Life After Death:
Techniques for the Prognostication of Coma Outcomes after Cardiac Arrest

by

Mohammad Mahdi Ghassemi

B.S. Electrical Engineering, New Mexico State University, 2008
MPhil. Information Engineering, University of Cambridge, 2011

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
IN ELECTRICAL ENGINEERING AND COMPUTER SCIENCE
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2018

Signature of Author _____

Department of Electrical Engineering and Computer Science
May 23, 2018

Certified by _____

Roger G. Mark
Professor of Electrical Engineering and Computer Science
Thesis Co-Supervisor

Certified by _____

Emery N. Brown
Professor of Brain and Cognitive Science
Thesis Co-Supervisor

Accepted by _____

Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students

LIFE AFTER DEATH:

Techniques for the Prognostication of Coma Outcomes after Cardiac Arrest

by

Mohammad Mahdi Ghassemi

Submitted to the Department of Electrical Engineering and Computer Science
on May 23, 2018 in partial fulfillment of the requirements for the Degree of
Doctor of Philosophy in Electrical Engineering and Computer Science

ABSTRACT

Electroencephalography (EEG) features are known to predict neurological outcomes of patients in coma after cardiac arrest, but the association between EEG features and outcomes is time-dependent. Recent advances in machine learning allow temporally-dependent features to be learned from the EEG waveforms in a fully-automated way, allowing for faster, better-calibrated and more reliable prognostic predictions. In this thesis, we discuss three major contributions to the problem of coma prognostication after cardiac arrest: (1) the collection of the world's largest multi-center EEG database for patients in coma after cardiac arrest, (2) the development of time-dependent, interpretable, feature-based EEG models that may be used for both risk-scoring and decision support at the bedside, and (3) a careful comparison of the performance and utility of feature-based techniques to that of representation learning models that fully-automate the extraction of time-dependent features for outcome prognostication.

Thesis Co-Supervisor: Roger G. Mark
Professor of Electrical Engineering and Computer Science

Thesis Co-Supervisor: Emery N. Brown
Professor of Brain and Cognitive Science

Contents

List of Figures

List of Tables

List of Recurring Acronyms

SCA: Sudden Cardiac Arrest

ICU: Intensive Care Unit

PAC: Post-anoxic Coma

CPR: Cardiopulmonary Resuscitation

EDF: European Data Format

VF: Ventricular Fibrillation

CPC: Cerebral Performance Category

DNN: Deep Neural Networks

LOOCV: Leave-one-out Cross Validation

AUROC: Area Under The Receiver Operator Curve

ROSC: Return of spontaneous circulation

EEG: Electroencephalography

qEEG: Quantitative Electroencephalography

PEA: Pulseless Electrical Activity

OHCA: Out of Hospital Cardiac Arrest

IHCA: In hospital Cardiac Arrest

SD: Standard Deviation

CRI: Cerebral Recovery Index

FNR: False Negative Rate

FPR: False Positive Rate

TPR: True Positive Rate

TNR: True Negative Rate

IQ: Information Quantit

ESE: Electrographic Status Epilepticus

BS: Burst Supression

DBD: Dynamic Bayesian Network

KF: Kalman Filter

IRB: Institutional Review Board

HIE: Hypoxia Induced Encephalopathy

Executive Summary

The goal of this thesis was to develop techniques for the rapid prognostication of post-anoxic coma (PAC) patients. Specific electroencephalogram (EEG) patterns are associated with eventual recovery from coma after cardiac arrest. However, existing EEG review practices rely on subjective visual analysis, which does not readily translate into reproducible quantitative predictions of neurological outcomes. PAC prognostication requires analytically grounded methods that assign risks of 'good' and 'bad' neurological outcomes to reduce clinical subjectivity, and avoid poor outcomes as a result of self-fulfilling prophecies. Quantitative EEG (qEEG) methods are capable of detecting early signs of neurological recovery while coma persists. Existing qEEG methodologies are limited by small sample sizes (< 100), limited qEEG features (ten or less), and tend not to account for changes in the prognostic significance of EEG features over the course of patient recovery. We addressed each of these challenges in the thesis:

- **We collected the world's largest dataset of PAC patients** from five university-affiliated hospitals between October 2009 and April 2016. The data contained over 35,000 hours of 21-channel continuous EEG recordings, a selection of clinical covariates, and an ordinal measure of neurological outcomes (Glasgow-Pittsburgh Cerebral Performance categories, CPC) for 950 adult patients diagnosed with in-hospital and out-of-hospital cardiac arrests.
- **We extracted 57 quantitative EEG features** that capture three signal properties:
  - complexity: the degree of randomness in the EEG signal,
  - category: qualitative descriptors of signal characteristics or behaviors and
  - connectivity: interactions between EEG electrodes.
- **We tested novel methods for time-sensitive classification** of patient outcomes:
  - penalized, sequential, logistic regression using 57 multi-scale features,
  - logistic regression using 10 qEEG features, with temporal dynamics constrained by a dynamic Bayesian network and
  - a variety of deep neural network architectures: convolutional, recurrent, and feed-forward.

We found that prognostication of bad outcomes (CPC > 2) may be performed robustly by spatio-temporal deep learning methods and feature-based approaches that account for time. The performance of our methods on unseen data was 0.8 AUROC, with a true positive rate of 25% at false positive rate of 0%. Our results provide strong motivation for the deployment of quantitative approaches to coma prognostication and highlight the immense potential of machine learning for health care problems, more generally.

"It is safer to stand next to that transformer, than it is to stand on the sidewalk, next to traffic. That's just statistics."

–A Role Model

Chapter 1

Introduction

Synopsis

Over 300,000 cardiac arrests occur every year in the United States. Approximately half of patients surviving to hospital admission sustain severe anoxic brain injury, resulting in coma. Despite consensus recommendations to postpone prognostication at least 72 hours post-arrest, the most common proximal cause of mortality among cardiac arrest patients is early withdrawal of life-sustaining therapies. There is a need for accurate methods to assign risk of 'good' and 'bad' neurological outcomes early after cardiac arrest, to reduce subjectivity and avoid poor outcomes as a result of self-fulfilling prophecies. Specific electroencephalogram (EEG) patterns are associated with eventual recovery from coma due to hypoxic-ischemic encephalopathy (HIE) after cardiac arrest. However, existing EEG review practices rely on subjective visual analysis, which does not readily translate into reproducible quantitative predictions of neurological outcome. Quantitative EEG (qEEG) methods are capable of detecting early signs of neurological recovery while coma persists. Hence, qEEG holds promise for reliably identifying patients likely to benefit from continued intensive supportive care. However, existing qEEG methodologies are limited by small sample sizes (< 100), limited qEEG features (ten or less), and tend not to account for changes in the prognostic significance of EEG features over the course of patient recovery. Our goal in this thesis is to overcome these limitations.

