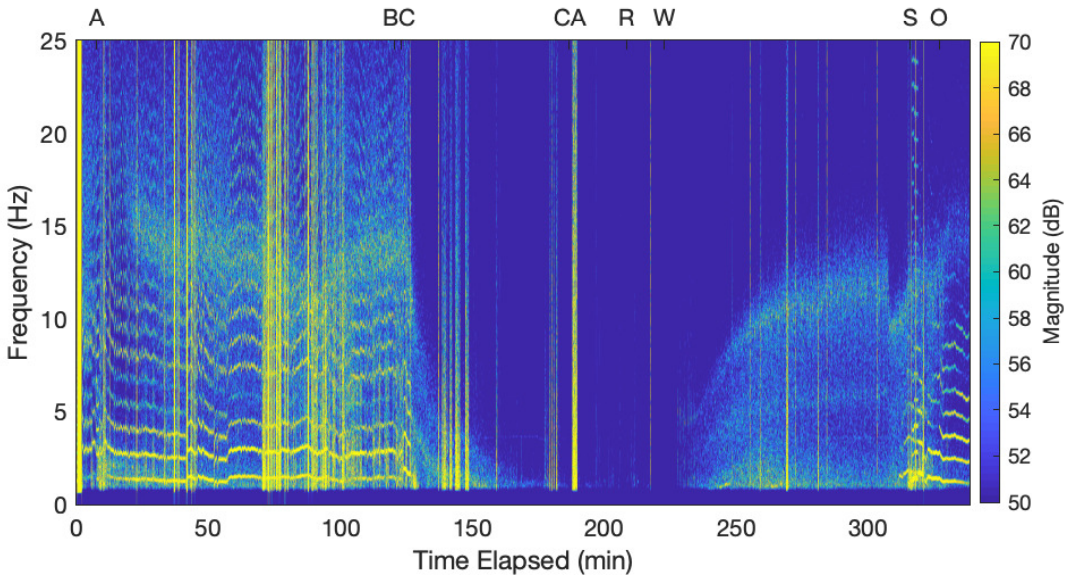

Figure 2.3: Spectrogram of patient 9's ECG recording. Heart activity can be recognized through the strong harmonic structure. The labels annotating the timeline include: A-Anesthetization, B-CPB on, C-cooling, CA-circulatory arrest, R-resume CPB, W-warming, S-cardiac shock, and O-CPB off.

while the pump drives circulation. The heart reacts quickly to cooling and slows down. After infusing the cardioplegic solution into the patient's blood, the heart stops beating and the harmonic structure vanishes. Some time passes before halting circulation because ECS has not been observed yet. After circulatory arrest, the heart does not start beating during the warming phase in this particular case. Not until a shock is applied does the harmonic structure return. After the patient is taken off bypass, the heart resumes its role and a heartbeat is maintained.

Various artifact signals are present throughout the recordings. EEGs are susceptible to recording artifacts. Filtering does not address these issues. Artifacts will corrupt derived measurements and mislead if left untreated. Identifying these responses is important for extracting meaningful information from the data. These artifacts are typically non-biological periodic signals, such as rectangle waves, with energy noticeably much larger than that of nominal neurological activity. Operating room environ-

ments contain many electromagnetic interferers originating from surgical instruments and monitors. Slight shifts in the electrodes while repositioning the patient, pumps, and ventilator can also produce artifacts that are impulse-like or consist of unnatural oscillations. Many of these artifacts occur prior to warming. Unfortunately, the cooling phase contains many artifacts that obscure neurological activity—in particular burst suppression. The data is, however, relatively uncorrupted once CPB resumes for the studied surgeries. Designing filters for these artifacts or attempting to recover the neural signals is difficult because these procedures may heavily distort the time-frequency structure of the remaining recording. Discrepancies in observations across channels complicate our ability to attempt signal recovery. Additionally, the extent and characteristics of the contamination vary between not only cases, but also over the course of the entire surgery. A severe example of artifact contamination is shown for patient 3 in figure 2.4, the spectrogram of their Cz recording. Instead, we identify when artifacts occur and treat them while postprocessing derived features.

Artifacts are detected by thresholding the windowed energy of each EEG channel. The log scale windowed energy is given by

$$\epsilon[n] = \log_{10}\left(\sum_{k=0}^{N-1}\left(w[k+n]x[k]\right)^2\right), \tag{2.1}$$

where $x[n]$ is the filtered time series data, $w[n]$ is a 10 s rectangle window, and N is the number of samples. Channel specific thresholds are set based on the histogram of energy observed from anesthetization until the cooling phase. Thresholding must be performed on a per channel and per case basis due to the diversity in observations. All histograms of the energy included a long tail, attributed to high energy outliers. Developing a detector based on a fixed order statistic threshold is not robust due to the variance in contamination severity. For example, 99^{th} percentile energy values may be true positives in some cases, but may be false positives in others. Consequently,

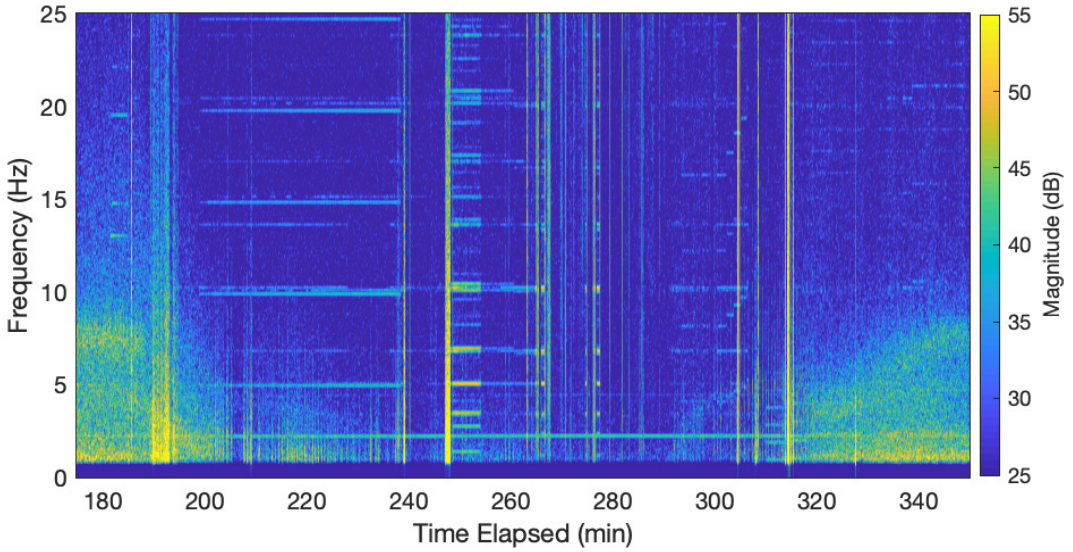

Figure 2.4: Spectrogram of patient 3's Cz recording encompassing a portion of patient 9's cooling, circulatory arrest, and warming phases. Shown here is an extremely contaminated case. Many of the artifacts are periodic in nature with strong harmonic structure. Other artifacts are impulsive and easily observed as broadband streaks within a short duration.

energy thresholds are chosen from histograms heuristically as

$$2\mu_\epsilon - \sigma_\epsilon \,, \tag{2.2}$$

where μ_ϵ is the mean determined by the histogram's maximum, and σ_ϵ denotes a "standard deviation" found by measuring the width of the portion where most observations lie. Selecting the threshold in this way generalizes this energy detector well enough to the entire data set. This method errs on the side of false negatives as opposed to false positives. An example of one such histogram and threshold selection is presented in Figure 2.5.

The ensuing detections have been labeled in the example voltage and corresponding energy time series shown in Figure 2.1 and Figure 2.6. Notice the remaining unlabeled artifacts that have energy close to nominal levels. Only the largest amplitude artifacts are treated through this method, but identifying them simplifies

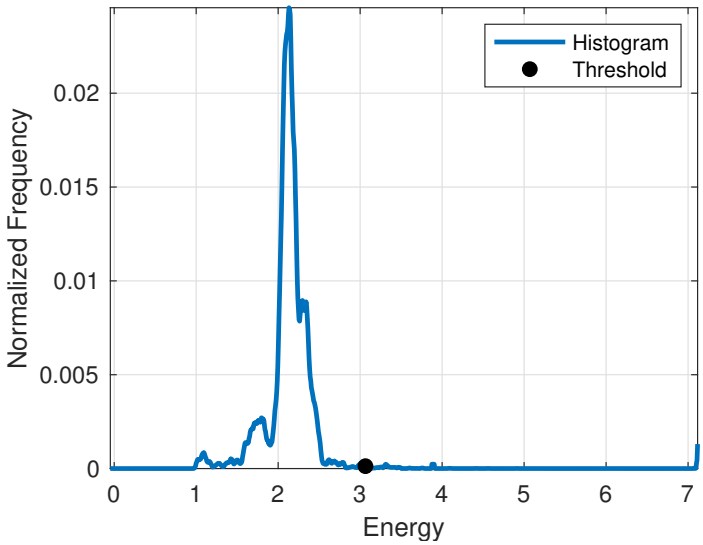

Figure 2.5: Histogram of energy computed from data recorded prior to the cooling phase. This recording, a part of patient 9's data set, was measured at EEG location Cz. The energy threshold used for declaring the occurrence of artifacts is indicated by the black dot.

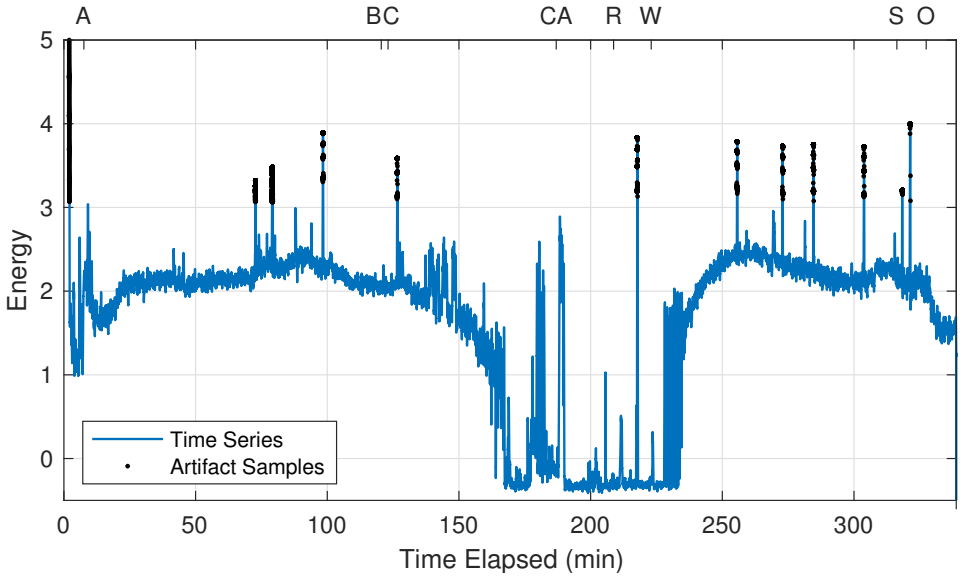

Figure 2.6: Windowed energy associated with the time series voltage shown in Figure 2.1. The labels annotating the timeline include: A-Anesthetization, B-CPB on, C-cooling, CA-circulatory arrest, R-resume CPB, W-warming, S-cardiac shock, and O-CPB off.

subsequent analysis. In the case shown in Figure 2.4, the neural signal has lower, if not comparable energy to the more subtle artifacts. We address more subtle artifacts by processing the data.

We perform a weighted averaging of the data through a spatiotemporal version of principal components analysis (PCA) to improve the neural signal quality and aggregate information recorded from multiple spatial locations. This analysis technique was used in a previous study of ours, but its purpose is different here [39]. The combination of information across channels helps address the issue that different channels have different observation qualities. Furthermore, our goal is to analyze burst suppression as a spatially global event. We attempt to improve our observation of this phenomenon by combining several diverse recordings.

A spatiotemporal combination, instead of a strictly spatial version, is used as a safety net to account for potential differences in arrival time across channels from any common sources. For each patient, a space-delay data matrix is constructed using the preprocessed data from channels Cz, FPz, Fz, Oz, and Pz. Any differences in activity across hemispheres are not considered in this study, but could be leveraged in future studies. The benefit of choosing these five channels was empirically verified when recognizing that the quality of the neural signals improved. We designate the non-delayed data matrix as

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_{\mathrm{Cz}} \\ \mathbf{x}_{\mathrm{FPz}} \\ \mathbf{x}_{\mathrm{Fz}} \\ \mathbf{x}_{\mathrm{Oz}} \\ \mathbf{x}_{\mathrm{Pz}} \end{bmatrix}, \tag{2.3}$$

where each row contains time series data from one of the five channels examined. Five delay taps are used to construct the space-delay matrix, where the taps are spaced

by 10 ms. Space-delay matrix $\mathbf{Z}$ is given by

$$\mathbf{Z} = \begin{bmatrix} \mathbf{X}_0 \\ \mathbf{X}_1 \\ \mathbf{X}_2 \\ \mathbf{X}_3 \\ \mathbf{X}_4 \\ \mathbf{X}_5 \end{bmatrix}, \tag{2.4}$$

where the subscripts indicate the tap index. Correlation matrices, $\mathbf{R}$, are estimated from data recorded prior to cooling, a subset of columns in $\mathbf{Z}$ denoted as $\mathbf{Z}'$. Consequently, the information recorded afterward is analyzed in reference to a baseline spatiotemporal covariance structure. Artifact segments are also temporarily excised from this subset prior to estimating the matrices. Leaving the artifacts in $\mathbf{Z}'$ would include unnatural correlations. Their large energy would also dominate the computation of the second order statistics. Artifact labeling was performed for each channel independently, and now several channels have been combined. Labels are reconciled across channels through a logical OR operation. The spatiotemporal correlation matrix is given by

$$\mathbf{R} = \frac{1}{N_{\mathbf{Z}'}} \hat{\mathbf{Z}}' \hat{\mathbf{Z}}'^{\mathrm{T}}, \tag{2.5}$$

where $N_{\mathbf{Z}'}$ is the remaining number of samples and $\hat{\mathbf{Z}}'$ is a row centered and normalized version of $\mathbf{Z}'$. PCA first decomposes the correlation matrix into

$$\mathbf{R} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}, \tag{2.6}$$

where $\mathbf{V}$ contains eigenvectors in each column and are organized according to its associated eigenvalues stored in the diagonal of $\mathbf{\Lambda}$. The eigenvalues are sorted from largest to smallest, where the largest eigenvalue is denoted by λ_1. The entire space-

delay data matrix $\mathbf{Z}$ is then row centered and normalized. The spatiotemporally weighted average is expressed by

$$\mathbf{s}_1 = \mathbf{v}_1^{\mathrm{T}} \hat{\mathbf{Z}}\,, \tag{2.7}$$

where $\mathbf{v}_1$ is the principal basis vector and $\hat{\mathbf{Z}}$ is $\mathbf{Z}$ centered and normalized by the same means and standard deviations yielded by $\mathbf{Z}'$.

We demonstrate the effect this averaging has on the severely contaminated data recorded during patient 3's procedure through a spectrogram of $s_1(t)$ shown in Figure 2.7, where $\mathbf{s}_1$ is simply the vectorization of the time series data. Many artifacts have receeded to the background, but others have become clearer. The neural signal compared to signal noise is much stronger now though. The eventual goal is to distinguish bursts from signal noise. In Figure 2.4, the bursts were difficult to observe, but are more distinct now in Figure 2.7. Perhaps there exists a more optimal subset of channels to exploit here, which would improve the outcome of this procedure. For

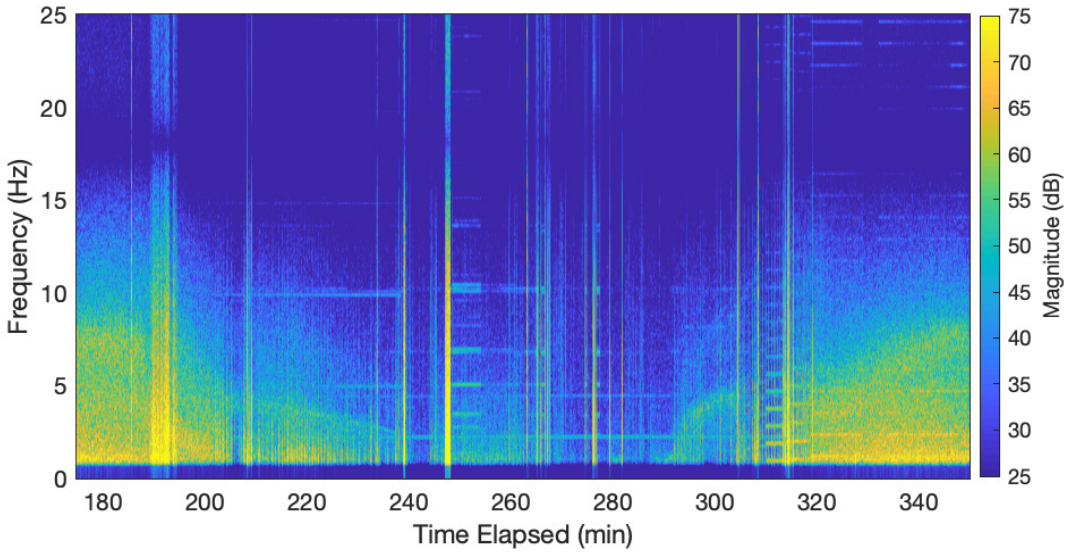

Figure 2.7: Spectrogram of patient 3's spatiotemporally averaged time series data. This segment contains a portion of patient 3's cooling, circulatory arrest, and warming phases. The neural signal has become significantly clearer. Some of the artifacts have been pushed to the background, while others have become more pronounced.

now, we proceed with the rest of the analysis using the principal component $s_1(t)$.