1.1 Setting the Scene

While jogging early on the beach one morning a runner begins to experience chest pains resulting from an abnormal cardiac rhythm, after which he collapses and enters cardiac arrest. Without spontaneous circulation, cells in the runner's brain begin to die due to a lack of oxygen. After some time, the runner is found by a passer-by who performs cardiopulmonary resuscitation (CPR), and calls the emergency medical services (EMS). Upon arrival, the EMS team defibrillates the runner, returning his cardiac rhythm, but the runner does not regain consciousness; he has entered an anoxia induced coma. EMS delivers the runner to an Intensive Care Unit (ICU), where he is promptly cooled to 34 degrees Celsius for approximately 24 hours. He is then slowly rewarmed to normothermia over a twelve to eighteen hour period. During his stay in the ICU, he may receive anesthetics and anti-epileptic drugs to suppress his brain's metabolism, and prevent seizures. While the patient lies unconscious, clinicians and family members are eager to know if the runner will wake up, and if does wake up, will he be able to resume normal activities? To answer this question, clinicians may perform a neurological examination of the patient, applying several physical stimuli, and checking for patient reactions that are reflective of brain damage. Additionally, the clinical team may consider laboratory measures or EEG recordings to help prognosticate the fate of the unfortunate runner.

1.2 The Problem

Each year, over 320,000 individuals in the United States (US) suffer from sudden cardiac arrests (SCA) [4]. Approximately 40% of these individuals (128,000) are successfully resuscitated and treated in intensive care units (ICU) [5]. Many SCA patients suffer from anoxic brain injury between the time of their arrest and return of spontaneous circulation (ROSC). 80% of those admitted to ICUs for SCA (100,000) enter an indefinite state of anoxia induced coma, representing approximately 1 in 40 patients treated in ICUs annually [6, 7]. Only 10% of these post-anoxic coma (PAC) patients survive (10,000) and only 3% to 7% of survivors regain normal neurological function (3,000 to 7,000) [6].

The rapid detection and treatment of PAC is associated with more favorable neurological

outcomes [8, 9]. An important goal of neurological prognostication is to avoid negative self-fulfilling prophecies; that is, to prevent patients who might have regained consciousness and achieved meaningful recovery given additional care from being prematurely withdrawn from life-preserving therapies. This goal must be balanced against the risk of encouraging undue optimism in the case of patients who are not on a path of recovery. Recommending unnecessary care can inflict a steep emotional and financial burden on families, with costs running up to $20,000 per day for ICU care [10]).

1.3 Current Clinical Prognostication Approach

The American Academy of Neurology's (AAN) recommended algorithm for PAC prognostication involves a sequence of clinical observations and auxiliary tests performed at specific time points following cardiac arrest. Failure on any of these tests is considered sufficient to render a poor prognosis, with a purported false positive rate of <1% [11]. Specifically, on day one, after excluding a clinical diagnosis of brain death, any signs of twitching (myoclonus) or overt epileptic convulsions (status epilepticus) indicates a poor prognosis. Otherwise, a missing N20 response[1] on days one through three indicates a poor prognosis. If N20 is present, then a serum sample is analyzed to measure the level of neuron-specific enolase (NSE), a protein whose level rises after neuronal injury. If the level of NSE is >33 micro-grams/liter, the physician declares a poor prognosis. Otherwise, on day three post-arrest, a neurological exam is performed. The absence of any purposeful motor response to noxious stimuli, papillary reflexes to light, or blink reflex to stimulation of the cornea during the exam are considered indicators of a poor prognosis.

Historically, the AAN guideline has been considered accurate in predicting poor neurological outcomes (death or severe disability) when severe deficits are present, but provides no guidance in cases where such clear-cut signs are lacking [12, 13]. Furthermore, the current guidelines do not indicate how to proceed for patients treated with therapeutic hypothermia or more recent normothermia protocols, which are known to impact the timing and probability of recovery [14]. These complexities have resulted in issues of incorrect classification

[1] N20: the scalp electroencephalogram for a cortical voltage response at a twenty millisecond delay over the location of the contralateral somatosensory cortex following a somatosensory evoked potential

of patient outcomes when using the standard exam process. Roest *et al.* reported that in as many as 20% of poor outcomes cases, the reliability of predictions is prone to error (that is, beneath 100%) [12].

1.4 Patient Outcomes

The most commonly employed outcome measure for PAC patients is the Cerebral Performance Category (CPC). The CPC is an ordinal measure of outcome (see Table 1). The most common CPC categories at discharge are 5, 1, 2, 3, and 4 respectively. For the purposes of most studies, The CPC is dichtomized into 'good' ($CPC \leq 2$) and 'bad' ($CPC > 2$) outcome classes.

Cerebral Performance Category	**Conscious** (Y/N)	**Cerebral Disability**	**Consequences**
1	Y	Mild	May resume independent activities with minimal complications
2	Y	Moderate	May resume independent activities while requiring some assistance
3	Y	Severe	Will require assistance to perform activities and may involve paralysis, or dementia.
4	N	Vegetative	Requires assistance to survive and will be minimally aware/responsive
5	N	Brain-Death	Requires assistance to survive and will be totally unaware/response

Table 1.1: **Patient Outcomes.** The Cerebral Performance Category, a common measure of neurological outcome for the PAC population

1.5 The Role of Electroencephalography in Prognostication

Investigators agree that the Electroencephalogram (EEG) is a valuable tool for the prognostication of PAC patients [15, 16]. The EEG uses several electrodes placed across the scalp to measure an ensemble of post-synaptic neural electrical field potentials (FP) generated from cortical neurons. FPs are a reflection of trans-membrane currents across collections

of neurons that vary in distribution, size and spiking frequency, making the FP far more complex in nature than spike trains of individual neurons[17]. The strongest component of neuronal activity measured by the EEG comes from the cerebral cortex, the area of the brain most sensitive to anoxic damage [18]. To maximize the prognostic utility of EEG, it must be assessed by neurologists with special training in the art of EEG interpretation [19]. Unfortunately, such neurologists are in low supply, and will be for the foreseeable future. The AAN reports that the supply of practicing neurologists will be 20% below national demands by 2020 [20], and the proportion of such neurologists with specialized training in EEG interpretation may be even lower.

The interpretation of EEG can be challenging even for those with specialized training given the confounding effects of anesthesia, medications, cooling protocols, patient age and other factors on the EEG findings. These complexities have been reported to impact the inter-rater reliability of clinicians interpreting the EEG. A 2015 meta-analysis by Gaspard *et al.* evaluated the inter-rater reliability of standardized EEG descriptors developed by the American Clinical Neurophysiology Society (ACNS) and found an average reliability of 72% (interquartile range of 22.5%) across all descriptors.

Given the lack of qualified experts, PAC patients are often cared for by clinical staff with little or no specialized training in EEG-based neurological examination. These well-meaning (but untrained) staff may utilize the EEG but struggle to meaningfully interpret the waveforms, providing limited benefits to patients while driving up care costs [20]. The challenges of interpretation may help explain the under-utilization of the technology itself, with one recent study reporting that EEG is indicated in roughly 20% of medical/surgical ICU patients, but only measured in 5% [21].

1.6 Quantitative Approaches to Prognostication

Given the number of patients suffering from post-anoxic coma, their high cost of care, the limited number of clinical EEG experts, and the current problems of qualitative prognostication, there is a strong motivation for the development of *quantitative solutions* to evaluation and prognostication.

Multiple quantitative approaches have been proposed to predict patient outcomes. Ex-

isting approaches use one or more of the following modalities: radiological imaging [22], electroencephalography [2], Somato Sensory Evoked Potentials [23], laboratory measures [24], and clinical tests [25]. The discriminative utility of existing approaches is comparable, depending on the precise time they are used for patient assessment. For example, Neuron-specific enolase[2] discriminates patient outcomes 72 hours after cardiac arrest [26] with efficacy similar to studies using quantitative electroencephalography (qEEG) based assessment 12 to 24 hours after cardiac arrest [2, 27]. Each data modality presents its own practical opportunities and challenges.

1.7 qEEG Opportunities and Challenges

A major opportunity of qEEG coma prognostication is its potential for fully-automated *continuous* assessment of patient outcomes at relatively low cost. There are, however, four central challenges, associated with the development of qEEG approaches for the coma prognostication problem:

1.7.1 Challenge 1: Small Sample Sizes

The development of quantitative solutions has been challenging historically given small sample sizes (typically < 100 patients), data heterogeneity steaming from differences in treatment protocols, and missing information (e.g. the time of arrest). Investigators of quantitative EEG approaches for PAC have previously addressed low data volumes by developing thoughtful model heuristics that reflect expert clinical knowledge about PAC prognosis [1].

1.7.2 Challenge 2: A Plethora of Possible Features

EEG signals are complex; they are dynamic, non-Gaussian, non-linear, non-stationary, highly correlated across channels, and state-dependent [28]. This complexity does not eliminate the information content of EEGs so much as it increases the difficulty in extracting it [29]. A majority of existing quantitative EEG features fall into three EEG signal property domains: (A) Complexity features quantify the degree of randomness or irregularity in the EEG signal,

[2] an enzyme that is released into the cerebrospinal fluid when neural tissue is damaged

(B) Category features are qualitative descriptors of signal characteristics or behaviors and (C) Connectivity features quantify interactions between EEG electrodes. Existing methodologies are limited by their use of a small number of QEEG features (ten or less) when predicting patient outcomes. [30–32].

1.7.3 Challenge 3: Relationships are Time-sensitive

With few exceptions, existing approaches tend not to account for changes in the prognostic significance of selected features over the course of patient recovery (using feature values at specific time frames instead). Models that account for the time-senstive nature of the EEG-outcome relationship re-estimate model coefficients in contiguous time-intervals [2]. While re-estimation of coefficients may improve prognostic performance, it also (incorrectly) assumes that the EEG-outcome relationship in a given time-interval is statistically independent from the relationship seen in its neighboring time-intervals. In reality, the EEG-outcome relationship is likely to change gradually, at rates determined by the underlying physiological process of recovery (or deterioration).

1.7.4 Challenge 4: Features are not Data-driven

Even when several features are available, there is always a possibility than an important feature may have been overlooked. It is for this reason that 'deep learning' techniques have gained popularity in recent years. They promise to learn features from data in a fully automated way, albeit at the cost of model interpretability. Deep learning methods have been applied to variety of problems in the areas of computer vision, speech recognition and natural language processing where they achieved impressive results. However, given the small sample sizes of PAC data, investigators have not yet been able take advantage of advanced deep learning paradigms, such as 3-dimensional convolutional neural networks, for this problem.

1.8 Specific Aims

Our specific aims in this work relate to the four challenges of qEEG prognostication described above. To address these challenges, we:

1. **Create a multi-center database** of PAC patients containing continuously recorded measures of EEGs and, when available, confounding factors including age, time until return of spontaneous resuscitation (ROSC), and location of cardiac arrest.

2. **Extract a comprehensive set of EEG features** that predict the categorical neurological outcomes of the patients at hospital discharge, and 6 months following discharge.

3. **Investigate the spatio-temporal evolution** of EEG features, and their relationship with patient neurological outcome at discharge, and 6-months following.

4. **Explore the use of 'deep' learning** for the extraction of novel EEG features of neurological outcome.

1.9 Thesis Outline

Here we provide a brief outline of the upcoming chapters in this thesis. In **Chapter 2**, we provide a brief an Overview of EEG Technology, and a comprehensive review of the PAC literature. In **Chapter 3**, we provide a discussion of the EEG features that may be useful for prognostication in this population based on what was found in Chapter 2. In **Chapter 4**, we review the current state-of-the-art, for the prognostication of coma, after cardiac arrest with an emphasis on qEEG. In **Chapter 5**, we provide an overview of the collected data, and the pre-processing approach used in the thesis. In **Chapter 6**, we describe a novel technique for time-sensitive classification of patient outcomes using a penalized, sequential logistic regression model, and 57 qEEG features. In **Chapter 7**, we deploy a dynamic Bayesian network that explicitly accounts for time-dependent nature of qEEG feature-outcome relationship. In **Chapter 8**, we deploy three deep learning approaches to the raw data for automated feature extraction and classification. Finally, we conclude the thesis in **Chapter 9**, with a discussion about broader implications of the work, deploy-ability and reproducibility.

We advise a reader familiar with (or uninterested in) PAC to skip Chapter 2 of the thesis. The features described in chapter 3, and the methodologies described in Chapters 6 through 8, have general applicability outside of the PAC prognostication domain.

"All of engineering is destined to be *dust on the pages of history*."

– A Nobel Laureate

Chapter 2

Review of The Literature

Synopsis

EEG measures an ensemble of post-synaptic neural electrical field potentials (FP) generated from cortical neurons. The desire to diagnose neurological ailments using EEG is decades old, but the usefulness of EEG has improved with more recent advances in signal processing and machine learning. Previous human and animal studies of PAC have investigated the evolution of EEG activity from the time of arrest to outcome. In general, human studies are typically less methodologically sophisticated than their animal counterparts due to data-driven constraints (e.g. small sample sizes). Human studies indicate that post-anoxic coma EEG recovery proceeds through a number of distinct phases as the patient either recovers or diminishes. To date, there is no known combination of features that authoritatively predict PAC outcomes in humans, although prior investigations have proposed several time- and frequency- domain features as possible indicators. There are several known confounders of outcome for the PAC population that make EEG prognostication more challenging including the quality and time of CPR, age, medications, cause of arrest, and other co-morbidities.

2.1 An Introduction to Electroencephalography

EEG is a collection of several electrodes placed across the scalp to measure an ensemble of post-synaptic neural electrical field potentials (FP) generated from cortical neurons. FPs are a reflection of trans-membrane currents across collections of neurons that vary in distribution, size and spiking frequency, making the FP far more complex in nature than spike trains of individual neurons[17].

The electrical signal measured by the EEG is affected by the conductivity of the skull, scalp, dura, and cerebrospinal fluid, all of which attenuate the current measured by the device. Scalp EEG measures space-averaged activity of 10^7 or more neurons, with a source area of at least a square centimeter. A typical EEG is sampled at 250 Hz and has an amplitude of 10 - 100 μVolts with spectral activity ranging between 1 - 100 Hz. The spectral form of the EEG contains several frequency bands of particular interest. These bands have been shown, over decades of research, to have strong correlations with clinically meaningful states (e.g., sleep/awake). The five bands of interest are delta (δ, 0.5-4 Hz) , theta (θ, 4-7 Hz), alpha (α, 8-15 Hz), beta (β, 16-31 Hz), and gamma (γ, $\geq$ 32 Hz). For clinical cases, the precise number (typically 21) and placement of the electrodes on the scalp are designated by the International Federation of Societies for Electroencephalography [33]. EEG recordings may be either mono-polar or bi-polar, with the distinction being the location of the reference electrode: the ear for mono-polar, and a neighboring scalp electrode for bi-polar.