2.3 Burst Suppression Duty Cycle

We choose to examine burst suppression through windowed energy, one of its more distinguishing features. One of the goals moving forward is to be able to distinguish burst activity from suppression phases. Computing the windowed energy allows us to label the time series based on burst envelopes. The windowed energy computed for the subsequent analysis uses parameters different from those used for artifact detection. We evaluate

$$e[n] = \log_{10}\left(\sum_{k=0}^{N-1}\left(w[k+n]s_1[k]\right)^2\right), \tag{2.8}$$

where $w[n]$ is a hamming window this time. Additionally, the window has been shortened to 2.5 s to preserve shorter term characteristics. An excerpt of the patient 9's synthesized time series and corresponding windowed energy, highlighting burst suppression, is depicted in Figure 2.8. Burst suppression can be recognized through the quasi-periodic bursts in energy. Burst suppression occurs at an atypical time scale compared to usual neural events. Aside from ECS, the next burst could occur tens of seconds after the previous cycle. As warming progresses, the time to retrigger shortens. One weakness of analyzing the windowed energy is that suppression phases eventually shorten below 2.5 s, making distinguishing some bursts from another difficult. The window length is a parameter that should be tuned in future iterations of this analysis. Bursts seemingly retrigger even during the previous burst toward the end of burst suppression. We assume burst suppression is a global event, but it is possible that bursts could be arriving at the EEG electrodes from different locations and overlapping with each other. Regardless, "continuous" EEG activity is eventually observed.

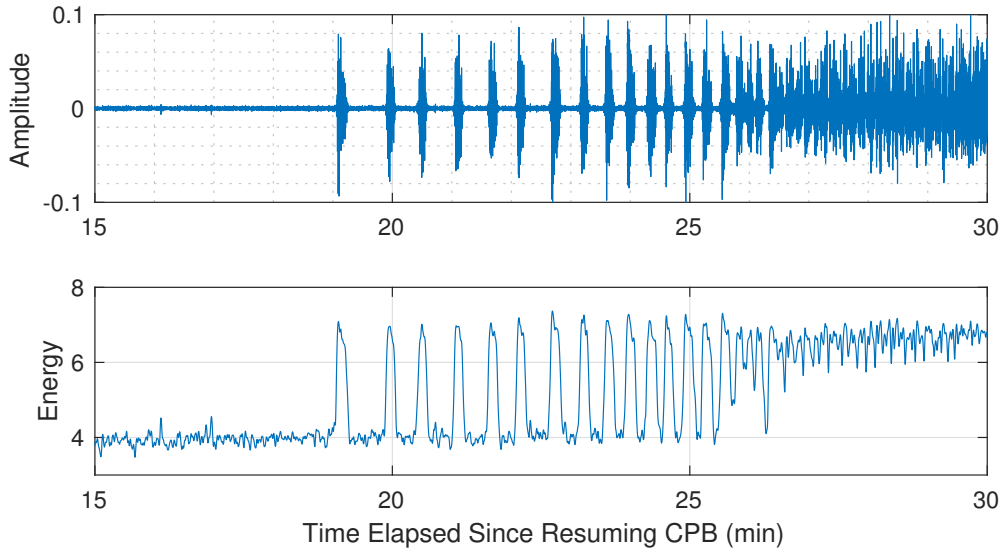

Figure 2.8: Burst suppression represented in spatiotemporally averaged time series data for patient 9 and its associated energy time series. Suppression phase duration decreases over time. Eventually, the endings and beginnings of sequential bursts becomes difficult to discern, indicating the end of burst suppression. A temporal structure of each burst can be observed roughly.

All energy time series starting when CPB resumes until the patients resume autonomous circulation are shown in Figure 2.9 and Figure 2.10. Times when burst suppression begins, when it ends, and when warming begins are marked in these figures. Note that the end of circulatory arrest was not documented for patient 14, so the timestamp indicating when rewarming begins is used instead. Quite a bit of variance in signal energy and burst suppression characteristics across patients can be observed. Some characteristics to note include the time elapsed until burst suppression begins, how long burst suppression lasts, and how long before warming began. The aforementioned features play a role in making predictions.

Changes in dynamics are tracked through a metric called burst suppression duty cycle (BSDC). BSDC quantifies the proportion of time within a window the brain spends in a burst phase. The quantity $1 - \text{BSR}$ is semantically the same as BSDC.

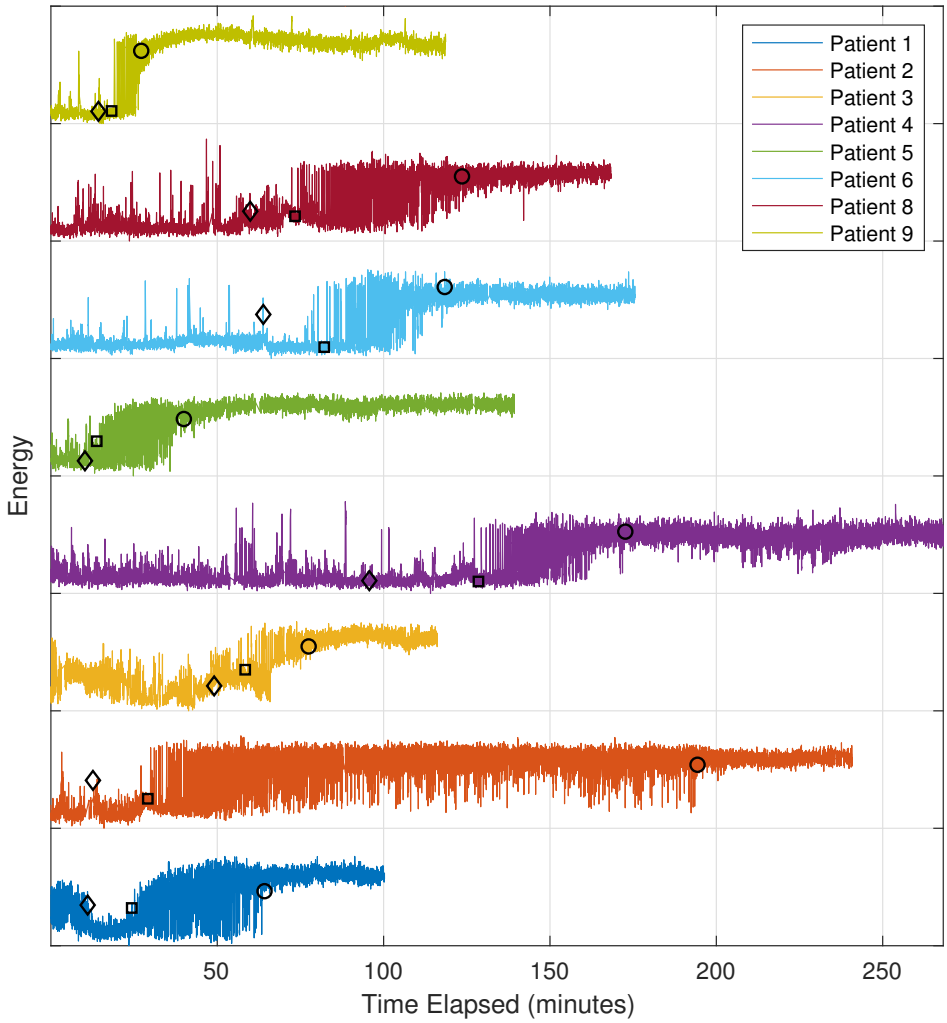

Figure 2.9: Energy time series for patients 1, 2, 3, 4, 5, 6, 8, and 9 starting when CPB resumes until it ends. Also highlighted are the beginnings (squares) and ends (circles) of burst suppression. These markers are found through analyzing BSDC. Warming starts at the time where the diamonds are located. Patients 2, 4, and 5 belong to the delirium positive group.

Computing BSDC is, however, explicitly different from evaluating $1 - \text{BSR}$, where BSR is attached to a standard. This difference stems from Rampil defining a suppression phase as a period where the voltage does not vary beyond ± 0.5 mV for

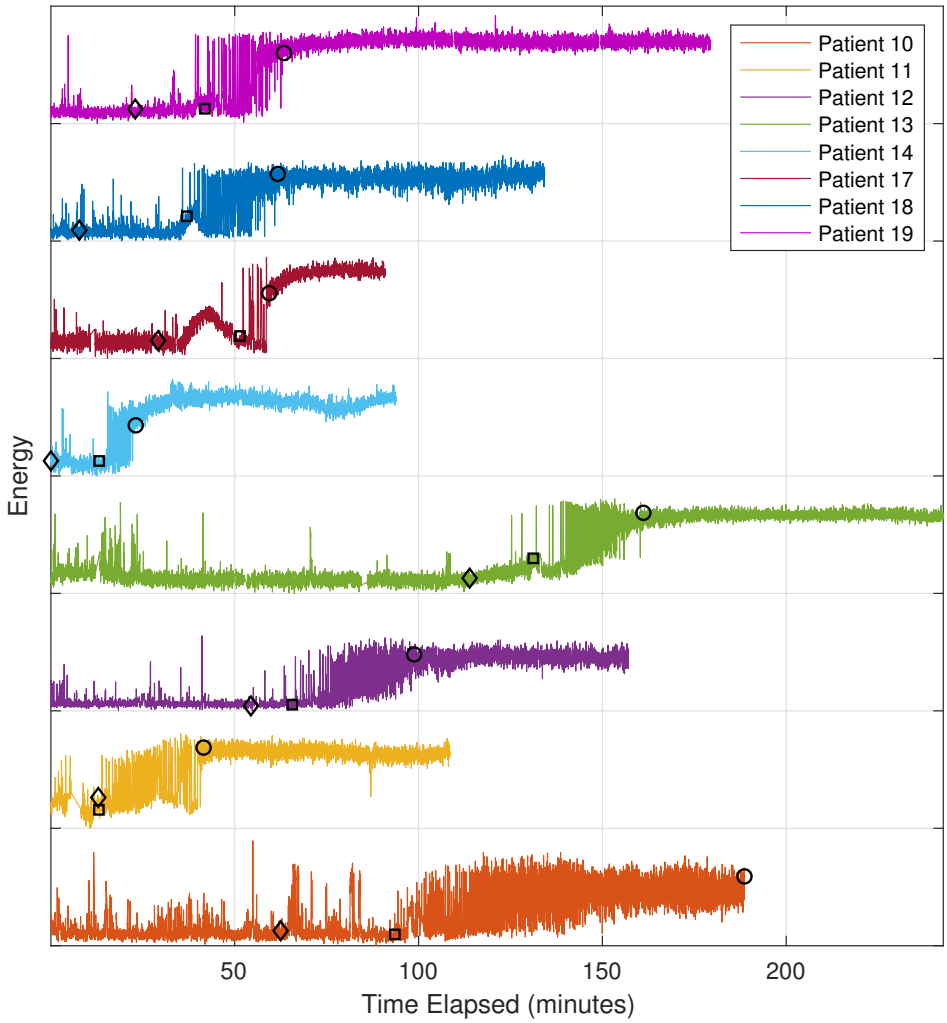

Figure 2.10: Energy time series for patients 10, 11, 12, 13, 14, 17, 18, and 19 starting when CPB resumes until it ends. Also highlighted are the beginnings (squares) and ends (circles) of burst suppression. These markers are found through analyzing BSDC. Warming starts at the time where the diamonds are located. Patients 10 and 13 belong to the delirium positive group. The end of patient 14's circulatory arrest was not annotated, so we default to using the rewarming time stamp.

longer than 0.5 s [33]. BSR is defined as the fraction of time the brain spends in a suppressed state over a 60 s window. Besides the window, we also define the rules of classification differently and consult the windowed energy instead at the risk of

losing more subtle characteristics. Consequently, a different name is coined to avoid confusion. Although several sources of modulation could exist, we notice that temperature primarily modulates the suppression duration. Therefore, BSDC will vary over time as conditions of the surgery change. Formulating BSDC provides a simple metric through which progress in transitioning out of ECS can be tracked.

BSDC is defined as the fraction of time within a window the brain spends in a high energy state, akin to a pulse width modulation duty cycle. The window length used to compute the proportion is 100 s long. The choice of this length is somewhat arbitrary but is long enough to be representative of the energy time series. In other words, a full BSDC should indicate that there is a high probability that suppression phases no longer occur. Tuning this window length should also be explored in future analysis iterations. Each energy sample $e[n]$ is labeled as a high or low energy state by thresholding. Observations during cooling are used as a model for classifying observations during warming. Artifacts aside, we assume the burst suppression activity during cooling is "ideal", granted no complications developed beforehand. A histogram is first constructed based on the signal energy observed specifically from the start of CPB until circulatory arrest begins. Two peaks, representing either nominal burst or suppression energy, are found. The midpoint between the two energy levels is chosen to be the threshold. An example of this procedure generated from patient 9's data is shown in Figure 2.11. The importance of identifying artifacts beforehand and combing information across channels is highlighted here. Obtaining less corrupted histograms improves our ability to determine a proper threshold and achieve more accurate classification. Lastly, the BSDC time series undergo postprocessing. Whenever an artifact has been identified, we linearly interpolate across them. Those windows are assumed to be short enough such that a linear model suffices. Also assumed is that patients do not drastically deviate from the trending evolution of burst

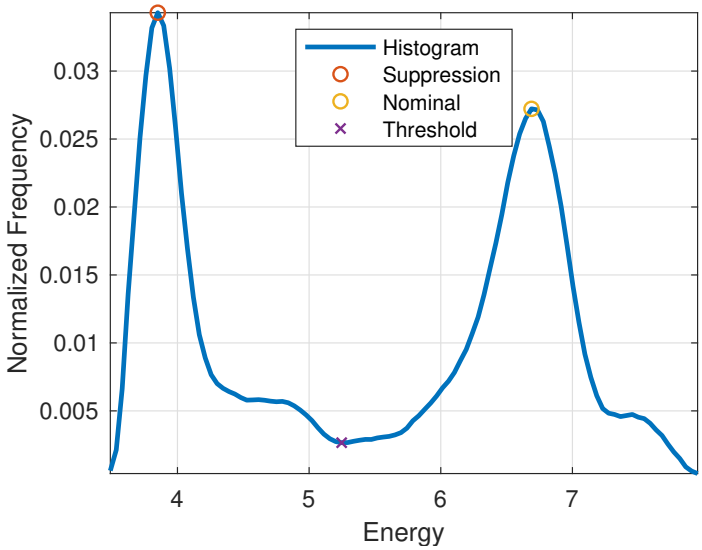

Figure 2.11: Histogram of cooling phase energy for patient 9. Each peak represents average energy for suppression or nominal activity. The midpoint is chosen as the threshold used for labeling each sample a high or low energy state.

suppression dynamics.

The time evolutions of BSDC for all patients since they resume CPB are shown in Figure 2.12 and Figure 2.13. Figure 2.9 and Figure 2.10 should be correspondingly referenced when interpreting the BSDC because this quantity is one layer of abstraction removed from the signal energy, which is already an abstraction of the original signal. When BSDC reads 0%, the patient exhibits ECS. Oppositely, no suppression activity is detected within the 100 s window when the metric hits 100%. We interpret this event as a biomarker indicating that "nominal" function has returned. Burst suppression has begun when the BSDC starts to steadily increase from 0%. Anticipating postoperative delirium involves tracking this transition between 0% and 100% BSDC.