The desire to diagnose neurological ailments using EEG is decades old, but the usefulness of EEG has improved with more recent advances in signal processing and machine learning. At the turn of the century, EEG was thought to be a highly nonspecific measure of brain function with limited use in populations with cerebrovascular disease [34]. Over the past 15 years (and especially the last 5), the utility of EEG has been more robustly established, with countless papers identifying EEG signatures of sleep stage, anesthetic drug level [35] and neurological ailment [36]. Investigators have also developed several diagnostic EEG methods for the detection of non-convulsive status epilepticus, level of sedation and nonconvulsive seizures, among other conditions[37–40].

2.2 Human vs. Animal Studies

As with most areas of clinical research, prior investigation of EEG-based PAC prognostication may be partitioned into animal and human research (where both prospective, and retrospective studies have been performed). There is evidence that patterns of ischemia-induced EEG alterations are similar across humans, primates, rats, and other mammals [41–43]. Qualitatively similar dynamics in EEG signals have also been observed in both humans and rats, [40, 44, 44]. However, animal studies in this area are often aimed at understanding the physiological mechanisms of post-anoxic injury, and not prognostication per-se. It follows that many animal studies assume the availability of features which are not practically available in a clinical population such as pre-arrest baselines measures of EEG, or precise measures of time until return of spontaneous circulation (ROSC). Hence, while animal-based research is often physiologically insightful, it is also difficult to translate into clinical practice. A subset of the proposed methods in the literature are translatable across species lines, but have not yet been tested in a clinical context. We believe that validation of the methods developed using animals on human subjects is an important and untapped opportunity of scientific exploration in this area.

The human research in this area is less methodologically advanced than its animal counterpart due, in large part, to data-driven constraints. To date, there is no known combination of features that authoritatively predict PAC outcomes in humans [45]. In 2014, Sandroni *et al.* performed a comprehensive review of the field, assessing over 73 works, and found that (due in large part to small sample sizes) the quality of evidence for identified features was 'low or very low' for nearly all surveyed studies [45]. Furthermore, Sandroni *et al.* identified several important knowledge gaps, including ambiguities on the pharmacokinetics of sedatives in temperature controlled patients. Recent reviews emphasize these points, highlighting the simplicity of existing methods, calling for the standardization of continuous time evaluation, and an extension of final patient evaluation to no earlier than 72 hours following normothermia [46, 47].

2.3 EEG Evolution From Arrest to Outcome

Borjingin *et al.* report that the EEG moves through four distinctive phases in rats following a cardiac arrest. Starting from the time of the last heartbeat (A) there is a loss of oxygenated blood pulse which is characterized by a decline in EEG amplitude with enhanced gamma oscillations. This is followed by (B) a burst of δ wave activity, along with θ and high frequency γ activity. Afterwards, (C) EEG amplitudes decline towards 10 μVolts, with most activity in the low-γ range, coupled to θ oscillations across channels thereafter. In the last phase, (D) EEG activity remains below 10 μVolts, with no characteristic changes [48].

Studies in animal models have shown that after resuscitation and resumption of oxygen delivery to the brain, recovery begins at the brain-stem, followed by activity in the deep structures of the brain, and finally a return of cortical activity. It is not until this last phase that consciousness returns [46]. Brain stem recovery could occur within hours but it is the recovery of the other areas that ultimately dictates the extent of resumed consciousness [46].

Human studies indicate that PAC EEG recovery proceeds through a number of distinct phases, beginning with isoelectricity (flat-line EEG), followed by burst-suppression, followed by an evolution towards continuous activity with varying degrees of complexity [40, 48].

Evoked potential responses (event-related potentials, ERP) in brain- injured animals have also demonstrated characteristic patterns of evolution predictive of neurological recovery from coma which are strongly associated with neurological outcomes [40, 49]. For a recent review of evoked potentials see Koenig *et al.* [50].

2.4 Important Rhythms

For over a decade investigators have repeatedly identified the ratio of α to δ band activity as a predictive feature of post anoxic coma (PAC) outcomes in humans. [3, 51, 52]. α activity arises from thalamic pacemaker cells in humans and has been associated with the occipital region of the brain while δ activity is strongly associated with depth of sleep. A recent study by Van *et al.* also identified infraslow activity (< 0.1Hz) as an important feature of brain pathology of humans. Infraslow activity has been associated with cortical excitation and the severity of postanoxic encephalopathy [53].

Colgin *et al.* described θ rhythms as an important feature of consciousness recovery in rats. θ rhythms are an indication of the convergence and synchronization of information in the brain, with discrete packets of sensory information corresponding to each θ cycle [54]. Given its role in sensory fusion, θ may be an important feature for coma prognostication in humans.

Rosetti *et al.* report that coma, which is dominated by α or θ activity (or transitions between the two) and higher activity levels in the posterior electrodes relative to the frontal electrodes, is commonly observed in human PAC patients. The observed α band dominance in these cases may arise from either drugs, cortical laminar necrosis, or deafferantation of the cerebral cortex. The continuous slowing towards a diffuse EEG in patients with α band dominance is associated with poor outcomes while α dominated EEG which is reactive to physical patient stimulus is associated with regained consciousness [13]. Coma which is dominated by activity in the 11- to 14 Hz band (known as spindle coma) is also commonly observed in PAC patients and is thought to have a mildly positive association with outcome [13].

Recently, Deng *et al.* reported that the rate of γ-band recovery shortly following return of spontaneous circulation (ROSC) was strongly associated with functional outcomes in rats [6]. γ band activity is thought to be related to cerebral metabolism which, when lowered, may protect the brain during arrest and thereby predict outcomes [48]. The prognostic value of the γ band is unclear in humans however, as γ is difficult to robustly measure using surface electrodes. Borjigin *et al.* identified coupling between (A) θ phase and low-γ (30–40 Hz) amplitude, and (B) between α phase and mid-γ (42–56 Hz) amplitudes in the first 20 mins of the arrest in rats [48].

2.5 Confounding Features of EEG Activity and Outcome

There are several known confounders of outcome for the PAC population that can make EEG prognostication more challenging. The severity of anoxic injury depends on several factors including the quality and time of CPR, age, medications, cause of arrest, and other co-morbidities [46]. The dependency of the EEG on age, gender, temperature and medications increase the complexity of deriving meaningful features that are predictively valid for all

patients, especially in the early hours of treatment [13]. Here we will discuss the effects of several of these confounders on ICU outcome and EEG.

2.5.1 Age and Gender

Not surprisingly, old age is associated with negative outcomes in the PAC population. Particularly, patients over the age of 50 are known to experience difficulties with survival and recovery [18]. Roest *et al.* found that patients younger than 50 tend to perform twice as well as patients older than 50 at ICU discharge, and 180 days after discharge [12]. It has been known for several decades that age influences the time and frequency domain characteristics of the EEG. As patients age, the amplitude of the EEG is attenuated. The dominant frequency changes from fast α to θ from 6-16 years of age and then linearly changes to become more α dominated again with age [55]. Gender is also known to affect the EEG with normal female activity exhibiting higher power density than men in δ, θ and low α bands during sleep [56].