An interesting variety in BSDC progressions are observed. Lengths of the traces denote how much time elapsed until the heart resumes driving circulation. The time between reaching full BSDC and the end of bypass varies, perhaps due to specific needs

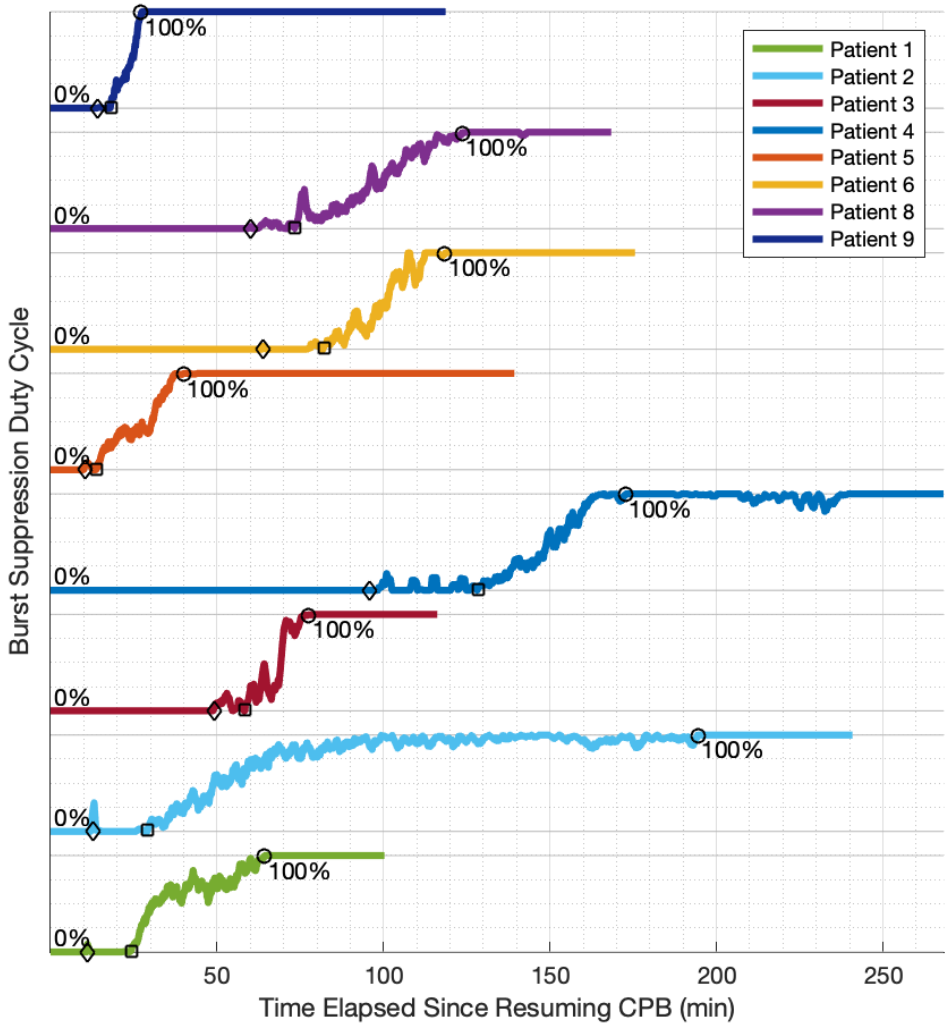

Figure 2.12: BSDC time courses of patients 1, 2, 3, 4, 5, 6, 8, and 9 since resuming CPB. Rewarming is marked by the diamonds. Circles denote T_0, the start of burst suppression. Squares represent T_1, which is interpreted as the end of burst suppression or when a patient's BSDC hits 100% with stability. Patients 2, 4, and 5 belong to the delirium positive group.

of each procedure. There does not appear to be an association with this amount of time and the procedures up to the level of specificity documented in Table 2.1. Time elapsed until warming begins also varies significantly. Some of these delays may be

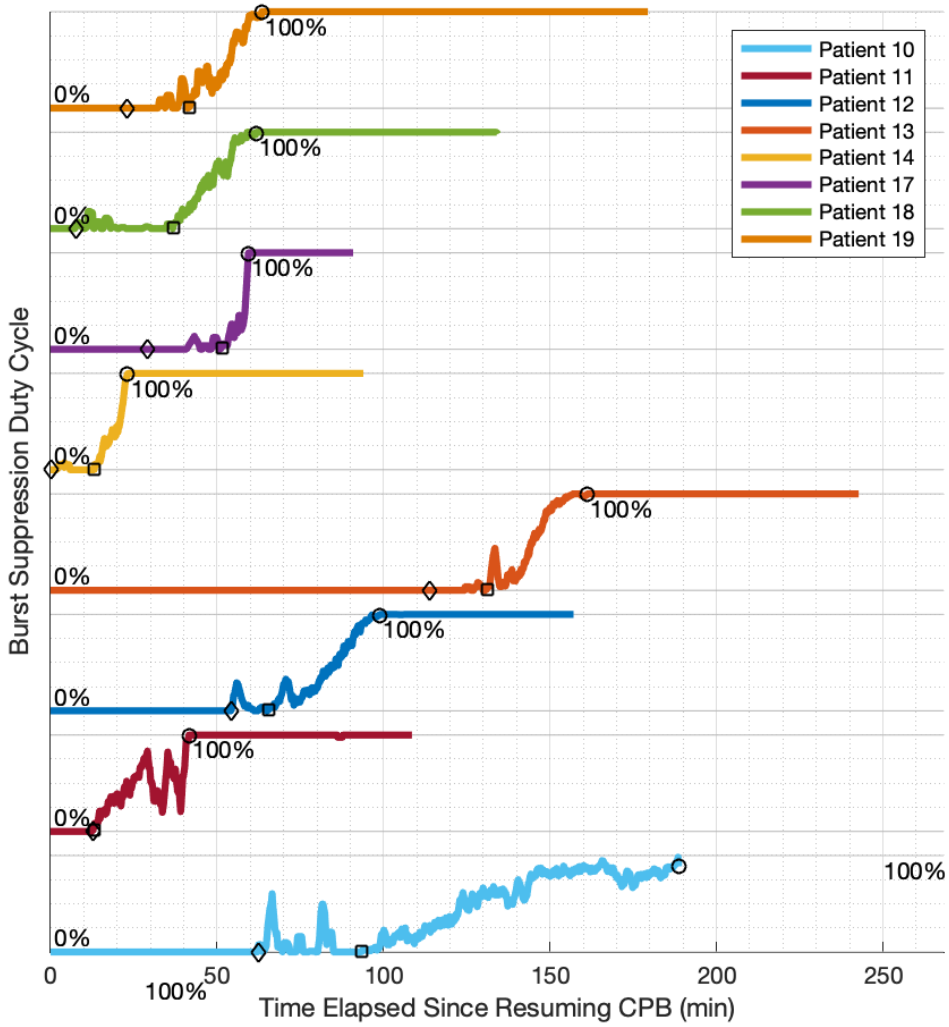

Figure 2.13: BSDC time courses of patients 10, 11, 12, 13, 14, 17, 18, and 18 since resuming CPB. Rewarming is marked by the diamonds. Circles denote T_0, the start of burst suppression. Squares represent T_1, which is interpreted as the end of burst suppression or when a patient's BSDC hits 100% with stability. Patients 10 and 13 belong to the delirium positive group. The end of patient 14's circulatory arrest was not annotated, so we default to using the rewarming time stamp.

attributed to the perfusion strategy. Patients 4, 6, 8, 10, and 12 all underwent both RCP and SACP. We observe that warming tends to begin later than other cases. In cases utilizing both strategies, the switch between RCP and SACP is not explicitly

annotated. Surgeons may have switched patients over to SACP to continue repairs after resuming CPB, which would account for the extended period of time before warming. The last source of variation that can be observed is the progression of BSDC over the course of warming, which is noted by the activity between the diamond and circle markers within Figure 2.12 and Figure 2.13. Many variables likely affect this phase, but body temperature seems to be the primary modulation source.

Several outliers can be identified within Figure 2.12 and Figure 2.13 in terms of these attributes. Patient 2 quickly reaches a high BSDC, just bumping up against 100%. They never manage to achieve full BSDC with stability until almost minute 200. In Figure 2.9 we observe suppression phases up until that same point. Patient 10 does not complete the transition out of burst suppression. In fact, bypass ends before full BSDC is achieved. Patient 13 is considered an outlier case due to the extreme delay before warming. Warming did not start until after some other patients have already completed transitioning from ECS. Consequently, they were late to reach full BSDC. Patient 4 experiences a similar progression, but much of the delay can be attributed to a late start in warming. These patients are all delirium positive, so there appears to be some correlation between the nature of transitioning from ECS to "nominal" activity and the postoperative delirium diagnoses.

Chapter 3

ANTICIPATING POSTOPERATIVE DELIRIUM

3.1 Prediction Procedure

Postoperative delirium forecasts are derived strictly from data recorded starting when CPB resumes. In doing so, we assume that the procedure prior to circulatory arrest did not cause injury. We suspect that prolonging the brain's recovery of nominal function may be detrimental. Specifically, we hypothesize that the transition from ECS (starting when CPB has resumed) to full BSDC takes significantly longer to complete for delirium positive patients when compared to delirium negative patients. By tracking the progress through BSDC, we can determine if the transition is projected to finish on time.

From Figure 2.12 and Figure 2.13 we recognize that the timing of the BSDC progression may provide warning signs. In particular, the time when full BSDC has been achieved is a strong indicator of postoperative delirium. Classifying the data once patients meet full BSDC is technically a prediction, but an earlier indicator is desired because the transition is effectively complete at that point. Whether intervening after achieving full BSDC would be effective remains to be determined, so we aim to accurately infer the outcome before this crucial transition period ends just in case. Consequently, classifying patients based on how fast they meet intermediate milestones is also explored. The prediction accuracy achievable at these earlier points in time determines how much foresight is available.

The general goal is to check if patients have met certain BSDC milestones within an acceptable margin of time. We monitor the time taken to reach the following BSDC

levels since resuming CPB: 1%, 5% to 95% in 5% increments, and 99%. We define T_0 as the time elapsed for BSDC to surpass 1%. This measurement is treated as the point when burst suppression begins. Similarly, T_1 designates the time when BSDC crosses 99% with stability, signaling that suppression activity has ceased. Stability is defined here as having crossed and stayed above the threshold for at least 600 s. This parameter is arbitrary, and there are likely more efficient and robust methods to determine measurement stability. If ever used clinically, visually examining BSDC and making a cautious judgement call may be sufficient if timing is crucial. Out of convenience, the other time stamps are designated correspondingly as $T_{0.05}$, $T_{0.1}$, $T_{0.15}$, etc. Let us define the BSDC milestones examined as $\gamma \in \{0, 0.05, \ldots, 0.95, 1\}$. These measurements for every patient are documented in Figure 3.1. We analyze the relationship between the delirium diagnoses and these milestone timestamps.

A prediction for each patient is made at each BSDC milestone based on T_γ. In other words, the inference about each patients' postoperative delirium diagnosis is updated as their transition progresses. Likelihoods that a measurement T_γ belongs to either diagnosis group inform these inferences. These quantities are given by

$$L(T_\gamma; \mu_{d,\gamma}, \sigma_{d,\gamma}) = (2\pi\sigma_{d,\gamma}^2)^{-\frac{1}{2}} \exp\left(-\frac{(T_\gamma - \mu_{d,\gamma})^2}{2\sigma_{d,\gamma}^2}\right), \tag{3.1}$$

where $\mu_{d,\gamma}$ and $\sigma_{d,\gamma}$ respectively denote the mean and standard deviation of T_γ for diagnosis group $d \in \{\mathrm{P}, \mathrm{N}\}$. Gaussian distributions are assumed to characterize the variation in observations of T_γ. This model is likely erroneous, but without sufficient samples, suggesting a more accurate one cannot be done. Therefore, using this simple, well-studied model capturing general notions of the distribution will suffice. If more data were to be obtained, then more suitable likelihoods could be formulated. By using such a Gaussian model though, we presume that a typical delirium negative progression, described through $\mu_{\mathrm{N},\gamma}$, is different from an average delirium positive pro-

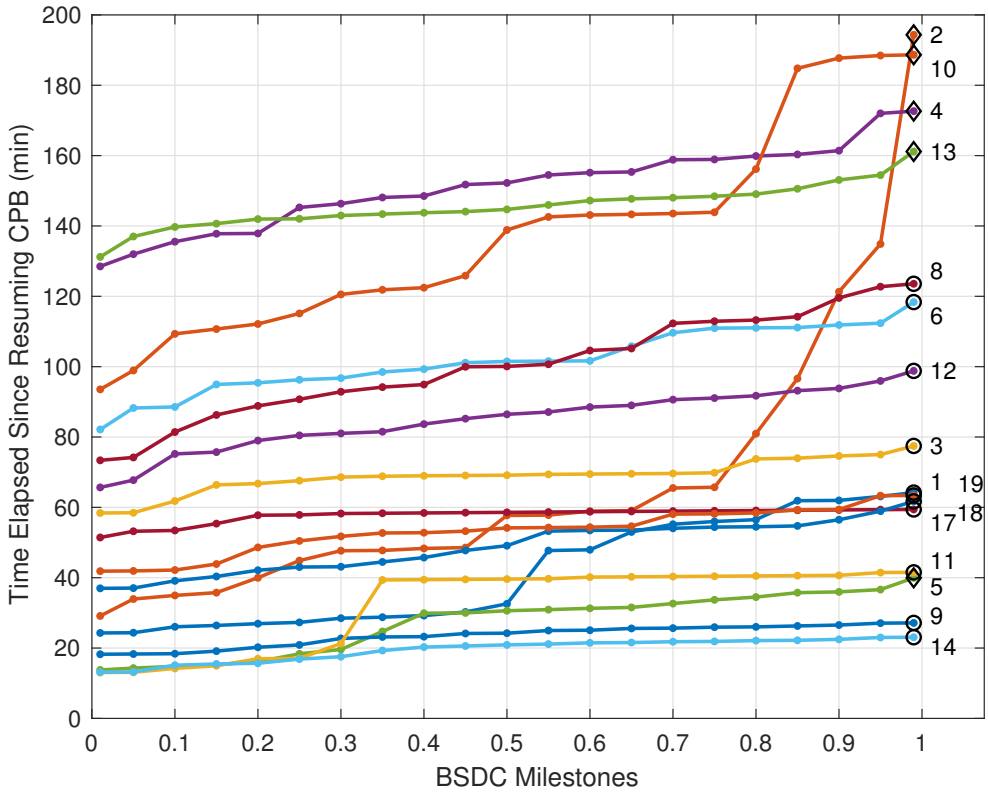

Figure 3.1: Time to reach each BSDC milestone for each patient. Observe at full BSDC that a line can be drawn between a group of four delirium positive patients and the rest. Patients are also classified based on intermediate measurements. Naturally, the prognosis becomes less clear when observing earlier phases of the transition. Each trace is labeled by the corresponding patient ID.

gression, represented by $\mu_{\mathrm{P},\gamma}$. Furthermore, the amount each group varies within can be different, which is characterized by $\sigma_{\mathrm{N},\gamma}$ and $\sigma_{\mathrm{P},\gamma}$. It is important, however, to not place too much significance on the notion of a typical delirium positive because there are many sources of complications that may affect this transition phase. Parameters $\mu_{\mathrm{P},\gamma}$ and $\sigma_{\mathrm{P},\gamma}$ characterize the group of delirium positive cases specific to this study. While monitoring the BSDC, we must determine if the progress transitioning from ECS resembles either a delirium positive procedure or a delirium negative procedure.

The generalized likelihood ratio test is employed to classify the data [43]. This

test is formulated as

$$\frac{L(T_\gamma; \mu_{\mathrm{N},\gamma}, \sigma_{\mathrm{N},\gamma})}{L(T_\gamma; \mu_{\mathrm{P},\gamma}, \sigma_{\mathrm{P},\gamma})} \gtrless \xi \,, \tag{3.2}$$

where ξ is a decision threshold. When the ratio is less than ξ, the null hypothesis, delirium negative, is rejected. Prior knowledge is required to construct the test. Characteristics describing typical delirium negative and delirium positive burst suppression sequences are required. In principle, the prediction procedure can function in real time if the necessary prior knowledge is available. This study is retrospective though. Consequently, we must utilize the data available to us currently. Typically in supervised learning studies, datasets must be partitioned into mutually exclusive sets. One of these partitions is used for estimating parameters needed to construct a classifier, and the other is used strictly to test it. The study population is small though, which makes obtaining non-degenerate parameter estimates difficult after partitioning the dataset. Furthermore, one desired property of these testing and training sets is that they must be balanced in the number of observations of either diagnosis. These issues must be addressed to properly evaluate the methodology.

Bootstrap sampling is employed to artificially increase the size of the training and testing data sets, so that useful estimates of the likelihood parameters and classifier performance can be obtained [44]. This method samples observations with replacement from a sampling pool to create a synthetic dataset. To estimate a parameter using bootstrap sampling, that parameter of interest is estimated from this synthetic dataset. A set of estimates is obtained for multiple realizations of these synthetic datasets. By the law of large numbers, averaging enough of these intermediate estimates produces a refined result that converges toward the true population parameter. Here, the performance of the classifiers is bootstrap estimated.

For each γ, the following sequence occurs for one iteration of the bootstrap supervised learning procedure. First, a synthetic training dataset is generated by bootstrap

sampling from a training set pool containing observations of T_γ. Sampling is controlled such that the number of delirium positive and delirium negative cases are equal. The learning step consists of estimating $\mu_{d,\gamma}$ and $\sigma_{d,\gamma}$. The means are estimated using the maximum likelihood estimator for a gaussian random variable, given by

$$\hat{\mu}_{d,\gamma} = \frac{1}{|\mathcal{S}_d|} \sum_{s \in \mathcal{S}_d} T_\gamma(s) \tag{3.3}$$

where s refers to a member of $\mathcal{S}_d$, the set of patients identified with the diagnosis d. The variances are also estimated through the maximum likelihood estimator for a gaussian random variable, formulated as

$$\hat{\sigma}^2_{d,\gamma} = \frac{1}{|\mathcal{S}_d|} \sum_{s \in \mathcal{S}_d} (T_\gamma(s) - \mu_{d,\gamma})^2 \,. \tag{3.4}$$

After obtaining parameter estimates, the detector described in Equations 3.1 and 3.2 is constructed. A separate synthetic testing set is generated by bootstrap sampling from a test set pool, such that the diagnosis groups are balanced. Fifty samples comprise each of the synthetic datasets. Patient 14, who tested negative for postoperative delirium, is only ever considered to be drawn for the testing group because the time when they resume CPB is not documented. Performance metrics of the detector are obtained by comparing the predictions with the true labels. This entire procedure is repeated 1000 times to obtain a set of intermediate estimates on the detector parameters and various performance metrics. Repeating this supervised learning procedure numerous times is important so that the reported metrics are not consequent of a particular realization of the synthetic training and testing sets.

The parameters are averaged over all the trials and reported. The bootstrap estimated means of $\mu_{\mathrm{P},\gamma}$ and $\mu_{\mathrm{P},\gamma}$ are plotted in Figure 3.2. Delirium positive progressions tend to lag behind that of delirium negative, especially toward the end of burst suppression. Regions within one standard deviation of these means have been marked

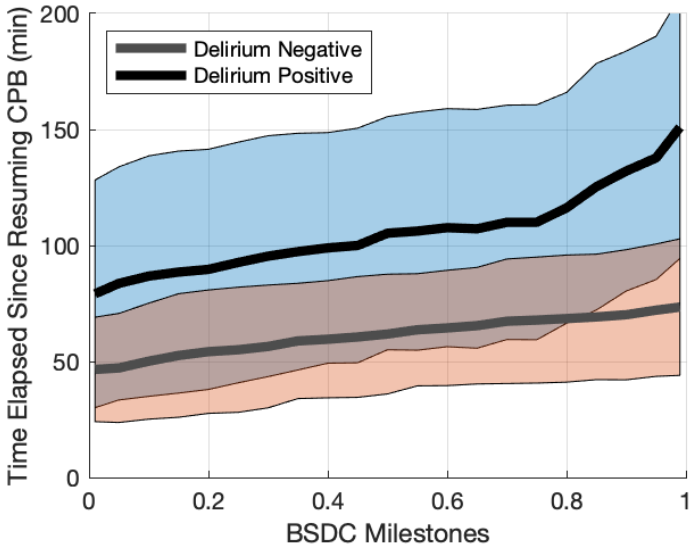

Figure 3.2: Average time required to meet each examined BSDC milestone for each diagnosis group. The mean times were estimated using a bootstrap estimation strategy. Delirium negative patients tend to transition to full BSDC faster than delirium positive patients. Furthermore, the transition tends to start earlier for the delirium negative group.

as well. This plot provides some sense of how the likelihood ratio test might classify each case when examining Figure 3.1. Delirium positive features vary drastically, but delirium negative features tend to vary less. The distinction between the two groups becomes clearer as the BSDC advances, especially toward the end where the intersection has minimized. Classifying the features correctly is easier when these regions are spread further apart.

We note that this procedure is different from what is described in [40]. In this earlier work, a Monte Carlo supervised learning simulation was implemented. A bootstrap method was not utilized. Instead, the dataset was repeatedly repartitioned into mutually exclusive training and testing sets, obeying the same caveat on patient 14. Only delirium negative patients are used in training though, which means prior knowledge of a delirium positive procedure is not considered. Similarly, a series of learning and testing is performed, producing a set of observed performance metrics. A

major benefit of this methodology is that some information is entirely withheld from the test set, so the robustness of the detector toward missing information is tested. A robust and accurate detector was constructed, but the detector has no frame of reference for how deviant cases can be. The lack of prior knowledge about delirium positive progressions results in a classification bias, even when bootstrap sampling is utilized to balance the test set. Thus, we employ the likelihood ratio test instead to assess the efficacy of the methodology after the bias in classification is removed.

During each trial, prediction accuracy, true positive rate (TPR), and false positive rate (FPR) are recorded for every BSDC milestone while varying ξ. The TPR is given by

$$\mathrm{TPR} = \frac{\mathrm{TP}}{|\mathcal{S}_{\mathrm{P}}|}\,, \tag{3.5}$$

where TP is the number of correct delirium positive predictions given threshold ξ. Whereas , the FPR is given by

$$\mathrm{FPR} = \frac{\mathrm{FP}}{|\mathcal{S}_{\mathrm{N}}|}\,, \tag{3.6}$$

where FP is the number of incorrect delirium positive predictions given threshold ξ. At the conclusion of the bootstrap supervised learning procedure, the observed accuracy measurements over all trials are averaged to obtain the reported estimated accuracy. The maximum mean accuracy observed over all choices of ξ at each BSDC γ is reported. All observed TPRs and FPRs are also averaged over all trials. The collection of FPRs and TPRs parameterized by the detection threshold ξ comprises a receiver operating characteristics (ROC) curve. This curve is a classical detection theoretical tool describing the performance of a detector as the decision threshold changes. A more comprehensive assessment of the detector is provided through the ROC curve than the accuracy, which is consequent of a singular detection threshold. The ROC curve provides a tool for the selection of ξ in the event that the detector

must be designed to satisfy certain requirements on the FPR. This design concept may be applicable in the case of these surgeries depending on the safety of the intervention. A ROC curve is produced for each γ examined.

The area under the curve (AUC) is a singular quantity that summarizes the ROC curve's shape. Computing the AUC, as the names suggests, involves integrating the ROC curve. The AUC is at maximum 1, which means the detector is never wrong for some choice of ξ when a detection is made. The worst AUC is 0.5 meaning the detector functions as a random chance detector. AUCs can be lower than 0.5, but this scenario simply means the test statistic can be reversed, though this may be nonsensical. One goal in constructing any detector is to aim for an AUC of 1, but this is practically unlikely. We observe how the AUC changes as a function of γ.

In addition to each ROC curve, its convex hull is also constructed. The convex hull of the ROC curve is a valid ROC curve, achievable in the statistical sense [45]. To produce this performance, decision thresholds corresponding to points on the ROC curve comprising its convex hull must be identified. Altering the proportion of time spent operating at the hull point thresholds sweeps out the hull line. Thresholds producing parts of the ROC curve inside the hull are suboptimal. The area under the hull (AUH) is equal to or greater than the AUC. The AUC and AUH are expected to improve as predictions are made later in the BSDC progression (i.e. as γ increases).

3.2 Performance Assessment

First, we observe the prediction performance achieved when the training and testing set pools, from which synthetic datasets are generated, both contain all available observations for the particular γ under examination. We pretend that the two synthetic datasets are two different study populations. They may share many of the same unique observations, but the composition is likely different. In this scenario,

the diversity in observations within each synthetic data sets is maximized.

Mean prediction accuracy as a function of the BSDC milestones is plotted in Figure 3.3. The first standard deviation above and below the mean is provided as well because the evaluation procedure is stochastic. As burst suppression begins, 78.10% can immediately be predicted correctly. Generally, in Figure 3.3 the accuracy remains approximately flat, creeping towards 80% accuracy for most of the transition. Not until passing 85% BSDC does the accuracy start to increase significantly. A maximum mean accuracy of 89.74% is observed when burst suppression has ceased. Accurate prediction has been achieved when monitoring the progression of BSDC once CPB has resumed. We notice that with the adjustments in the testing methodology, a curve similar to that which is presented in our previous work [40] was produced. The difference here is an improvement in accuracy at earlier stages in the BSDC progression.

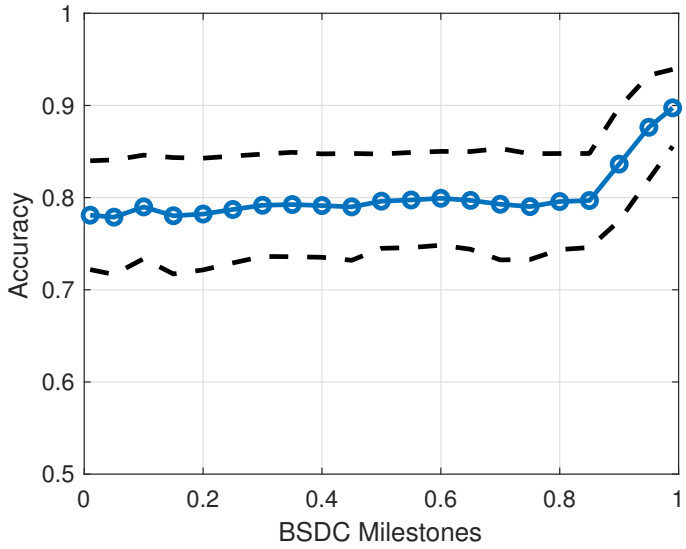

Figure 3.3: Maximum mean accuracy observed at each BSDC milestone over all choices of threshold ξ. The accuracy one standard deviation above and below are also provided. Generally, accuracy improves as predictions are made later along the transition to full BSDC. Accurate predictions are achievable by tracking BSDC progress.

Referring to Figure 3.1, we can easily understand the trend observed in Figure 3.3. Patient 5 is as a false negative across all choices of BSDC. In some sense, they may be considered a deviant of the delirium positive group in terms of post-DHCA BSDC measurements. Patients 4, 10, and 13 are easily distinguished from the rest of the patients. Their features are outliers from the first burst until burst suppression has concluded. Patient 2's progression does not begin to clearly deviate until 90% BSDC. Their progress begins to significantly slow down starting at 75% BSDC though. All other patients do not veer off track in the way that patient 2 or even patient 10 did. The 80% accuracy refers to situations where patients 2 and 5 are misclassified in these balanced test sets. Prediction accuracy improves to 90% because the status of patient 2 has officially deviated significantly. The gradual increase is attributed in part due to the bootstrap sampling.

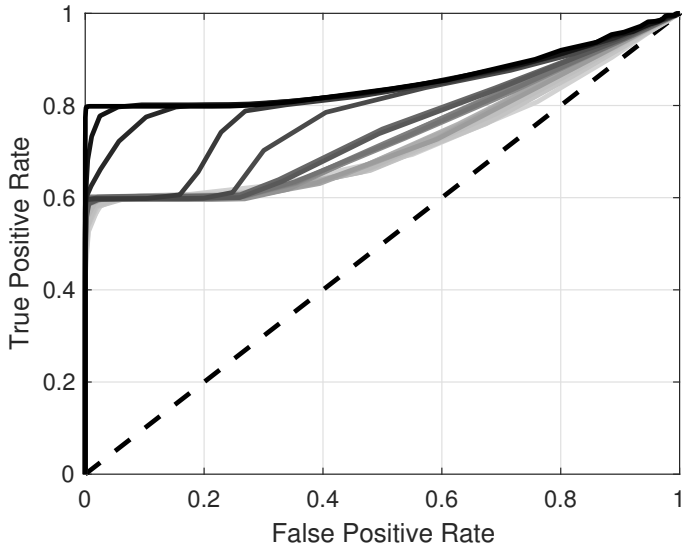

Figure 3.4: ROC curve produced when making predictions based on each BSDC milestone. The shade of each curve denotes how advanced the transition is. The darkest line and incidentally the ROC curve with the largest AUC represents the performance achieved when predicting based on T_1. Similarly, the lightest shade represents the performance curve achieved when the prediction is made at the earliest juncture.

ROC curves produced from the bootstrap supervised learning simulation are presented in Figure 3.4. The gradient in line color represents how advanced into the progression patients are. The darkest shade represents performance once patients are at full BSDC. Oppositely, the lightest shade denotes the classifier based on timing of the first burst. We note that as the BSDC advances, the ROC curve area increases, which is sensible because the status of the patient should become clearer as time passes. Additionally, the TPR increases from 0.6 to 0.8 with γ in a 0 FPR region, where the darkest three lines correspond with the last three points in Figure 3.3.

A more optimistic estimate of the detector performance characteristics are represented in Figure 3.5 through the convex hulls of the ROC curves. We observe a significant increase in the area covered across each BSDC milestone monitored. If the situation arises at all where a worse FPR could be traded for a better TPR, then we observe the benefits in performance that could be theoretically achieved. This means the viability for using earlier milestones increases.

To concretely convey the improvement in prediction performance as time passes, the AUC and AUH as a function of BSDC milestones is plotted in Figure 3.6. A maximum AUC of 0.8557 was achieved at full BSDC and the minimum 0.7196 was achieved at the first burst. The corresponding AUHs observed were 0.8988 and 0.7889. The AUCs and AUHs are closer in agreement when predicting from 75% BSDC and higher than otherwise. After accounting for sources of bias in the classifier construction and testing procedure, slight improvements over the performance documented in [40] are observed.

The observed performance demonstrates that not only can accurate predictions be made, but they can also be produced before burst suppression even ends. Making predictions earlier incurs a penalty in accuracy, naturally. Moderately accurate predictions can be made as soon as burst suppression begins until 80% BSDC, be-

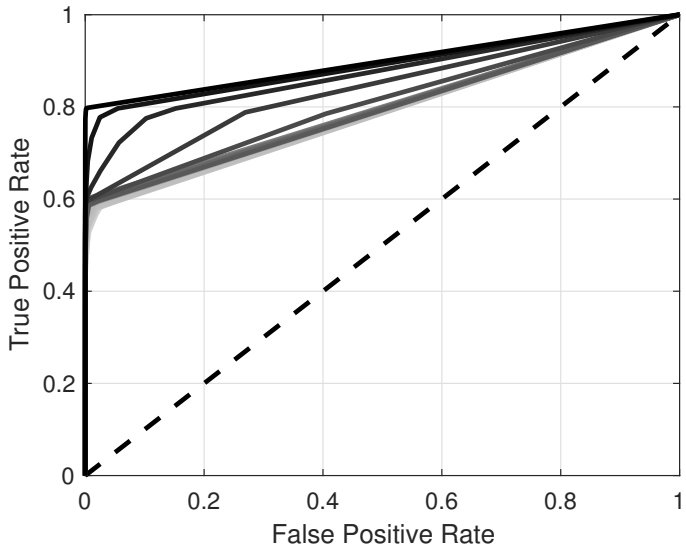

Figure 3.5: Convex hull of the ROC curves, statistically achievable by time sharing amongst various thresholds. The shade of each curve denotes how advanced the transition is. The darkest line and incidentally the ROC curve with the largest AUC represents the performance achieved when predicting based on T_1. Similarly, the lightest shade represents the performance curve achieved when the prediction is made at the earliest juncture.

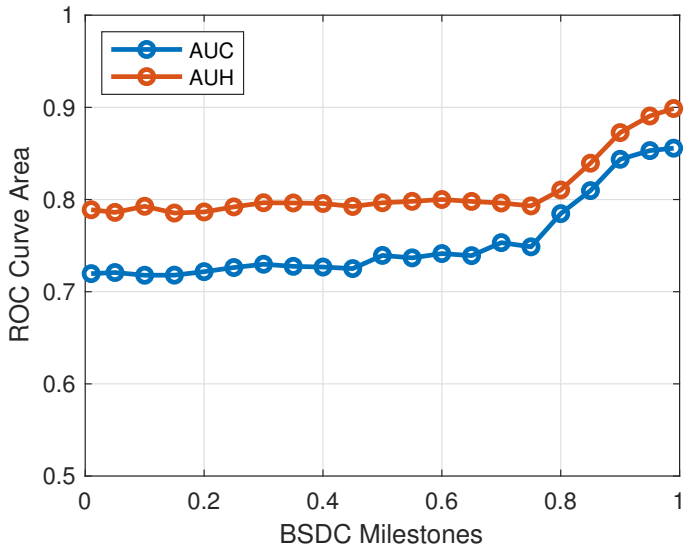

Figure 3.6: AUC and AUH as a function of the BSDC milestone. The robustness of the detector improves when the BSDC progression is queried later. The convex hull of the ROC curves demonstrate an improved viability in making predictions earlier in the burst suppression.

yond which the accuracy and robustness of the classifiers improve considerably. The accuracy achieved demonstrates a correlation between the speed at which burst suppression dynamics evolve as measured with respect to when CPB resumes.

Results obtained when using the two dimensional feature vector described in [40] with the revised learning and testing method does not yield accuracy as strong as T_γ alone. Hence, why this work only documents results produced using a single feature, which encodes approximately the same information conveyed by the two dimensional feature vector used previously. The benefits of using additional measurements would need to be assessed with a revised set of features.

3.2.1 Sampling Pool Disjointedness

In typical supervised learning settings, training and testing sets must contain mutually exclusive observations. Of course, this is not strictly satisfied in the sampling variation first presented. Fortunately, the learning stage only requires estimating means and standard deviations, which depends in part on the composition of the generated training dataset. Contrast this with constructing a classifier that is sensitive to the precise observations, such as a nearest-neighbor classifier. We must address how changing the overlap in sampling pools affects prediction performance.

Consider first the scenario in which the testing and training sampling pools are completely disjoint. The synthetic datasets will contain mutually exclusive unique observations after the bootstrap sampling procedure. To accomplish this while enforcing the balanced groups property, three observations from each diagnosis group are placed into the training sampling pool, and two observations from each diagnosis group are placed into the testing sampling pool. The dataset generation, training, and testing procedure commences afterward. For each iteration of the simulation, the sampling pools are redrawn. The average performance for this variation is evaluated.

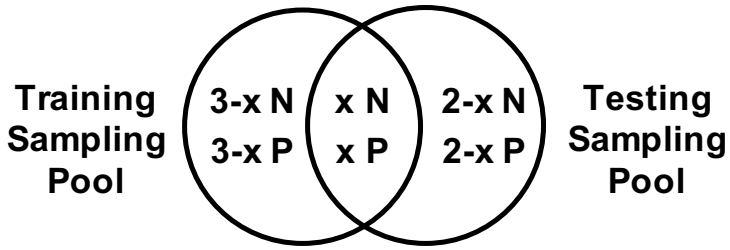

Figure 3.7: Analysis variations for testing the effects of sampling pool overlap on prediction performance. The variable x denotes the number of samples the sampling pools share in common. From these sampling pools, the synthetic datasets used for training and testing are drawn through bootstrap sampling.

AUCs and AUHs are used as a point of comparison. Keeping the same number of observations within each sampling pool, the average performance is evaluated while adjusting the number of samples common between pools. The sampling pool overlap variations are summarized in Figure 3.7, where x denotes the number of samples in common between either pool.