2.5.2 Time-Related

The length of time until return of spontaneous circulation (ROSC) is another strong predictor of outcomes for this patient population. In general, the longer a patient is without circulation, the greater the extent of the anoxic damage. Many studies consider 20 or more minutes without circulation as the critical threshold after which the chances of normal neurological function are severely curtailed. While the length of ROSC is strongly associated with outcomes, it it certainly not definitive. Cerebral neurons have been shown to recover electrical and chemical activity even 60 mins after circulatory arrest [57]. The duration of non-medically induced unconsciousness is another indicator of outcome for this population, with one report stating that 88% of patients with unconsciousness lasting more than 6 hours have negative outcomes [18].

2.5.3 Temperature

It is a well known fact that even small changes in temperature can significantly influence the outcomes of patients suffering from ischemic brain injuries [11, 58, 59] with elevated

temperatures associated with negative outcomes. It has become standard practice to therapeutically cool PAC patients for 12 to 24 hours after resuscitation. The current guidelines recommend cooling patients to anywhere between 32 and 34 degrees Celsius for 24 hours following an arrest, as this slows brain metabolism and is thought to help prevent further brain injury. Despite these guidelines, a precise understanding of the effects of temperature on outcome are not well understood. [59]. Therapeutic hypothermia is known to improve outcomes while simultaneously complicating the exam process [13]. It has been hypothesized that hypothermia stretches out the effects of drugs, such as sedatives and there have been calls to revise treatment guidelines, including delaying the time of prognostication beyond 72 hours following an arrest [46]. Temperature also has known effects on the EEG, attenuating amplitude and modifying spectral properties. Increased body temperature has been associated with faster α rhythms while cooling is associated with EEG slowing in multiple animal models [60]. Simulations by Deboer *et al.* estimate that temperature decrease in humans should increase spectral activity around 12Hz and decrease activity around 15Hz, with smaller reductions in the 4Hz and 8Hz bands.

2.5.4 Rhythm at Arrest

The type of arrhythmia detected at arrest is another feature which is predictive of outcome, with several investigators reporting ventricular fibrillation and ventricular tachycardia as less strongly associated with poor outcome than other arrest rhythms (asystole, pulseless electrical activity, or unknown) [61].

2.5.5 Medications

Sedatives, anti-psychotics and anti-epileptic agents are also commonly deployed during treatment of the PAC population. In general, these drugs are used to sedate patients (to decrease cerebral metabolic rates), and reduce seizure activity (to prevent additional damage). Although their utility is well established in animal models, there remains confusion and conflicting reports about the precise effects of these drugs on outcomes in humans. [62, 63]. Like age and temperature, drugs are known to confound the time and frequency characteristics of the EEG. To complicate matters further, the pharmacodynamics of these drugs are actually

influenced by the age and temperature of the patients themselves [64]. In Table 2.1, we illustrate the known effects of several drugs on the EEG [65]. The nature of drug interaction with subject behavior, physiological response, and possible neural circuit mechanisms are extensively discussed in Brown *et al.* [66].

Drug	**Use**	**Initial EEG Effects**	**General EEG dose-response**	**Additional effects**
Barbiturates Propofol Etomidate	Sedation	Reduced occipital alpha activity, Increased frontal beta-activity	Larger amplitude, Lower frequency, Frontal spindles with activity in 1-3 Hz, Isoelectric state	Burst supression, Epileptiforms
Ketamine	Sedation	Frontally dominant theta activity Increased amplitude	Intermittent high amplitude delta-activity and low-amplitude beta-activity	No isoelectric state Epileptiforms
Benzodiazepines	Seizure management	Reduced occipital alpha activity, Increased frontal beta-activity	Larger amplitude, Lower frequency	No burst supression No isoelectric state Epileptiforms
Opioids	Sedation	Reduced occipital alpha activity, Increased frontal beta-activity	Larger amplitude Lower frequency	No isoelectric state Epiletiforms

Table 2.1: **Effects of Common Anesthetics on EEG.**

2.6 PAC Outcome Measures

The single most commonly employed outcome measure for this population of patients is the Cerebral Performance Category (CPC). The CPC is an ordinal measure of outcome. Importantly, CPC scores at ICU discharge have been associated with longer-term outcomes for the PAC population [67, 68]. The most common CPC categories at discharge are 5, 1, 2, 3, and 4 respectively. In a survey of nearly 1000 PAC patients, Phelps *et al.* identified that 82% of patients in CPC categories 1-4 survived one year following discharge. CPC at

discharge is associated with longer term CPC measures at 6 and 12 months but the causal factors driving recovery remain unclear [68].

Other less commonly used measures include the Glasgow Outcome Score and the modified Rankin Scale (mRS). The Neurological Deficit Score is another established estimate of neurological outcome in clinical settings [69–71]. The score provides a value between 0 and 80 (with 0 being the worst, and 80 the best) on the basis of subject behavior, arousal levels, seizure activity and brain-stem function among other things. The scale was developed by Thakor *et al.* from a combination of existing animal and human scales [72–75].

"How many features are there?
HOW. MANY. FEATURES?"

– A Professor

Chapter 3

Quantitative EEG Features

Synopsis

EEG signals are complex; they are dynamic, non-Gaussian, non-linear, non-stationary, highly correlated across channels, and state-dependent [28]. This complexity does not eliminate the information content of EEGs so much as it increases the difficulty in extracting it [29]. A majority of quantitative EEG features attempt to describe complexity (signal chaos), category (clinically relevant signal behaviors) and connectivity (cross-channel interactions). We extracted a total of 57 EEG features: 29 measuring EEG category, 21 measuring EEG complexity and 7 measuring EEG connectivity.

3.1 Overview

For the purposes of PAC prognostication, EEG features may be divided into three EEG signal property domains: (A) Complexity features quantify the degree of randomness or irregularity in the EEG signal; (B) Category features quantify the degree to which brain states fall into certain key EEG patterns likely to carry prognostic significance and (C) Connectivity features quantify interactions across EEG electrode. In Figure 3.1 we provide a example features from each signal property domain.