First, the AUC and AUH as a function of the BSDC milestone achieved for the completely disjoint case is plotted in Figure 3.8. Both area measurements still increase as the transition advances. The increase is inconsistent though. Additionally, far poorer performance is observed. The AUCs are not much better than a random chance predictor until later in the transition. The convex hull provides an immediate significant increase in the area, but the AUHs still show modest performance. Achieving promising results with only three unique observations per diagnosis in the training set and two unique observations per diagnosis in the testing set is perhaps overly demanding though. Second, the metrics achieved when the sampling pools have one case in common is plotted in Figure 3.9. Again, the general trend in improving accuracy as the transition advances is observed. Furthermore, a substantial improvement is seen in the AUCs, especially in the earlier stages of the transition. The AUH does not provide much more optimism in this case. Lastly, the AUC and AUH as a function of the BSDC milestone achieved where the sampling pools had two

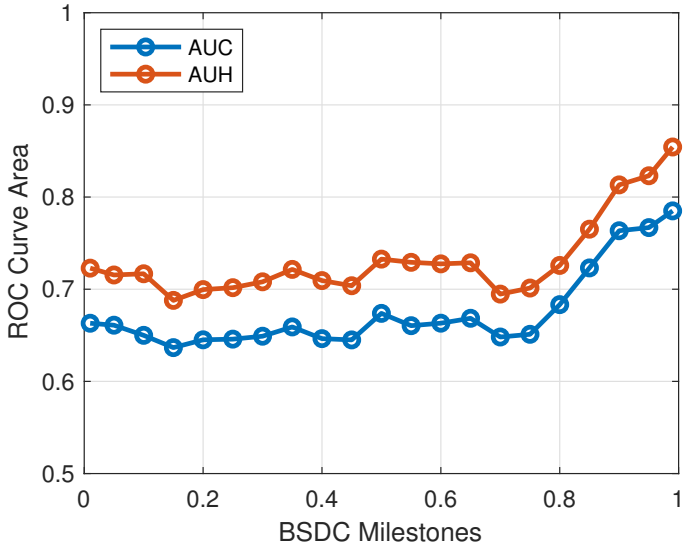

Figure 3.8: Areas under the ROC curves and their convex hulls when sampling pools are completely disjointed. The same improvement trend is observed, but performance is significantly worse.

cases in common is plotted in Figure 3.10. Another substantial improvement from Figure 3.9 in the areas is observed here. The performance observed in this scenario is quite close that which was observed in Figure 3.6. The AUCs and AUHs tend to agree with each other better in this testing variation though. The AUHs are lower in the early stages of the transition here than in Figure 3.6, but improve to approximately the same by the end phase. Comparing Figure 3.10 and Figure 3.8, we see quite a large discrepancy in performance—almost a 0.1 unit area gap. Extrapolating to Figure 3.6 and the rest of the performance metrics reported, the performance is overestimated when both sampling pools were comprised of the entire dataset. Presumably, however, performance observed in Figure 3.6 should be achievable while enforcing disjoint sampling pools if more data were obtained. Achieving even better performance after obtaining more data is plausible too.

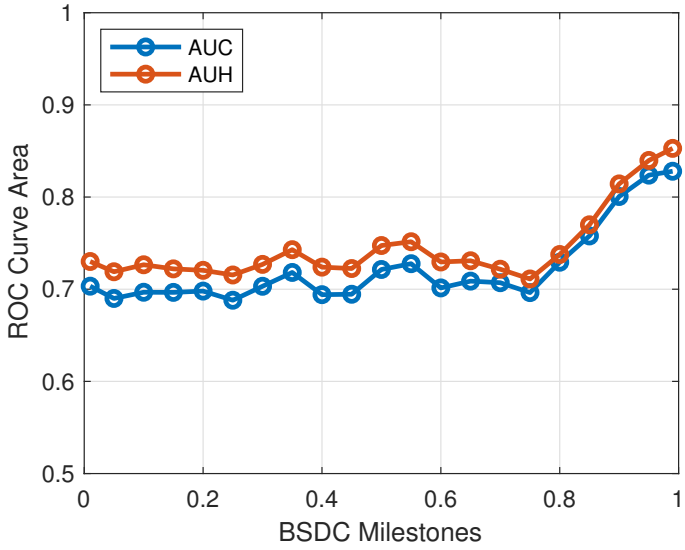

Figure 3.9: Areas under the ROC curves and their convex hulls when sampling pools contain one common case. A significant improvement over the completely disjoint variation is achieved.

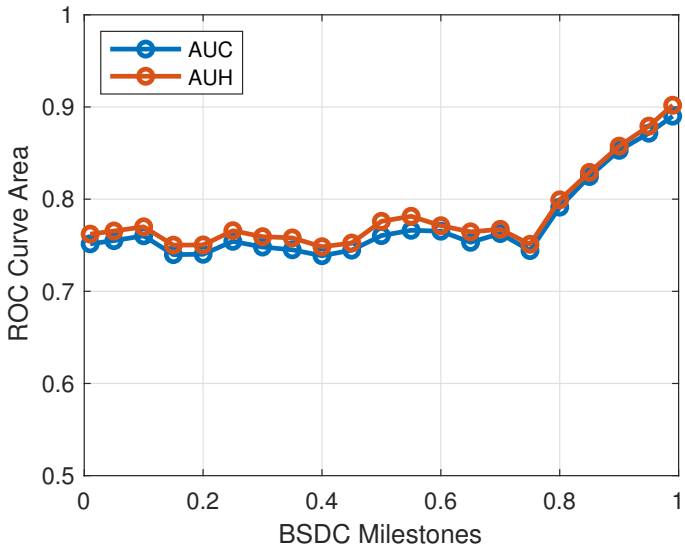

Figure 3.10: Areas under the ROC curves and their convex hulls when samplings pools completely overlap. Another significant improvement over the single overlap variation is achieved. Overlapping the sampling pools causes the performance to be overestimated.

3.3 False Classifiers

As is the case with any statistical learning problem, there is a concern that the performance achieved is a fluke. Another worry is that the methodology would function just as well no matter how the data was labeled, meaning the data is not actually informative. The small size of our dataset only compounds on these concerns. The bootstrap method for assessing our methodology only partially addresses the sensitivity to the data size. Consequently, we consider false models to clarify these questions. We construct false models by randomizing the training set labels before the learning phase. Thus, $\mu_{d,\gamma}$ and $\sigma_{d,\gamma}$ are all mischaracterized. After generating this erroneous model, the detector is evaluated by testing properly labeled data. Observing accurate predictions even with these false models would be alarming and indicate the presence of possibly several problems.

One reason accurate classification could be achieved using a false classifier is the existence of a bias, which was the problem in the strategy described in [40]. This bias can be found in either the dataset or the classifier construction. Bootstrap sampling was incorporated to correct for an inherent bias that can result from the distribution of diagnoses. The performance metrics will incur an inherent inflation or deflation in imblanced datasets. For instance, if a diagnosis group that tends to be correctly classified was overrepresented, then the classification accuracy would be overestimated. The creation of balanced datasets, facilitated by bootstrap sampling, helps correct for this problem. Using the likelihood ratio test attempts to correct for a bias in the classifier because knowledge from both diagnosis groups informs its construction. Additionally, the repetition and redrawing of datasets involved in the bootstrap method removes the sensitivity to particular drawings that cause flukes. Lastly, one other way predictions can still be accurate using these false detectors is an

excessive amount of degrees of freedom. If the dimensionality of the data were large enough, then constructing some manifold that still classifies the data appropriately may be possible. This scenario is not an issue here because the feature on which learning and testing is based is one dimensional. Ideally, these false models would perform as if decisions were sampled as a Bernoulli random variable and an AUC of 0.5 would be observed.

We must verify the observed performance was not a fluke by confirming the performance using false classifiers is poor. The maximum—over all choices of thresholds—mean accuracy achieved at each BSDC milestone is plotted in Figure 3.11. We observe consistent accuracy over all milestones. Making a prediction at any one milestone is no better than predicting at any other milestone. Furthermore, that accuracy is poor, where just shy of 60% of the cases were correctly classified. The maximum mean accuracy observed is 57.62%, occurring at the first burst. These false classifiers perform just slightly better than randomly assigning labels to the test dataset. Although not closer to 50%, the accuracy achieved using false models helps validate the methodology.

For brevity, only the convex hulls of the ROC curves are presented. The curves in Figure 3.12 hug closely to the 0.5 AUC boundary. Ideally, the curves would rest on top of the random labeling detector curve represented by the dashed line. The raw ROC curves actually fluctuate above the line before 0.5 FPR, and fluctuate under beyond it. Using the convex hulls of the ROC curves makes validating the methodology more difficult because AUHs are at least as large as AUCs. Constructing the convex hulls does not provide the false classifiers much benefit in this case though.

Both the AUCs and AUHs observed at each BSDC milestone are documented in Figure 3.13. The AUC is consistently averaging approximately 0.5, which is ideal. The highest achieved was 0.5082, whereas the lowest was 0.477. Due to the structure

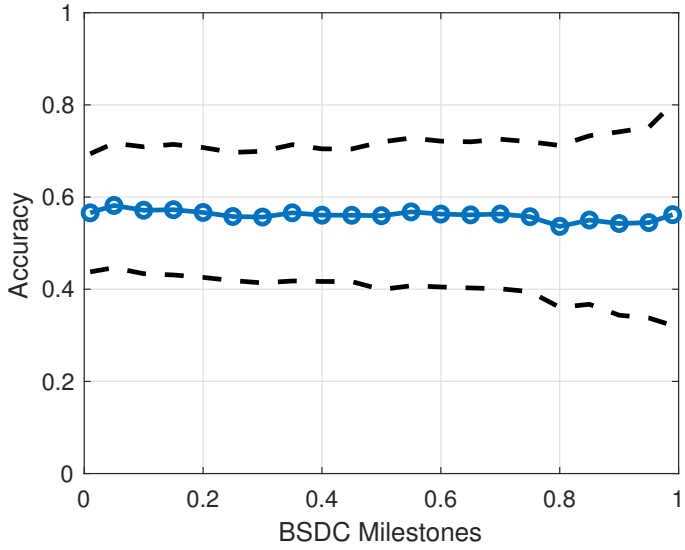

Figure 3.11: Accuracy of the classifiers when labeling of the training datasets have been randomized. Querying one milestone performs no better than any other. Most importantly, the accuracy is approximately 60%. The detectors do not function precisely as if random labels were assigned to the test set. Accuracy of these detectors is still poor though, which helps validate the methodology.

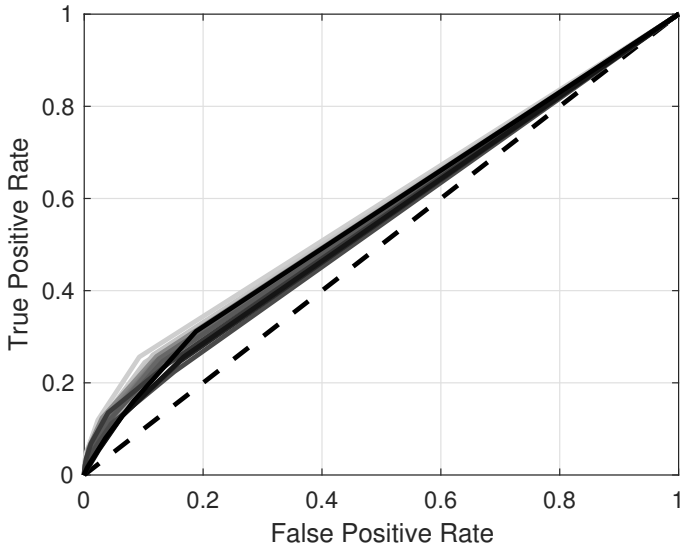

Figure 3.12: Convex hulls of the ROC curves produced when classifier construction is informed by mislabeled training data. These curve do not provide much more of an optimistic evaluation of the classifiers, which is desired in this case. The hull line still hugs close to the 50% AUC line.

of these ROC curves, the convex hull yields a slight gain in the area. The largest AUH achieved is 0.5974, and the lowest is 0.5484. Although the AUHs achieved are not closer to 0.5, the observed performance helps validate our methodology. Additionally, the performance of the properly trained classifiers is far better than that which is observed here.

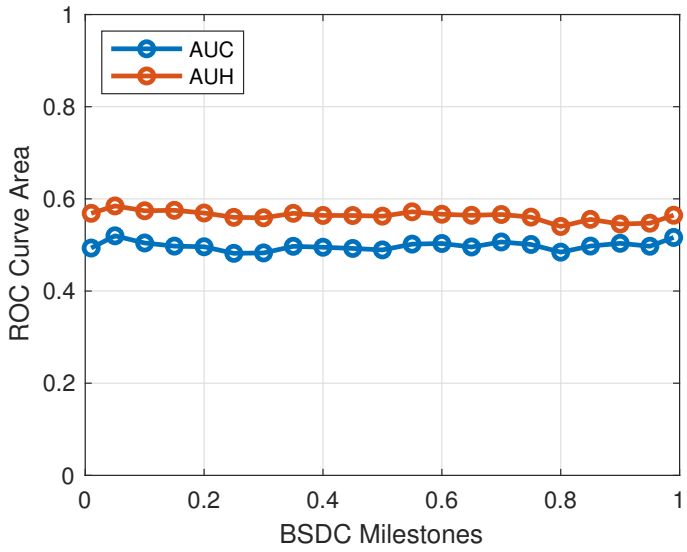

Figure 3.13: AUC and AUH as a function of BSDC milestones. The AUC for the false construction is approximately 50% over all milestones. The convex hull provides just a small improvement, where the AUH is approximately 60% across the entire progression. Predicting at any one BSDC milestone is no better than predicting at any other.

Revising the learning and testing methodology has minimized classification bias. The learning and testing methodology is now sound, and we have obtained an honest assessment of the prediction procedure. When this label scrambling analysis was performed with the procedure described in [40], the classifiers still performed well—perhaps as well as when properly trained classifiers were used. The poor classification accuracy achieved here using false learning demonstrates that the data is in fact informative. Further, the classifier structure learned based on properly labeled data effectively captures information required to make accurate predictions.

Chapter 4

SUPPLEMENTARY ANALYSIS

4.1 The Influence of Perfusion Techniques

The relationship between other attributes of the cases and the postoperative delirium diagnoses must be determined. Understanding these other relationships will clarify the significance of making predictions by monitoring the BSDC. We are particularly curious about the relationship between the perfusion strategy and postoperative delirium. Observing a correlation would suggest that the diagnoses may be consequent of the surgical techniques rather than complications in the neurophysiology—though, we recognize from previous studies that different levels of efficacy are associated with the various neuroprotection strategies. Many sources of complications (i.e. confounding variables) exist. Performing this analysis would determine if the perfusion strategy is one of them.

We first determine whether or not the rate at which the BSDC evolves is indicative of the perfusion technique. The correlation strength between milestone timestamps and perfusion strategies is determined by setting up a classification problem. The same bootstrap learning and testing strategy used when predicting the delirium diagnoses based on T_γ is employed, but the datasets are labeled based on the perfusion strategy instead. We expect there to be a nontrivial correlation between these two variables because maintenance incorporating both RCP and SACP naturally delays warming.

The mean accuracy achieved at each BSDC milestone, maximized over all threshold choices, is plotted in Figure 4.1. The proportion of cases correctly classified begins

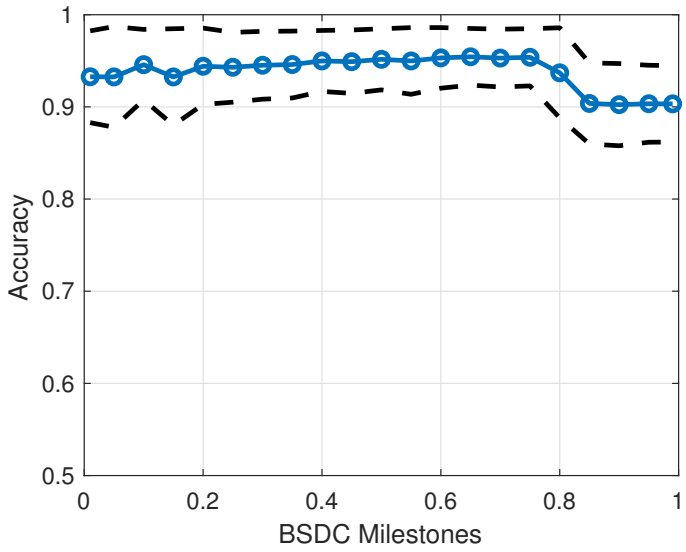

Figure 4.1: Mean accuracy of using T_γ to predict the perfusion strategy employed. The accuracy plus or minus one standard deviation is also plotted. The predictions are even more accurate than attempting to predict incidences of postoperative delirium. A strong correlation between the perfusion strategy used and BSDC progress is confirmed.

at 93.27%, which is already accurate. This accuracy is even higher than when BSDC progress is used to predict postoperative delirium. Interestingly, the accuracy begins to decrease at 75% BSDC. At full BSDC, the accuracy is 90.33$. Accuracy decreases because not all cases taking longer to reach full BSDC underwent both forms of cerebral perfusion. These cases are patients 2 and 13.

Convex hulls of the ROC curves yielded when testing these classifiers are presented in Figure 4.2. The achieved AUCs of the raw ROC curves and AUHs associated with their hulls are presented in Figure 4.3. Again, we observe that the classifiers are accurate and robust across all BSDC milestone polled. The AUC and corresponding AUH once burst suppression begins are 0.9227 and 0.9444. Once burst suppression ends the AUC and AUH are 0.8582 and 0.9055, respectively. In conjunction with Figure 4.1, these results indicate that the time to reach each BSDC milestone is associated with the perfusion technique, but outliers can occur.

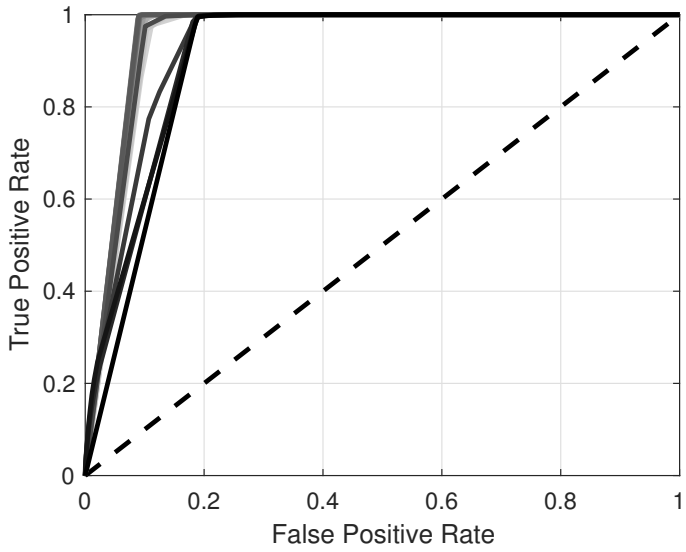

Figure 4.2: Convex hulls of the ROC curves produced when classifying cases according to the perfusion strategies used based on time elapsed to meet BSDC milestones since resuming CPB. The hulls deviate far from the 0.5 AUC line. In fact, these curves demonstrate a stronger classification accuracy than what was achieved when predicting postoperative delirium.