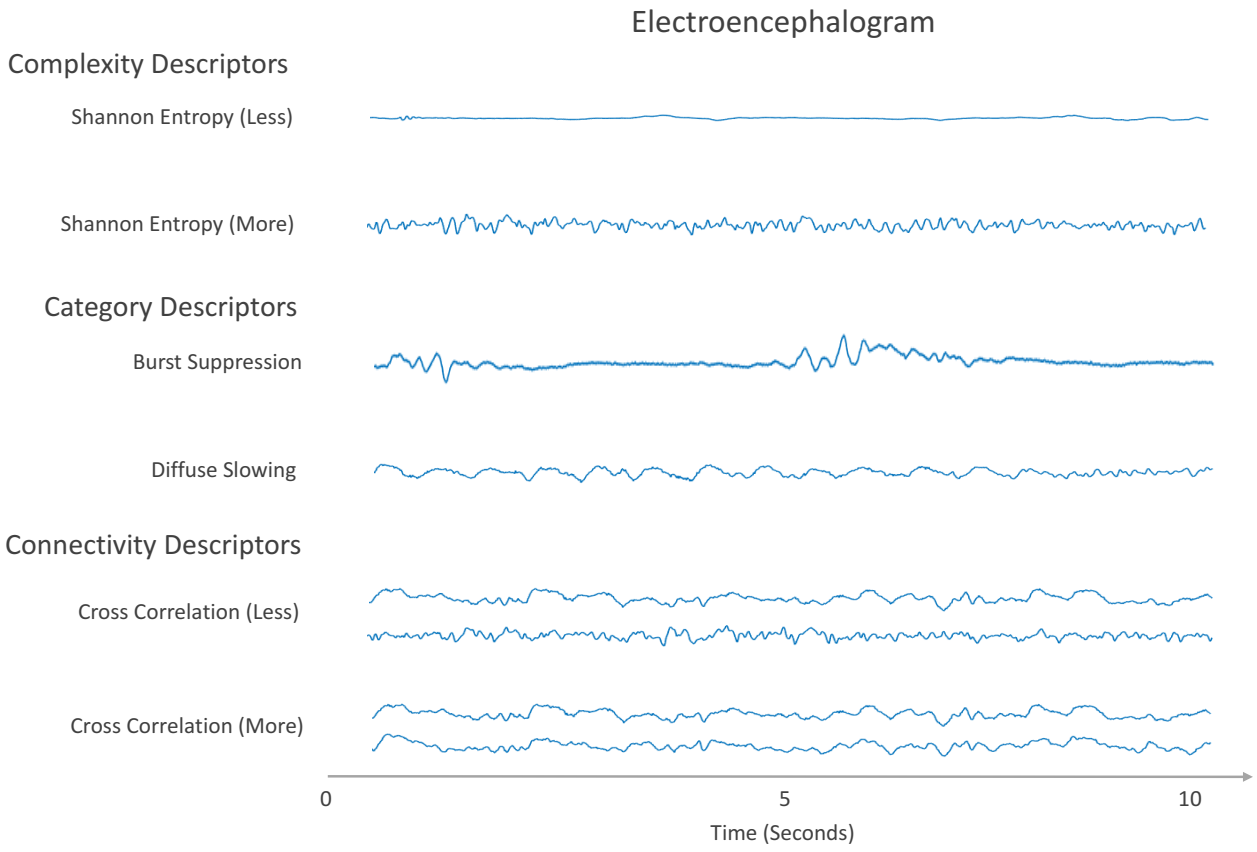

Figure 3.1: **Exemplary EEG Features.** Features (descriptors) quantify the complexity of EEG waveforms (e.g. *Shannon entropy*), dependencies between channels (e.g. *cross-correlation magnitude*), and clinically relevant categories of activity (e.g. *burst suppression*).

In Table 7.1, we provide a list of all features that were extracted for the purposes of this thesis. A detailed discussion of quantitative EEG features is presented thereafter. Code to

generate the features shown in Table 7.1 are publicly available in an on-line repository[3]

Signal Descriptor	Ref.	Brief Description
Complexity Features		degree of randomness or irregularity
Shannon Entropy	[76]	additive measure of signal stochasticity
Tsalis Entropy (n=10)	[49]	non-additive measure of signal stochasticity
Subband Information Quantity	[77]	entropy of a wavelet decomposed signal
Cepstrum Coefficients (n=2)	[78]	rate of change in signal spectral band power
Lyapunov Exponent	[79]	separation between signals with similar trajectories
Fractal Embedding Dimension	[80]	how signal properties change with scale
Hjorth Mobility	[81]	mean signal frequency
Hjorth Complexity	[81]	rate of change in mean signal frequency
False Nearest Neighbor	[82]	signal continuity and smoothness
ARMA Coefficients (n=2)	[83]	autoregressive coefficient of signal at (t-1) and (t-2)
Category Features		clinically grounded signal characteristics
Median Frequency		the median spectral power
δ band Power		spectral power in the 0-3Hz range
θ band Power		spectral power in the 4-7Hz range
α band Power		spectral power in the 8-15Hz range
β band Power		spectral power in the 16-31Hz range
γ band Power		spectral power above 32Hz
Median Frequency		median spectral power
Standard Deviation	[84]	average difference between signal value and it's mean value
α/δ Ratio	[1]	ratio of the power spectral density in α and δ bands
Regularity (burst-suppression)	[1]	measure of signal stationarity / spectral consistency
Voltage $< (5\mu, 10\mu, 20\mu)$		low signal amplitude
Normal EEG	[85]	Peak spectral power >= 8Hz
Diffuse Slowing	[85]	indicator of peak power spectral density less than 8Hz
Spikes	[85]	signal amplitude exceeds μ by 3σ for 70 ms or less
Delta Burst after spike	[85]	Increased δ after spike, relative to δ before spike
Sharp spike	[85]	spikes lasting less than 70 ms
Number of Bursts		number of amplitude bursts
Burst length μ and σ		statistical properties of bursts
Burst band powers $(\delta, \alpha, \theta, \beta, \gamma)$		spectral power of bursts
Number of Suppressions		number of segments with contiguous amplitude suppression
Suppression length μ and σ		statistical properties of suppressions
Connectivity Features		interactions between EEG electrode pairs
Coherence - δ	[1]	correlation in in 0-4 Hz power between signals
Coherence - All	[86]	correlation in overall power between signals
Mutual Information	[87]	measure of dependence
Granger causality - All	[88]	measure of causality
Phase Lag Index	[89]	association between the instantaneous phase of signals
Cross-correlation Magnitude	[90]	maximum correlation between two signals
Crosscorrelation - Lag	[90]	time-delay that maximizes correlation between signals

Table 3.1: **EEG Features.** The 57 QEEG features fell into three EEG signal property domains: Complexity features (21 in total), Category features (29 in total), Connectivity features (7 in total)

[3]https://github.com/deskool/ComaPrognosticanUsingEEG

3.2 Features of Complexity

An EEG can be modeled as a combination of deterministic and stochastic signal components. At the core of complexity-based EEG analysis is the assumption that the information content in the signal is driven primarily by its stochastic nature. In general, more complexity is considered a sign of a healthier brain [1]. Our selected features of EEG complexity included: Shannon Entropy, Tsalis Entropy (with q ranging from 1:10), Cepstrum coefficients, Sub-band Information quantity, Lyaponov exponent, fractal dimension, Hjorth mobility/complexity, false nearest neighbor embedding dimension and the coefficients of a second order auto-regressive moving model. In Figure 3.2, we illustrate the estimated complexity of several toy signals using our selected complexity features. We described the features in greater detail, thereafter.

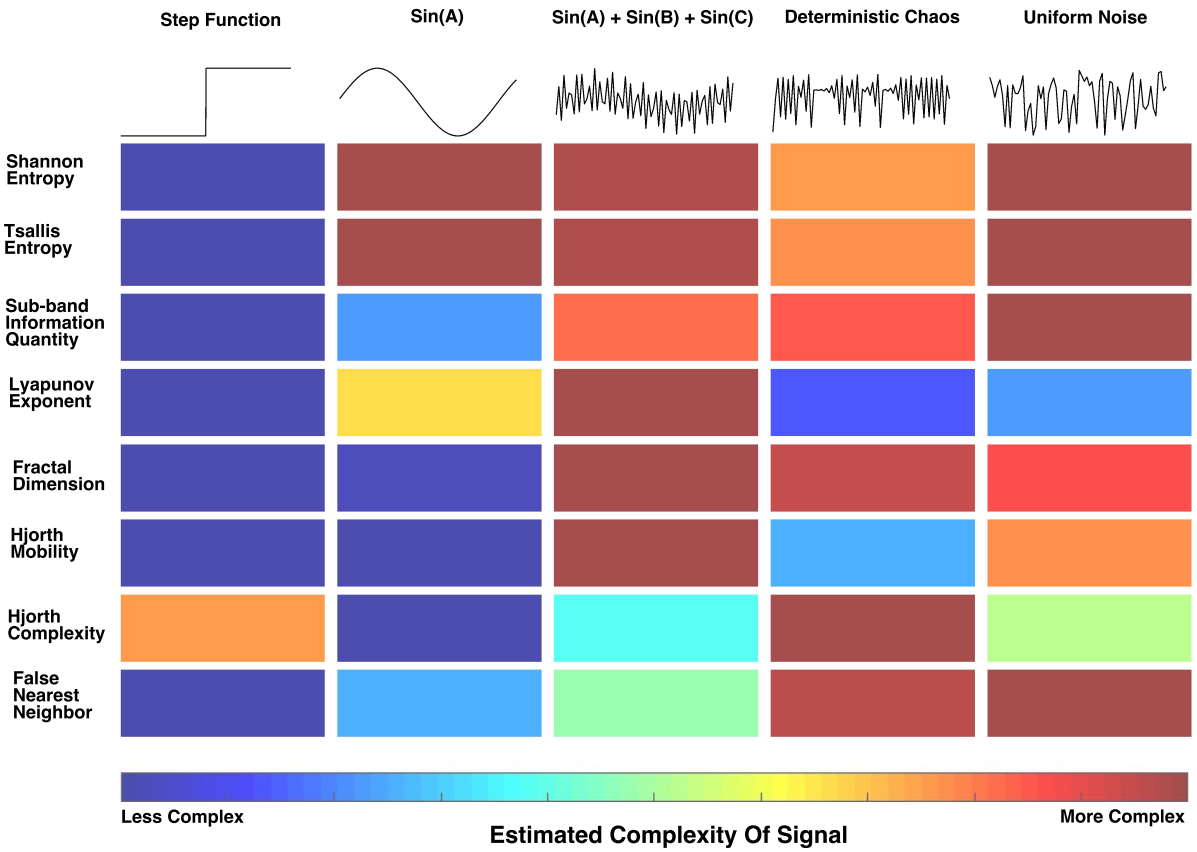

Figure 3.2: **Complexity Features.** The estimated complexity of several toy signals using our selected features of complexity.

3.2.1 Entropy

The earliest, and perhaps best-known measure of complexity is additive Entropy (Shannon's Entropy), which is known to be associated with neurological outcomes in PAC patients [1].

$$H_{sh}(X) = -\sum_{x} p(x) log_2 p(x)$$

One fundamental assumption of additive entropy is the property of entropic additivity. That is, for two random channels of the EEG X_1 and X_2, Shannon entropy assumes a factorizable joint distribution $p(x_1, x_2) = p(x_1)p(x_2)$, which results in

$$H_{sh}(X_1, X_2) = -\sum_{x_2}\sum_{x_1} p(x_1)p(x_2)(log_2[p(x_1)] + log_2[p(x_2))])$$

$$= -\sum_{x_2}\sum_{x_1} p(x_1)p(x_2) log_2[p(x_1)] - \sum_{x_2}\sum_{x_1} p(x_1)p(x_2) log_2[p(x_2))]$$

$$= -\sum_{x_1} p(x_1) log_2[p(x_1)] - \sum_{x_2} p(x_2) log_2[p(x_2))]$$

$$= H_{sh}(X_1) + H_{sh}(X_2)$$

The assumption of entropic additivity may cause additive entropy to over-estimate the complexity of a system with dependent random variables, where $p(x_1, x_2) = p(x_1)p(x_2|x_1)$ [91]. As indicated by Thakor *et al.* over a decade ago, this limitation makes additive entropy ill-suited for the measurement of complexity in systems with highly correlated signals, long-range interactions, memory, or abrupt changes, all of which are common in a standard EEG signals [28].

Given the limitations of Shannon entropy, Thakor's group proposed Tsallis entropy as a potential alternative for the robust estimation of complexity in EEG signals [49]. Tsallis Entropy is practically similar to additive entropy but does not assume entropic additivity. That is, for two random channels of an EEG X_1 and X_2, The Tsalis entropy evaluates to

$$H(X_1, X_2) = H_q(X_1) + H_q(X_2) + (1 - q)H_q(X_1)H_q(X_2)$$

Where the parameter q may be understood as a measure of how strong the correlations are between the two signals. While Tsalis entropy has been validated in animal PAC models, it has not yet been utilized in human subjects [92], presenting a opportunity for investigation.

Given the non-stationary nature of EEG signals, it is common to compute time-dependent entropy, which is simply a measure of entropy computed within multiple subsequent data epochs. Bezerianos *et al.* used time dependent entropy to quantify evolving brain injury levels in PAC rats and reported that changes in entropy over time could be used to effectively discriminate between various levels of injury [92]. These findings provide evidence that changes in EEG complexity over time may provide important prognostic features in humans.

3.2.2 Information Quantity

Shin *et al.* reported that the entropy of wavelet decomposed EEG signals, a feature termed 'Information Quantity' (IQ), was superior to entropy of the raw EEG waveform in its ability to predict PAC outcomes in rats [93]. The IQ is computed by applying the discrete wavelet transform (DWT) to the EEG signal to first extract sub-bands of clinical interest. The IQ is then computed as the Shannon entropy of the decomposed signal. More recently, Jia *et al.* used the IQ to investigate the effects of temperature on PAC outcomes in rodents [8]. The study reported an ability to reliably predict neurological outcomes in as little as 60 minutes after return of spontaneous circulation in PAC rats when using the measure. More recently, Deng *et al.* proposed an extension to the IQ, denoted the 'Sub-band Information Quantity' (SIQ) which is the average entropy over the individual components of the DWT decomposed EEG. Deng et al. reported that the SIQ was superior to the IQ for prognostication of PAC outcomes in rats [6].

3.2.3 Cepstrum

Entropy-based features are one of many ways to measure system complexity. To our knowledge, other techniques in linear/nonlinear dynamical analysis of EEG have not been sys-

tematically applied to this patient population although they are highly prevalent in EEG analysis in general. One exception is the work of Geocardin *et al.* over a decade ago, who proposed a method of estimating the cerebral damage of a post-anoxic brain injury in rats [39] using a differential cepstrum measure [94] (the rate of change in spectral bands), and showed that the cepstrum was superior for prognostication in rats as compared to other conventional spectral features.

3.2.4 Measures of Choas

Other features of complexity that may be of value and have been applied to EEG in other contexts include: Lyapunov exponents [95], fractal and correlation dimension [96], non-linear forecasting error [97], and false nearest neighbor [98].