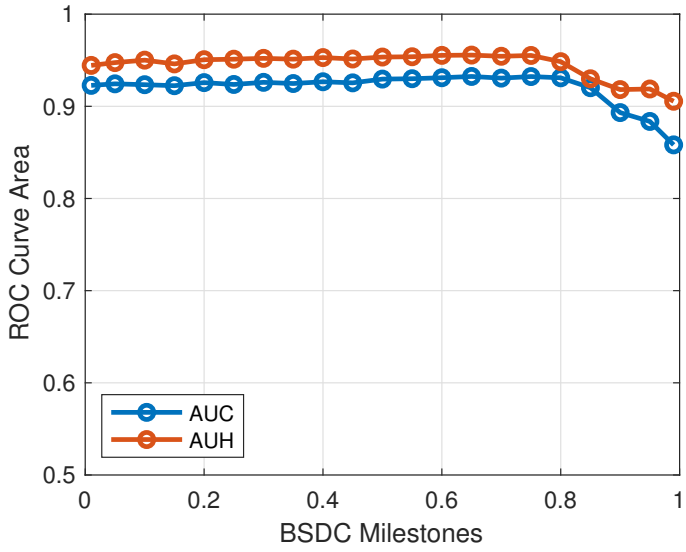

Figure 4.3: AUC and AUH as a function of the BSDC milestone polled associated with the ROC curves produced when predicting perfusion strategy. The classifiers become more accurate as BSDC increases.

Assessing the correlation strength between perfusion strategy and postoperative delirium involves formulating a classification problem. Balanced test sets are generated through bootstrap sampling. If RCP and SACP are used in a case, then that patient is predicted to be diagnosed with postoperative delirium. Otherwise, that patient is predicted to be delirium negative. The rationale behind this decision strategy is that cases utilizing both RCP and SACP tend to prolong patients meeting full BSDC. This trend makes performing this analysis especially important because the postoperative delirium prediction decision depends on observing if the transition out of ECS is progressing quickly enough. No learning is involved in this test, and a ROC curve cannot be generated because there is no decision threshold that can be adjusted. Therefore, the accuracy will determine the correlation. A low correlation would be indicated by poor classification accuracy.

The mean accuracy, within one standard deviation, achieved when classifying cases based on perfusion strategy is $0.5651 \pm 0.0667\%$. This accuracy performs similarly to a random labeling detector. Therefore, the perfusion strategy is a poor predictor of neurological outcome. According to Table 2.1, only two of the five cases employing both perfusion techniques were delirium positive. Based on the earlier reported findings in this study, different perfusion techniques could present situations that drive the features to indicate increased risk because restoring nominal function becomes prolonged. It is important to remember though that such situations may simply be required due to the extent of the necessary repairs.

The perfusion strategy classification based on BSDC progress reveals a strong correlations between the two variables—perhaps even stronger than the correlation between BSDC progress and postoperative delirium. Despite this, we observe that the perfusion strategy is minimally predictive of the diagnosis. Therefore, the perfusion strategy is not a confounding variable in this study, provided the caveat that these

results are made based on a small study population. The perfusion technique is not necessarily a confounding factor because progress can lag due to reasons besides the perfusion technique. For now though, we are comfortably able to further validate the monitoring of a patient's BSDC to predict postoperative delirium.

4.2 Descriptive Statistics

We compute descriptive statistics about the timing of other surgical events to determine if any of them are correlated with postoperative delirium. These measurements are documented in terms of standard deviations away from the delirium negative group's mean. Note that patient 14 was excluded from all benchmark computations requiring the time when CPB resumes. Examining these quantities in relation to diagnosis groups elucidates other potential confounding variables that may contribute to incidences of postoperative delirium.

Durations of several different phases of DHCA, including cooling, circulatory arrest, transitioning to full BSDC (T_1), and burst suppression ($T_1 - T_0$), are reported in Table 4.1. Means and standard deviations of each of these durations are listed in the table footnotes. The patients have been grouped by diagnosis so that the quantities can be interpreted better in the context of postoperative delirium.

The cooling phase is defined here as the time from when body temperature starts being actively affected until circulatory arrest begins. These estimates were derived from the annotations provided. The mean and standard deviation for cooling are 53.1 min and 5.8 min, respectively. Aside from the two outlier cases, patients 4 and 5, a distinction between diagnosis groups in terms of this measurement cannot be drawn. Examining this phase is important because it resembles a dual to T_1, which is well-correlated with postoperative delirium.

Some studies have suggested a correlation between the duration of circulatory

Table 4.1: Duration of DHCA Phases

Patient	Cooling (σ^{a})	Arrest ($\sigma^{b\dagger}$)	T_1 ($\sigma^{c\dagger}$)	$T_1 - T_0$ (σ^{d})
2*	0.508	-0.466	3.847	10.188
4*	6.541	-1.358	3.156	1.361
5*	2.867	-0.191	-1.070	0.053
10*	1.898	2.173	3.666	5.077
13*	-1.736	1.165	2.790	0.328
1	-0.045	-0.787	-0.298	1.053
3	-0.337	-1.324	0.124	-0.466
6	0.144	1.995	1.427	0.784
8	1.170	1.274	1.593	1.803
9	1.844	-0.265	-1.478	-1.207
11	-1.306	-0.531	-1.020	0.220
12	0.124	0.467	0.804	0.559
14	-1.174	2.194	-1.608	-1.132
17	-1.097	0.108	-0.451	-1.275
18	0.869	-0.474	-0.378	-0.054
19	-0.194	-0.463	-0.324	-0.287

* Delirium positive
† Computed without patient 14
[a] $\mu = 53.1$ min, $\sigma = 5.8$ min
[b] $\mu = 17.7$ min, $\sigma = 15.3$ min
[c] $\mu = 73.6$ min, $\sigma = 31.4$ min
[d] $\mu = 25.5$ min, $\sigma = 13.7$ min

arrest and poor prognosis [7, 8]. The mean circulatory arrest duration is 17.7 min for the delirium negative group, and the standard deviation is 15.3 min. The duration for which circulatory arrest was induced in patient 14 is over estimated because the

moment when CPB resumed was not documented. The time stamp used instead was when warming started, which of course begins after resuming CPB. As a result, they actually deviate less from the mean. Despite this discrepancy, a distinction between the two diagnosis groups based on the normalized measurements is not clear. This conclusion is supported by a moderate variation in most delirium positive patients from the delirium negative mean. In fact, some members of the delirium negative group varied more than patients in the other group.

The designation T_1 represents the time elapsed until burst suppression stops since resuming CPB. This quantity is treated as the dual to the cooling phase duration because the measurement starts when CPB resumes and lasts until full BSDC, when we believe sufficient warming has been achieved. Note, this quantity is not formally treated as the warming phase because measurements do not begin at the warming time stamp. The mean value of T_1 duration is 73.6 min and the standard deviation is 31.4 min. All diagnoses but one (patient 5) can be correctly predicted by considering T_1 alone. Patient 5 is a false negative in this case. The correlation between T_1 and the postoperative delirium was already established in the bootstrap supervised learning analysis.

Burst suppression duration was observed to be predictive of postoperative delirium for cardiac surgeries that did not employ DHCA [38]. A longer than normal burst suppression duration was hypothesized to be moderately correlated with postoperative delirium, where the AUC achieved in the ROC curve was 0.73 [38]. In this study, the mean duration of burst suppression for delirium negative cases is 25.5 min, and the standard deviation is 13.7 min. oOnly two clear outliers with respect to typical delirium negative measurements are found. These cases are patients 2 and 10 according to Table 4.1. On the other hand, some delirium negative cases deviated from the mean more than the remaining delirium positive cases. Patient 8's

burst suppression duration might be considered long for a delirium negative case, but perhaps not significantly. Burst suppression duration is just one element of a more correlated measure.

Durations of various extracorporeal support phases are documented in Table 4.2. These periods include total extracorporeal support, overall time spent on CPB, CPB prior to circulatory arrest, and CPB following circulatory arrest. Total extracorporeal support includes overall CPB and circulatory arrest as metabolic support was still provided during that time. Again, delirium negative means and standard deviations of the measurements are listed in the table footnotes. Several potential confounding variables are identified in this table.

Krähenbühl et al. document a correlation between CPB duration and temporary neurological dysfunction incidences [10]. We document the same metrics here to observe if the same correlation is observed for this particular set of surgeries. Average duration of total extracorporeal support duration is 218.7 min with a standard deviation of 47.5 min. A strong correlation between diagnosis groups and the total duration of extracorporeal support can be observed. The average time delirium negative patients spent on CPB was 194.4 min, varying with a standard deviation of 38.5 min. A slightly weaker distinction can be made here.

We also examine both stints of CPB separately. Mean delirium negative duration of the CPB phase prior to circulatory arrest is 87.12 min. Measurements varied according to a standard deviation of 21.15 min. We observe that only patient 4 strongly deviates. Patient 5 may also be considered to deviate significantly. Despite these observations, the duration of the first CPB stint does not seem to be strongly correlated with postoperative delirium. The second stint of CPB seems to be much more correlated with postoperative delirium though. Second stint CPB duration of the delirium negative group is on average 135.01 min and varies with a standard

Table 4.2: Duration of Extracorporeal Support and Cardiopulmonary Bypass

Patient	Total Support (σ^{a})	CPB (σ^{b})	CPB Stint 1 (σ^{c})	CPB Stint 2 ($\sigma^{d\dagger}$)
2*	2.188	2.847	0.586	3.234
4*	3.205	4.458	4.815	4.069
5*	0.380	0.505	1.988	0.132
10*	2.066	1.645	1.317	1.639
13*	2.666	2.788	0.082	3.291
1	-0.995	-0.956	-0.108	-1.060
3	-0.886	-0.608	-0.411	-0.575
6	1.690	1.253	1.073	1.239
8	1.464	1.259	1.587	1.021
9	-0.248	-0.241	1.003	-0.498
11	-0.922	-0.968	-1.176	-0.807
12	0.660	0.589	0.212	0.674
14	-0.330	-1.323	-1.054	-1.254
17	-1.116	-1.461	-1.400	-1.331
18	-0.040	0.098	0.688	-0.021
19	0.722	1.036	-0.415	1.357

* Delirium positive
† Computed without patient 14
[a] $\mu = 218.68$ min, $\sigma = 47.50$ min
[b] $\mu = 194.41$ min, $\sigma = 38.46$ min
[c] $\mu = 87.12$ min, $\sigma = 21.15$ min
[d] $\mu = 135.01$ min, $\sigma = 32.74$ min

deviation of 32.74 min. We notice in Table 4.2 that patient 2, 4, 10, and 13 would be correctly classified as delirium positive if this metric were thresholded.

The duration of total extracorporeal support, overall CPB, and the second CPB

stint are correlated with postoperative delirium. The duration of the second stint of CPB seems to drive the correlation that overall CPB duration has with neurological integrity following the surgery. Considering the findings of previous studies, these variables are deemed potentially confounding. We note that the time needed to recover nominal EEG activity may affect the duration of the second CPB stint to some extent. The amount of time elapsed after burst suppression ends, however, could depend on numerous independent factors. Despite the presence of potential confounding factors, monitoring the BSDC at the very least provides a significantly earlier correlate upon which accurate predictions can be made about incidences of postoperative delirium—in fact, the same accuracy obtained by tracking the duration of CPB according to this study.

Chapter 5

DISCUSSION

We demonstrate in this work a capability to predict postoperative delirium based on the neurological activity observed through EEG recordings during cardiac surgeries involving DHCA and supplementary metabolic support techniques. This is accomplished by examining a specific neurophysiological phenomenon, known as burst suppression, within which the brain's metabolic state is embedded. Analyzing this phenomenon in further detail is important because the neuroprotective basis of DHCA is to control the brain's metabolic rate. Assuming no complications developed prior to circulatory arrest or when aorta repairs begin, the earliest opportunity to detect warnings of complications through electrophysiological measurements is immediately following circulatory arrest when CPB resumes. A diverse collection of electrophysiological behavior, especially in terms of burst suppression, was observed in this study. BSDC is suitable as a basis for comparing this neurological activity across several patients. Finding such metrics for intrapatient analysis is usually quite difficult. Furthermore, the metric is physiologically interpretable. Strategies were formulated to robustly process the data in a way that makes this neurophysiological characterization descriptive.

Accurate predictions are made by tracking through BSDC how fast patients recover nominal neurological function, signified by the conclusion of burst suppression activity. Prolonging the patient's recovery of nominal neurological function is correlated with postoperative delirium. More foresight has been achieved through this analysis. Predictions are most accurate once burst suppression has ceased. Naturally, prediction accuracy suffers slightly when observing earlier points in the transition.

Some delirium positive patients were observed to be deviant cases as soon as burst suppression began. These cases are patients 4, 10, and 13. Patient 10 is far less deviant at the beginning but eventually develops into a significant outlier. Patient 2 appears as delirium negative at the beginning of their burst suppression sequence, but also develops into a significant outlier over time. Consequently, BSDC must be closely monitored during the transition to observe if cases become outliers. This prediction capability potentially provides a better opportunity to act and attempt to avert complications during the surgery.

This work revises the methodology described in [40]. The feature vector used in [40] is not as robust as using T_γ alone with the updated learning and testing scheme. Motivation for using the two dimensional feature vector is detailed later in this chapter. Ultimately, slightly improved prediction performance was achieved—particularly when inferring from activity in the early stages of burst suppression. We produced a methodology that ensures results are sound by reducing the influence of classification bias. This is verified by the false classifier analysis. The affirmed results are further supported by appending related analyses. A lack of correlation between the perfusion technique used and the diagnosis signifies that the former is not a confounding factor in this study.

Discussed in Chapter 1 are three other studies particularly relevant to the findings of this work. First, a correlation between burst suppression duration and postoperative delirium for cases that do not employ DHCA was reported in [38]. A clear correlation was not determined in this study, but the measurements here do not disprove that one exists. Two cases, which resulted in postoperative delirium, were significant outliers in burst suppression duration though. Significance of burst suppression duration in a non-DHCA setting carries a different interpretation. Recall that burst suppression is observable not only in hypothermic conditions, but also in

normothermic conditions. This phenomenon can be observed while patients are anesthetized. BIS, which was formulated to track the depth of anesthesia, incorporates measurements characterizing burst suppression. In DHCA settings, we found burst suppression duration to be a piece of a more significant correlate. An even stronger correlation is observed when incorporating the time elapsed until burst suppression begins since resuming CPB.

Second, we identified in this study a similar, if not the same, finding reported in [37]. Stecker et al. identified a correlation between the time to recover continuous EEG activity since resuming full body perfusion after circulatory arrest. Observing when full BSDC has been met since resuming CPB is a similarly defined biomarker. In this study, however, we provide a framework for accurately predicting cases even before nominal EEG activity is observed once again. Additionally, the methods developed in this work assess the progress in transitioning from ECS to nominal activity quantitatively, as opposed to qualitatively. Qualitative evaluation of the BSDC in real time should not be ruled out though.

Lastly, this study improves on the first analysis we performed on this dataset documented in [39]. The detector constructed in [39] produced a different set of classification errors than those observed in the current work. Originally, patients 1 and 17 were classified as false positives, whereas all other patients were correctly classified. Instead, patient 5 and patient 2 (for part of the observation window) are classified as false negatives in this study. Patient 5 is consistently classified as a false negative based on monitoring BSDC, but correctly classified when observing EEG spatiotemporal covariance structure. Even though a patient may not be an outlier in terms of the BSDC progression, there may be other outlying characteristics of the electrophysiology. This may not be so surprising due to the various sources of possible complications. The significance of different forms of outlying electrophysiological

characteristics would need to be evaluated through further research incorporating various EEG based features. The results of both studies indicate that a richer description of the intraoperative neurophysiology can still be derived through EEG recordings and leveraged in prediction.

5.1 Limitations

A strong correlation was identified between postoperative delirium and the rate at which nominal neurophysiological function is recovered, but this study does not conclude a causal relationship between them. BSDC could possibly only contain a footprint for a causal mechanism of postoperative delirium. The current study design is not equipped to determine any causal relationships. Even then, this study is limited in concluding the correlation found. The first and foremost limitation is the small study population. We must determine if the documented methodology still works with a separate set of data. Additionally, many different features can still be extracted from this data. There is potentially an even more informative biomarker that can be found once a larger study is performed. Still, the results of this study motivate a further analysis determining if BSDC can capture the presence of an event or condition causing postoperative delirium. Naturally, an analysis of prolonging the recovery of full BSDC as a causal factor is prompted.

Many other intraoperative and preoperative factors can contribute to the patient's neurological state after the surgery. The operating notes report that several of the examined cases were already in delicate condition. Some of these procedures were prescribed to redo a previous repair. Whether the same sources of postoperative delirium arose in different cases of this specific dataset is uncertain. As a result, the second major limitation of this study is its retrospective nature. Many observations within the data prompt clarification as to what happened in the operating room, but

knowledge of those specific details is limited. The goal of the operating reports is not to provide every bit of detail regarding each decision made. There are several possible confounding factors, many of which are unknown to us for these cases. Fortunately, we were able to determine that the perfusion technique did not have a significant impact on the findings. Limited observations and knowledge of every surgical event prevents us from disentangling all other possible factors contributing to postoperative delirium for each case. A larger, more controlled, and more detailed study is required to identify these factors.

CPB duration is one potentially confounding factor identified with the currently available information. Previous studies identified that a longer CPB duration is correlated with incidences of postoperative neurological dysfunction. Therefore, the validity of the correlation between BSDC progress and postoperative delirium is called into question. In other words, there is a concern T_γ happens to be correlated with the diagnoses in this study but bears no actual significance. One potential strategy to remove this confounding variable in future studies is to ensure the CPB duration is kept within a duration known to be correlated with safer outcomes. Such control over the duration of CPB may not be feasible though. Analyzing cases with similar CPB durations would be equally helpful in clarifying the influence of CPB duration and milestone timestamps on postoperative delirium. This confusion is precisely why more data is need. Diversity in observations is needed to further validate the use of BSDC in predicting postoperative delirium.

Several pieces of information would have been useful for answering current questions in the context of this study. First is a more regularly and frequently sampling of body temperature. Having a denser, consistent sampling of body temperature would allow us to determine more precisely the effect warming has on BSDC. From this knowledge we might be able to develop a model that predicts BSDC given the

current temperature. Creating such a model may allow us to identify when BSDC progress does not follow the expected course and whether or not that indicates the occurrence of a complication. Second is the knowledge of when patients were switched between perfusion techniques, which would clarify the delays in warming. Perhaps the data and features from these patients must be treated in a different manner. Lastly, we are limited by a lack of knowledge about the long term outcome of these surgeries. Although postoperative delirium is correlated with poor long term prognosis [5], this study would benefit from a more direct measurement of patient recovery. More important than postoperative delirium, we must determine if long term outcome can be predicted. We expand on other desired pieces of information in Chapter 6.

5.2 Warming as a Driving Force

Change in body temperature in large part drives burst suppression dynamics [32]. From the EEG recordings and features, we observe that the delirium positive group varies almost uniquely between cases. Consequently, we want to better understand what relationship body temperature has with BSDC characteristics observed in the delirium positive cases. This discussion strictly relates the observed neurophysiology characteristics with warming. Many other factors can play a role in the electrophysiology. Whether or not signals of complications manifest in the BSDC in addition to the influence of warming remains to be determined. In other words, it is important to not derive a causal relationship between warming and postoperative delirium as a result of this discussion.

Two different ways of examining this relationship are presented. The first of which focuses on the change in esophageal temperature starting when the temperature is actively being affected. This is plotted in Figure 5.1. Delirium negative patient warming curves are plotted in gray. Delirium positive patient warming curves are

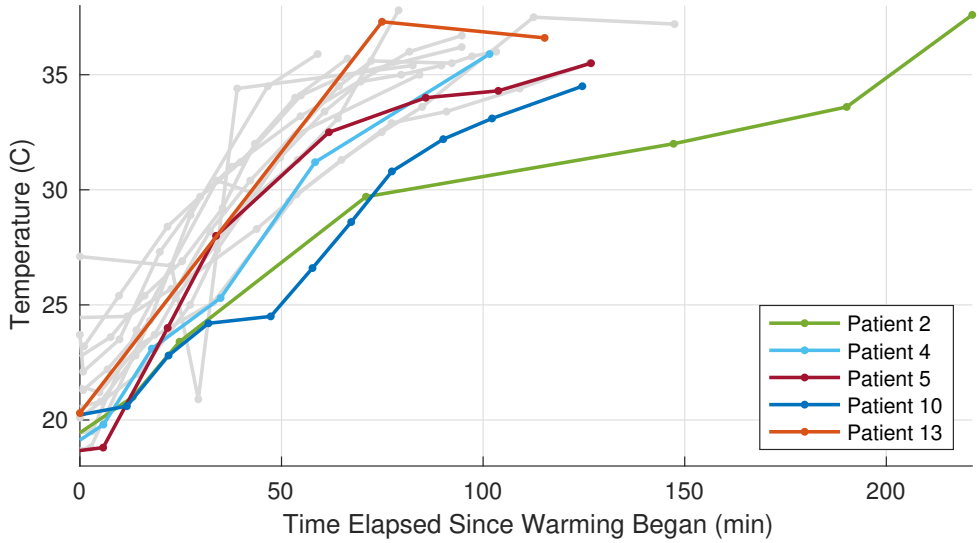

Figure 5.1: Esophageal temperature time series starting when rewarming begins. Temperature drives changes in burst suppression dynamics. Consequently, interpreting Figure 2.9, Figure 2.10, Figure 2.12, and 2.13 is clearer while referencing body temperature. The noticeable outliers in warming are patients 2 and 10.

overlaid and color coded to highlight any deviations from the other procedures. The second way is by observing the change in body temperature with respect to when CPB resumes. We compare these curves in terms of the diagnosis groups. Figure 5.2 contains all delirium negative patients. A somewhat more consistent progression can be observed in this group than the delirium positive patients shown in Figure 5.3. We relate these curves with Figure 2.12 and Figure 2.13.

Patient 2 deviated most from the delirium negative group in terms of burst suppression duration—the longest observed of all patients studied. Warming began shortly after CPB resumed as noted in Figure 5.3. Burst suppression began not long after according to Figure 2.12. Despite the timing of these events, T_1 had deviated significantly. In Figure 2.12, patient 2's BSDC advances regularly until the warming curve starts to level out prematurely. The BSDC quickly approaches and barely touches the ceiling but never settles into full BSDC. We believe this behavior is

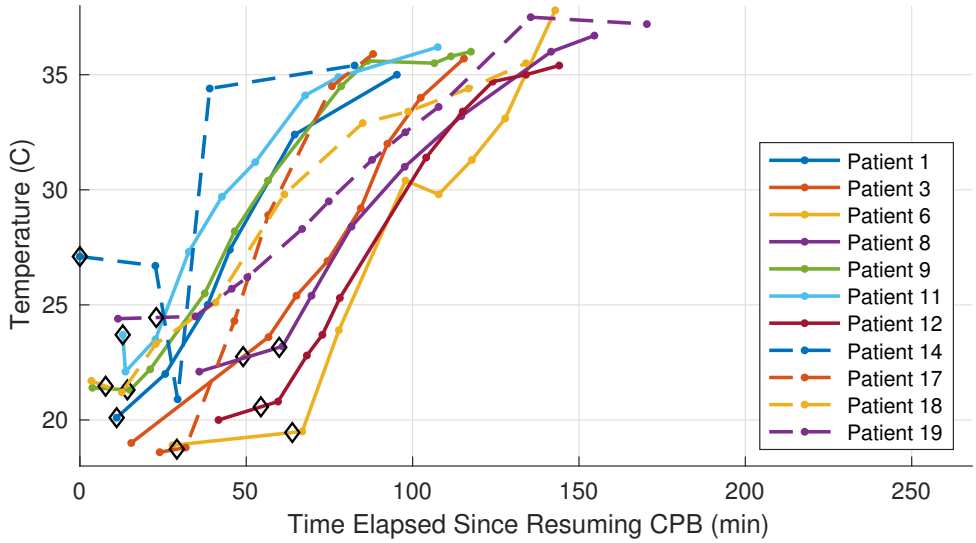

Figure 5.2: Change in esophageal temperature since resuming CPB for delirium negative patients. Patients 6, 8, and 12 underwent both SACP and RCP.

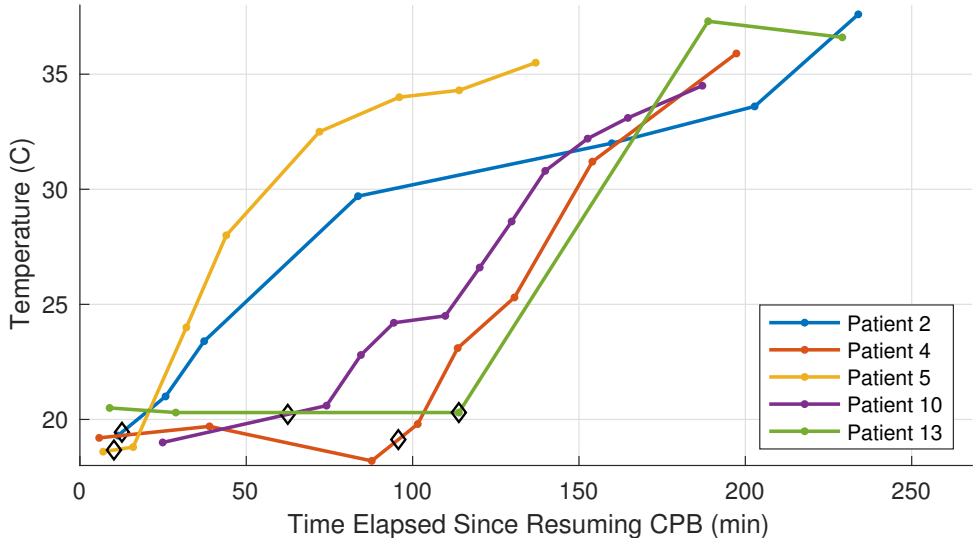

Figure 5.3: Change in esophageal temperature since resuming CPB for delirium positive patients. Patients 2, 4, 10, and 13 appear to be outliers when observing temperature change with respect to when CPB resumes. Each of these patients are outliers in different ways though. Patients 13 and 4 begin warming much later than the other patients.

dictated primarily by the deviant warming procedure observed in Figure 5.1. The rate of warming is much slower for this patient than other cases. At 83.6 min after resuming CPB, when esophageal temperature reads 29.7 C, the warming rate slows down even further. At this point the BSDC is 87.3%. Full BSDC was reached at minute 194.3. Recall that patient 2 was officially classified as delirium positive when much later BSDC milestones were met. This is due to warming slowing down in the later stages of BSDC progression. Why warming proceeded in the manner documented in Figure 5.1 is not documented. This deviant warming procedure is particularly curious because their cooling phase was normal with respect to the delirium negative group. Complications may have dictated such an adjustment in warming.

Patient 4's warming rate observed in Figure 5.1 does not appear irregular in terms of the delirium negative trend. According to Figure 5.3, warming began extremely late though. This patient underwent both RCP and SACP. This delay could be attributed to the additional use of SACP, if in such cases repairs to the aorta continued when resuming CPB. Still, warming began late even in terms of other cases needing both techniques. Once burst suppression activity has started, the rate at which BSDC approaches the ceiling is within a normal amount of time when examining Figure 2.12. Despite this, full BSDC is reached 172.6 min after resuming CPB, which is significantly longer than the delirium negative mean. This case may already have been in a delicate state though, noted by annotations regarding a pacemaker, which may have influenced the sequence and timing of events.

Patient 5 appears to be an outlier of the delirium positive group in terms of BSDC characteristics following DHCA. In fact 100% BSDC was reached well before normothermic temperatures. This behavior is reflected by the documented warming procedure. We observe that the rate at which warming progresses does not appear irregular in Figure 5.1. The time at which warming began is also normal. Conse-

quently, we suspect other complications may have occurred that are not observable in the BSDC. This case highlights complexity in determining effective ways of identifying compromised neurophysiology in EEG signals.

Patient 10's warming is slightly irregular, slowing down considerably toward the beginning of the transition, but resuming a more normal rate shortly after. This patient underwent both RCP and SACP, a potential reason for the delay in start of warming. This delay, however, does not appear to be uncharacteristic of delirium negative cases involving both forms of cerebral perfusion. Aside from this delay, we observe the most irregular BSDC progression of the entire dataset in this case. Patient 10 never reached full BSDC before CPB ends. We are unable to precisely determine what influence the warming has over this characteristic, aside from noting the general upward trend that results from increasing body temperature. We suspect a factor other than warming is significantly affecting the BSDC, which implies the metabolic state of the brain may be affected. We hypothesize that factor has some relation to a causal mechanism of postoperative delirium.

Patient 13's warming rate is presumed to be regular considering the available timestamps, but the precise shape of the curve is unknown because the temperature was measured sparsely. As can be observed in Figure 2.13, their burst suppression duration was shorter than most patients studied. Warming began considerably late though. The reason for this abnormal delay is not documented. The delay cannot be attributed to the possible use of SACP because only RCP was employed. Repairs may still have been occurring during this delay, but this cannot be corroborated at this time. The length of the delay was certainly the primary contributor to the outlying correlate, T_γ, but remember that it is possible some other factor contributed to this incidence of postoperative delirium instead.

The correlation BSDC has with postoperative delirium prompts a more focused

study on the relationship between warming and BSDC. Obtaining a better understanding of how patients respond to warming could help us identify abnormalities in the BSDC cued by other factors. We are curious if expediting warming to shorten T_1 would be beneficial. Even if a longer T_1 is not a direct causal mechanism, but instead increases the chances of ischemic and embolic events, we would like to determine if the probability of such events could be reduced by speeding up this transition. At the very least, BSDC could be used to monitor the warming procedure in a similar way to how EEGs are used to guide the cooling process. This might be particularly useful because each patient can respond differently to warming, especially considering the possible side effects of the repairs. If need be, the warming procedure could be tailored to the patient's needs by monitoring BSDC.

5.3 Real Time Functionality

The formulated procedure can function without modification in real time, granted the prior knowledge necessary to generate the likelihood ratio tests. All the required parameters to estimate BSDC are computed before DHCA. Monitoring the progress simply involves recording the time elapsed as each BSDC milestone is met, as in Figure 3.1. At each milestone, likelihoods the current observation indicates either diagnosis are generated. Based on those likelihoods, the prognosis of the patient is predicted, and the projected outcome is updated over time. This assessment should be corroborated through other forms of physiological monitoring if available. Then an intervention, if one were applicable, could take place.

Consider patient 2 for example. Despite not becoming clearly classified as delirium positive until they cross the 90% BSDC milestone, we observe in Figure 3.1 that their progress takes a noticeable turn. They begin to drastically deviate starting at 75% BSDC. In fact, this observation is what [40] leverages when attempting to obtain

further foresight. This is accomplished by separating T_γ into T_0 and the progress in relationship to when warming begins $T_\gamma - T_{\mathrm{W}}$, where T_{W} is the warming timestamp offset by when CPB restatrts. The set of these two features better captured this form of deviation, but as stated earlier is not robust. In a real time situation, a warning may be raised at some point by observing T_γ over time before the likelihood ratio actually classifies the case as delirium positive.

Otherwise, the dual question could be asked. In this study, we determine if the patient is arriving at certain milestones on time. The dual approach involves checking if patients have met a minimum BSDC at certain times. A new set of statistics on the average BSDC level and variance would be generated at each monitoring interval. Then, new likelihood tests can be generated. In other words, the BSDC level is thresholded instead of the time elapsed.

Other studies have shown promise in intervening based on cues provided by EEGs. This study has not identified specific cues for intervention in the BSDC. Considering BSDC as a peripheral measure of the brain's metabolic demand, however, we make a conjecture about avenues of possible intervention. A natural choice for intervention is to adjust metabolic support through either oxygen or blood delivery. This may mean adjusting perfusion parameters, which could be strongly restricted. We are curious if adjusting warming to control the brain's metabolic rate in a more suitable manner is possible, but there are also constraints on warming to which a surgical team must adhere [15]. Additional research is required to better understand what the BSDC conveys and how warming affects it before any actual detailed directions are suggested. At the very least, the results by [14] instill optimism by demonstrating the existence of effective interventions for correcting course. The optimism motivates continued research in improving our ability to detect warning signs that postoperative delirium may result.

Chapter 6

CONCLUSION

This study aims to predict postoperative delirium through intraoperative EEG during cardiac surgeries involving DHCA. We specifically monitor burst suppression activity, which is the earliest opportunity to search for warning signs of neural injury through EEG after DHCA. Postoperative delirium can be accurately predicted not only by the time burst suppression has ended, but also while the brain transitions through this behavior back to nominal activity. Consequently, the devised methodology provides further foresight when anticipating postoperative delirium, allowing for a better opportunity to intervene so that neural injury can be minimized. A correlation is observed between the evolution of burst suppression dynamics and postoperative delirium. Some delirium positive cases begin and continue as outliers in terms of this evolution, whereas others can develop into them. This result motivates further research in determining how causal factors and signs of neurological complications can be observed through BSDC. We are also motivated to determine if the timing of the burst suppression progression is a causal mechanism in future studies. Hopefully this research is informative for future efforts made toward reducing the risk of such surgeries.

6.1 Further Analysis of Currently Available Data

Just one of many possible characteristics of the neurophysiology observed through EEGs is analyzed in this study. Between the results of the current work and our previous studies, we recognize the potential for achieving even better accuracy when anticipating DHCA. Analysis on classification strategies based on multidimensional

features must continued to be performed. This was attempted in [40], but a more robust feature set needs to be extracted. Perhaps a similar decomposition of T_γ would reveal an ability to raise warnings sooner. Some other candidates to include in a reformulated feature vector consist of multiple BSDC timing measurements, spatiotemporal covariance, and energy. Several possible avenues of further characterization methods can still be pursued with the available dataset though.

One direction to pursue is analyzing the timing of burst and suppression phases. We seek a better understanding about how durations of either phase depend on attributes of the patient, body temperature, and incidences of neural injury. A deeper analysis on the the temporal structure of the bursts, such as the speed of the rising transient and the decay rate, should be conducted. In other words, we need to determine if signs of neural injury are also encoded in each individual burst. The current study identifies if signs are encoded in ensembles of bursts instead.

Further analysis regarding the spatial variation in EEG activity, especially burst suppression, should be performed. On a "macroscopic" level, burst suppression appears global and temporally closely synchronized. Perhaps, there are micro-shifts of which we are not cognizant. If there are such micro-shifts, then we must determine if they express the presence of a complication. Aside from the apparent temporal synchronicity observed, the coherence of bursts across channels should be analyzed. Naturally, whether or not abnormalities related to neural injury can be identified within this characteristic should be determined.

The current study only considers EEG channels measured along the zenith axis of the head. Only five of the 23 channels are used here, so plenty of information has been ignored. Every channel is used in [39], but the individual contributions from each of them is not well understood. All that is understood is that the deviation from baseline in covariance structure during warming is different between diagnosis

groups. In addition to incorporating more channels, we can also analyze differences in activity across hemispheres, which may provide cues of postoperative delirium. Soehle et al. examine asymmetries in activity across hemispheres [38]. Asymmetries should be characterized in terms of not only burst suppression, but also other features, such as energy.

6.2 Future Prospective Studies

The ultimate goal of this study is to improve the expected long term prognosis of cardiac surgeries involving DHCA by ensuring the brain is properly protected. We believe that intraoperatively monitoring brain electrophysiology and identifying situations of heightened risk will provide a pathway to reduce incidences of neurological injury. Although great strides were made, the current study is limited in understanding how to accomplish this goal. At this stage, a new study is necessary to answer further questions. We identify some pieces of information that would be useful to accomplishing that goal.

Many confounding factors prohibit a proper understanding of how monitoring brain electrophysiology can anticipate injury. Two of these variables are surgery complexity and preoperative health of patients. Obtaining a larger sample size would help minimize the effects of confounding variables. A larger data set increases the diversity of observations and the chances of observing procedures containing similar types of deviations. With this increased knowledge base, we can better interpret the electrophysiology. Additionally, more meticulous documentation of the procedure, including reasons for certain decisions, would create a better understanding of the observable electrophysiological phenomenon. In particular, documenting when and if patients are switched between RCP and SACP would be useful. One confounding variable analyzed in this study is the duration of extracorporeal support and CPB.

We had mentioned in Chapter 5 potential methods for reducing the influence of this factor. Obtaining data on more cases would allow us to effectively parameterize the analysis to account for the influences of various sources of complications. Care must be taken when designing a future study to minimize the number of confounding variables ahead of time if at all feasible.

Due to the correlation found between BSDC and postoperative delirium, we suggest conducting a more focused study on how warming affects BSDC. Body temperature should be sampled at a higher and more regular rate than what was observed in this study. Sampling at 500 Hz is not necessary, but a rate of 1 Hz would likely be sufficient. Furthermore, an additional temperature measurement site, such as the nasopharynx, would be useful. Having this data would allow us to determine precisely what contribution warming has on the BSDC and how much of the current reading could be explain by other sources. The utility of guiding the warming procedure based on BSDC—in a similar way that EEGs are already used to guide the cooling process—should also be assessed. This also prompts determining how parameters controling warming and cooling affect risk.

Intraoperative measurements aside from EEG would be useful in tracking patient status especially because EEGs are blind to any complications that may happen during circulatory arrest or ECS. This lack of knowledge is one of the major limitations of this current study. There is no opportunity for intervention if monitoring is not possible. Consequently, hemodynamics should be monitored throughout the surgery. This mode of measurement may be the most descriptive of the neurophysiology during circulatory arrest. The amount of metabolic support can be monitored this way. Furthermore, microembolic events can be detected. Hemodynamics should be monitored concurrently with EEG signals. Efficacy in anticipating complications using multimodal monitoring and intervening based on those cues was demonstrated by [14].

Determining a relationship between the observations of either modes of monitoring would also be useful.

Lastly, long term outcomes, such as patient mortality and extent of any neurological dysfunction, should be documented. Long term neurological health can be determined in part by a cognitive battery. The type of neurological dysfunction such as, memory or motor control issues, should be stated. Other forms of complications, including organ failure and sepsis, should be reported as well. We also propose that the patient's recovery progress should be recorded at several intervals past surgery. An example set of observation intervals may consist of the end of the first month, first six months, and first year. Future studies should seek to identify a connection between intraoperative data and long term outcome.

REFERENCES

[1] Centers for Disease Control and Prevention, National Center for Health Statistics, "Underlying cause of death 1999-2018." `http://wonder.cdc.gov/ucd-icd10.html`, 2020. Retrieved from CDC WONDER Online Database.

[2] Mayo Clinic, "Thoracic aortic aneurysm." `https://www.mayoclinic.org/diseases-conditions/thoracic-aortic-aneurysm/symptoms-causes/syc-20350188`, 2018.

[3] A. Gega, J. A. Rizzo, M. H. Johnson, M. Tranquilli, E. A. Farkas, and J. A. Elefteriades, "Straight deep hypothermic arrest: Experience in 394 patients supports its effectiveness as a sole means of brain preservation," *The Annals of Thoracic Surgery*, vol. 84, no. 3, pp. 759–767, 2007.

[4] M. A. Ergin, S. Uysal, D. L. Reich, A. Apaydin, S. L. Lansman, J. N. McCullough, and R. B. Griepp, "Temporary neurological dysfunction after deep hypothermic circulatory arrest: A clinical marker of long-term function deficit," *The Annals of Thoracic Surgery*, vol. 67, no. 6, pp. 1887–1890, 1999.

[5] C. H. Brown IV, "Delirium in the cardiac surgical intensive care unit," *Current Opinion in Anaesthesiology*, vol. 27, no. 2, pp. 117–122, 2014.

[6] D. L. Reich, S. Uysal, M. Sliwinksi, M. A. Ergin, R. A. Kahn, S. N. Konstadt, J. McCullough, M. R. Hibbard, W. A. Gordon, and R. B. Griepp, "Neuropsychologic outcome after deep hypothermic circulatory arrest in adults," *The Journal of Thoracic and Cardiovascular Surgery*, vol. 117, no. 1, pp. 156–163, 1999.

[7] B. A. Ziganshin and J. A. Elefteriades, "Deep hypothermic circulatory arrest," *Annals of Cardiovascular Surgery*, vol. 2, no. 3, pp. 303–315, 2013.

[8] J. N. McCullough, N. Zhang, D. L. Reich, T. S. Juvonen, J. J. Klein, D. Spielvogel, M. A. Ergin, and R. B. Griepp, "Cerebral metabolic suppression during hypothermic circulatory arrest in humans," *Annals of Thoracic Surgery*, vol. 67, no. 6, pp. 1895–1899, 1999.

[9] C. Hagl, M. A. Ergin, J. D. Galla, S. Lansman, D. McCullough, J. Spielvogel, P. Sfeir, C. A. Bodian, and R. B. Griepp, "Neurologic outcome after ascending aortaaortic arch operations: Effect of brain protection technique in high-risk patients," *The Journal of Thoracic and Cardiovascular Surgery*, vol. 121, no. 6, pp. 1107–1121, 2001.

[10] E. S. Krähenbühl, F. F. Immer, M. Stalder, L. Englberger, F. S. Eckstein, and T. P. Carrel, "Temporary neurological dysfunction after surgery of the thoracic aorta: A predictor of poor outcome and impaired quality of life," *European Journal of Cardio-thoracic Surgery*, vol. 33, no. 6, pp. 1025–1029, 2008.

[11] Y. Okita, K. Minatoya, O. Tagusari, M. Ando, K. Nagatsuka, and S. Kitamura, "Prospective comparative study of brain protection in total aortic arch replacement: Deep hypothermic circulatory arrest with retrograde cerebral perfusion or selective antegrade cerebral perfusion," *The Annals of Thoracic Surgery*, vol. 72, no. 1, pp. 72–79, 2001.

[12] E. Kumral, M. Yüksel, S. Büket, T. Yagdi, Y. Atay, and A. Güzelant, "Neurologic complications after deep hypothermix circulatory arrest," *Texas Heart Institute Journal*, vol. 28, no. 2, pp. 83–88, 2001.

[13] J. G. Augoustides, T. F. Floyd, M. L. McGarvey, E. A. Ochroch, A. Pochettino, S. Fulford, A. J. Gambone, J. Weiner, S. Raman, J. S. Savino, J. E. Bavaria, and D. R. Jobes, "Major clinical outcomes in adults undergoing thoracic aortic surgery requiring deep hypothermic circulatory arrest: Quantification of organ-based perioperative outcome and detection of opportunities for perioperative intervention," *Journal of Cardiothoracic and Vascular Anesthesia*, vol. 19, no. 4, pp. 446–452, 2005.

[14] P. Zanatta, S. M. Benvenuti, E. Bosco, F. Baldanzi, D. Palomba, and C. Valfrè, "Multimodal brain monitoring reduces major neurologic complications in cardiac surgery," *Journal of Cardiothoracic and Vascular Anesthesia*, vol. 25, no. 6, pp. 1076–1085, 2011.

[15] S. C. Conolly, J. E. Arrowsmith, and A. A. Klein, "Deep hypothermic circulatory arrest," *Continuing Education in Anaesthesia, Critical Care & Pain*, vol. 10, no. 5, pp. 14–21, 2010.

[16] B. Braathen and T. Tønnessen, "Cold blood cardioplegia reduces the increase in cardiac enzyme levels compared with cold crystalloid cardioplegia in patients undergoing aortic valve replacement for isolated aortic stenosis," *The Journal of Thoracic and Cardiovascular Surgery*, vol. 139, no. 4, pp. 874–880, 2010.

[17] T. D. Yan, P. G. Bannon, J. Bavaria, J. S. Coselli, J. A. Elefteriades, R. B. Griepp, G. C. Hughes, S. A. LeMaire, T. Kazui, N. T. Kouchoukos, M. Misfeld, F. W. Mohr, A. Oo, L. G. Svensson, and D. H. Tian, "Consensus on hypothermia in aortic arch surgery," *Annals of Cardiothoracic Surgery*, vol. 2, no. 2, 2013.

[18] M. L. James, N. D. Andersen, M. Swaminathan, B. Phillips-Bute, J. M. Hanna, G. R. Smigla, M. E. Barfield, S. D. Bhattacharya, J. B. Williams, J. G. Gaca, A. M. Husain, and G. C. Hughes, "Predictors of electrocerebral inactivity with deep hypothermia," *The Journal of Thoracic and Cardiovascular Surgery*, vol. 147, no. 3, pp. 1002 – 1007, 2014.

[19] D. L. Reich, S. Uysal, M. A. Ergin, and R. B. Griepp, "Retrograde cerebral perfusion as a method of neuroprotection during thoracic aortic surgery," *The Annals of Thoracic Surgery*, vol. 72, no. 5, pp. 1774–1782, 2001.

[20] Y. Ueda, "A reappraisal of retrograde cerebral perfusion," *Annals of Cardiothoracic Surgery*, vol. 2, no. 3, 2013.

[21] J. Bachet, D. Guilmet, B. Goudot, G. D. Dreyfus, P. Delentdecker, D. Brodaty, and C. Dubois, "Antegrade cerebral perfusion with cold blood: a 13-year experience," *The Annals of Thoracic Surgery*, vol. 67, no. 6, pp. 1874–1878, 1999.

[22] D. Harrington, A. Walker, H. Kaukuntla, R. Bracewell, T. Clutton-Brock, M. Faroqui, D. Pagano, and R. Bonser, "Selective antegrade cerebral perfusion attenuates brain metabolic deficit in aortic arch surgery," *Circulation*, vol. 110, no. 11_suppl_1, pp. II–231–II–236, 2004.

[23] D. Spielvogel and G. H. L. Tang, "Selective cerebral perfusion for cerebral protection: what we do know," *Annals of Cardiothoracic Surgery*, vol. 2, no. 3, 2013.

[24] D. R. Busch, C. G. Rusin, W. Miller-Hance, K. Kibler, W. B. Baker, J. S. Heinle, C. D. Fraser, A. G. Yodh, D. J. Licht, and K. M. Brady, "Continuous cerebral hemodynamic measurement during deep hypothermic circulatory arrest," *Biomedical Optics Express*, vol. 7, no. 9, pp. 3461–3470, 2016.

[25] G. W. Fischer, H.-M. Lin, M. Krol, M. F. Galati, G. D. Luozzo, R. B. Griepp, and D. L. Reich, "Noninvasive cerebral oxygenation may predict outcome in patients undergoing aortic arch surgery," *The Journal of Thoracic and Cardiovascular Surgery*, vol. 141, no. 3, pp. 815–821, 2011.

[26] W. C. Pearcy and R. W. Virtue, "The electroencephalogram in hypothermia with circulatory arrest," *Anesthesiology*, vol. 20, no. 3, pp. 341–347, 1959.

[27] W. J. Levy, "Quantitative analysis of eeg changes during hypothermia," *Anesthesiology*, vol. 60, no. 4, pp. 291–297, 1984.

[28] W. J. Levy, E. Pantin, S. Mehta, and M. McGarvey, "Hypothermia and the approximate entropy of the electroencephalogram," *Anesthesiology*, vol. 98, no. 1, pp. 53–57, 2003.

[29] J. Bruhn, H. Röpcke, and A. Hoeft, "Approximate entropy as an electroencephalographic measure of anesthetic drug effect during desflurane anesthesia," *Anesthesiology*, vol. 92, no. 3, pp. 715–726, 2000.

[30] M. M. Stecker, A. T. Cheung, A. Pochettino, G. P. Kent, T. Patterson, S. J. Weiss, and J. E. Bavaria, "Deep hypothermic circulatory arrest: I. effects of cooling on electroencephalogram and evoked potentials," *The Annals of Thoracic Surgery*, vol. 71, no. 1, pp. 14–21, 2001.

[31] F. Amzica, "Basic physiology of burst-suppression," *Epilepsia*, vol. 50, no. 12, pp. 38–39, 2009.

[32] M. B. Westover, S. Ching, V. M. Kumaraswamy, O. Akeju, E. Pierce, S. S. Cash, R. Kilbride, E. N. Brown, and P. L. Purdon, "The human burst suppresion electroencephalogram of deep hypothermia," *Clinical Neurophysiology*, vol. 126, no. 10, pp. 1901–1914, 2015.

[33] I. J. Rampil, "A primer for eeg signal processing in anesthesia," *Anesthesiology*, vol. 89, no. 4, pp. 980–1002, 1998.

[34] M. Hayashida, H. Sekiyama, R. Orii, M. Chinzei, M. Ogawa, H. Arita, K. Hanaoka, and S. Takamoto, "Effects of deep hypothermic circulatory arrest with retrograde cerebral perfusion on electroencephalographic bispectral index and suppression ratio," *Journal of Cardiothoracic and Vascular Anesthesia*, vol. 21, no. 1, pp. 61–67, 2007.

[35] S. Ching, P. L. Purdon, S. Vijayan, N. J. Kopell, and E. N. Brown, "A neurophysiological-metabolic model for burst suppression," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 8, pp. 3095–3100, 2012.

[36] G. Santarpino, R. Fasol, J. Sirch, B. Ackermann, S. Pfeiffer, and T. Fischlein, "Impact of bispectral index monitoring on postoperative delirium in patients undergoing aortic surgery," *HSR proceedings in intensive care & cardiovascular anesthesia*, vol. 3, pp. 47–58, 01 2011.

[37] M. M. Stecker, A. T. Cheung, A. Pochettino, G. P. Kent, T. Patterson, S. J. Weiss, and J. E. Bavaria, "Deep hypothermic circulatory arrest: Ii. changes in electroencephalogram and evoked potentials during rewarming," *The Annals of Thoracic Surgery*, vol. 71, no. 1, pp. 22–28, 2001.

[38] M. Soehle, A. Dittmann, R. K. Ellerkmann, G. Baumgarten, C. Putensen, and U. Guenther, "Intraoperative burst suppression is associated with postoperative delirium following cardiac surgery: a prospective, observational study," *BMC Anesthesiology*, vol. 15, pp. 1–8, 2015.

[39] O. Ma, A. Dutta, D. W. Bliss, and A. Z. Crepeau, "Predicting postoperative delirium in patients undergoing deep hypothermia circulatory arrest," in *2017 51st Asilomar Conference on Signals, Systems, and Computers*, pp. 1313–1317, Oct 2017.

[40] O. Ma, A. Z. Crepeau, A. Dutta, and D. W. Bliss, "Anticipating postoperative delirium during burst suppression using electroencephalography," *IEEE Transactions on Biomedical Engineering*, 2020.

[41] R. Oostenveld and P. Praamstra, "The five percent electrode system for high-resolution eeg and erp measurements," *Clinical Neurophysiology*, vol. 112, no. 4, pp. 713–719, 2001.

[42] E. W. Ely, "Confusion assessment method for the icu." `http://www.icudelirium.org/docs/CAM_ICU_training.pdf`, 2014.

[43] S. Kay, *Fundamentals of Statistical Signal Processing, Volume II: Detection Theory*. Englewood Cliffs: Prentice Hall, 1998.

[44] B. Efron and R. Tibshirani, "Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy," *Statistical Science*, vol. 1, no. 1, pp. 54–75, 1986.

[45] F. Provost and T. Fawcett, "Robust classification in imprecise environments," *Machine Learning*, vol. 42, no. 3, pp. 203–231, 2001.

APPENDIX A

PREDICTION ON GRADED OUTCOMES

Included in our data set is the time elapsed following the operation before the patient was extubated. This metric is an alternative measure on a patient's short term postoperative recovery—earlier than postoperative delirium. There is an opportunity for a graded evaluation of the result here, where there is now a spectrum for a patient's prognosis. That is, we would like to determine how the intraoperative observations map to a variation of outcomes. In a new prospective study, determining how this measurement on postoperative recovery relates to the long term outcome would be useful, where complications can vary widely in severity.

Table A.1 documents the time required until the patients were able to be extubated. Patients 4, 5, and 13, who tested positive for postoperative delirium, were extubated two days after their procedures. Patients 2 and 10, the other members of the delirium positive group were intubated for one day beyond surgery. Several delirium negative patients, however, were intubated for an additional day. All patients extubated day of the surgery were delirium negative. Consequently, there appears to be a slight correlation. To properly determine a relationship between BSDC and time to extubation, we believe the resolution of the latter variable would at least need to be on the order of hours. A finer resolution would permit the use of regression analysis. This is just one alternative postoperative variable amongst many to measure. The correlation observed here prompts us to be understand how the intraoperative electrophysiological characteristics relate to these other measures of postoperative recovery.

Table A.1: Time to Extubation

Patient	Post Operative Day
1	0
2*	1
3	1
4*	2
5*	2
6	1
8	1
9	0
10*	1
11	0
12	0
13*	2
14	1
17	0
18	1
19	0

* Delirium positive