3.3 Features of Connectivity

Central to normal brain function is the interaction of functionally specialized areas of the brain. Features of connectivity aim to quantify these interactions. Robustly measuring connectivity using EEG signals can be challenging due to the effects of field potential spread ("volume conduction") on sensor level activity [17]. Nevertheless, there is ample literature discussing approaches to measure functional connectivity using EEG [99]. These techniques have been deployed to measure connectivity in the context of mental illness [100], perception [101]. However, there is relatively less work which has measured the prognostic significance of connectivity changes in the PAC population. Our selected features of connectivity include: coherence in δ band, coherence in all bands, phase lag index, cross correlation magnitude, cross correlation lag, mutual information and granger causality. In Figure 3.3, we illustrate the estimated connectivity of several toy signals using our selected connectivity features. We described the features in greater detail, thereafter.

3.3.1 Phase Lag Index

A notable recent paper by Beudel *et al.* performed a graph theoretic analysis of PAC patients and showed a statistically significant correlation between increased EEG connectivity and

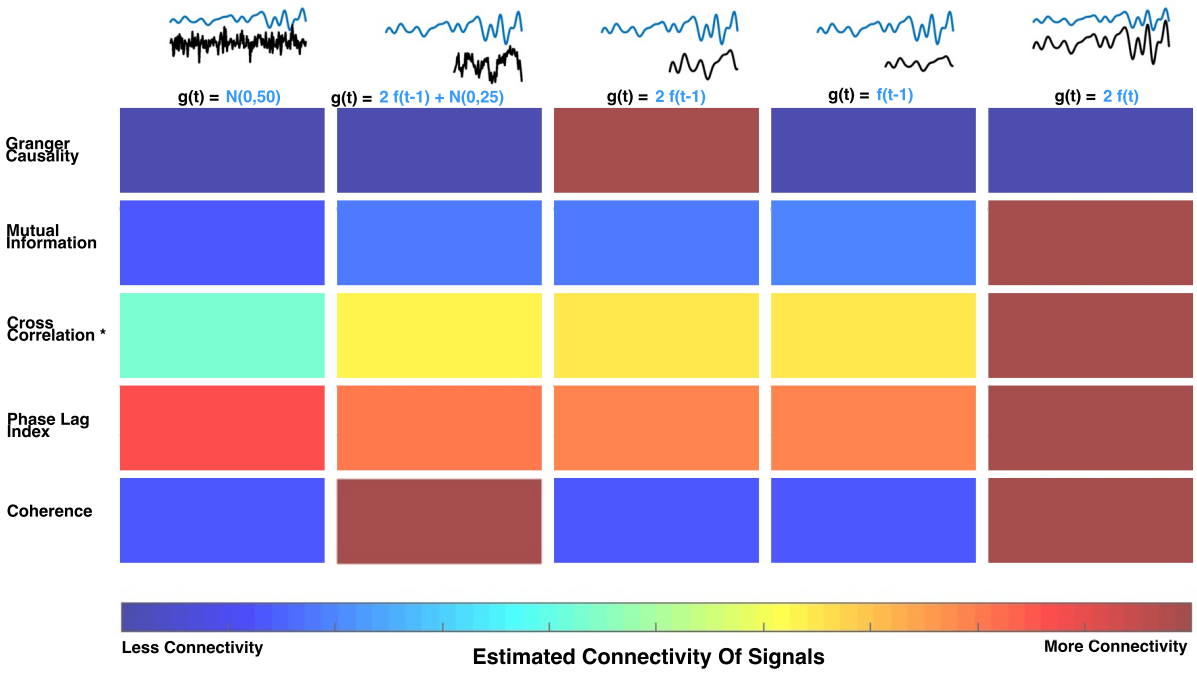

Figure 3.3: **Connectivity Features.** The estimated connectivity of several toy signals using our selected features of connectivity.

positive outcomes (as measured by cerebral performance category - CPC) [102]. Stam *et al.* defined a measure of EEG connectivity, the Phase Lag Index, which quantifies the difference in the distributions of phases between two signals[89] and was also found to be predictive of PAC outcomes.

3.3.2 Cross Correlation and Coherence

Signal cross-correlation and coherence (a measure of spectral cross-correlation) are symmetric (i.e. direction-less) measures of functional connectivity that have also been used for prognostication in PAC [3, 99]. Cloostermans *et al.* found that early increased coherence in the δ band, specifically, is associated with negative outcomes.

3.3.3 Other Connectivity Features

There are several other EEG connectivity metrics that have not yet been applied to the PAC population but are prevalent in other areas of EEG signal processing. The *Phase locking value* is a symmetric measure of phase synchronization that, unlike coherence, is insensitive

to differences in signal amplitudes. It has been used in EEG source connectivity analysis [99], albeit not for PAC prognostication. Mutual information is another commonly employed feature of connectivity in EEG analysis. Mutual Information (MI) is a measure of dependency between sources that can capture non-linear relationships. While also symmetric, MI can also be used to determine the direction of connections via time-lagging. Both Granger causality and multivariate auto-regression can (and have) been used to gauge connectivity and the direction of information flow in EEG. Independent Component Analysis (ICA) is another technique with a long history of use in the EEG community that may be utilized to gauge connectivity [103].

3.4 Features of Category

EEG activity for the PAC population often falls into one of several qualitatively distinct, named states: isoelectric, low-voltage, burst-suppression, generalized periodic discharges, electrographic status epilepticus, diffuse irregular slowing ,or normal. Over the years, there has been extensive investigation into the characteristics of these states, variations within them, and their associations with patient outcomes [45]. Juan *et al.* performed a useful meta-analysis of the literature to determine EEG features of category that have been shown to be relevant for PAC prognosis. For negative prognostication, these include: burst suppression, isoelectric/low-voltage activity, limited reactivity, status epilepticus, and epileptiform transients. In this section we will provide a brief overview of these states, as they pertain to PAC prognostication. Our selected features of category include: standard deviation, regularity, EEG frequency band power (δ, θ, α, β, γ, μ), the ratio of α/δ band power, signal amplitude less than 5 μV, signal amplitude less than 10μV, signal amplitude less than 20μV, "normal" EEG, diffuse slowing, number of epileptiform spikes, epileptiform peaks followed by increase in delta band power [104]. We quantified burst suppression using the following features: burst length (mean and std), suppression length (mean and std), number of bursts, number of suppressions. In Figure 3.4, we illustrate the estimated category of several toy signals using our selected category features. We described the categories in greater detail, thereafter.

3.4.1 Normal:

The EEG is a complex, state and age dependent signal, even in normal subjects. Hence, defining a "normal" EEG is non-trivial. The spectral content of conscious, normal, awake patients is reported to be above 13Hz (of course, this is modulated by factors including gender, age and others). This spectral content is modulated when patients close their eyes, adding additional power to the 8-13Hz (α) band. A normal EEG is also uniformly distributed, with amplitude above 20 μVolts. In prior studies of PAC, a "normal" EEG, especially early in the stay, has been strongly associated with positive outcomes. For our purposes, normal EEG will be defined defined as a continuous activity exceeding 20 μVolts, with dominant spectral content above 8Hz and no periodic discharge.

3.4.2 Low Voltage, and Isoelectric:

An EEG is considered to be in a low voltage state when activity remains below 20 μVolts and is considered to be isoelectric when there is no appreciable EEG activity. Cloostermans *et al.* reported that the presence of isoelectric or low-voltage EEGs 24 hours following resuscitation is nearly twice as sensitive as somatosensory evoked potential responses for

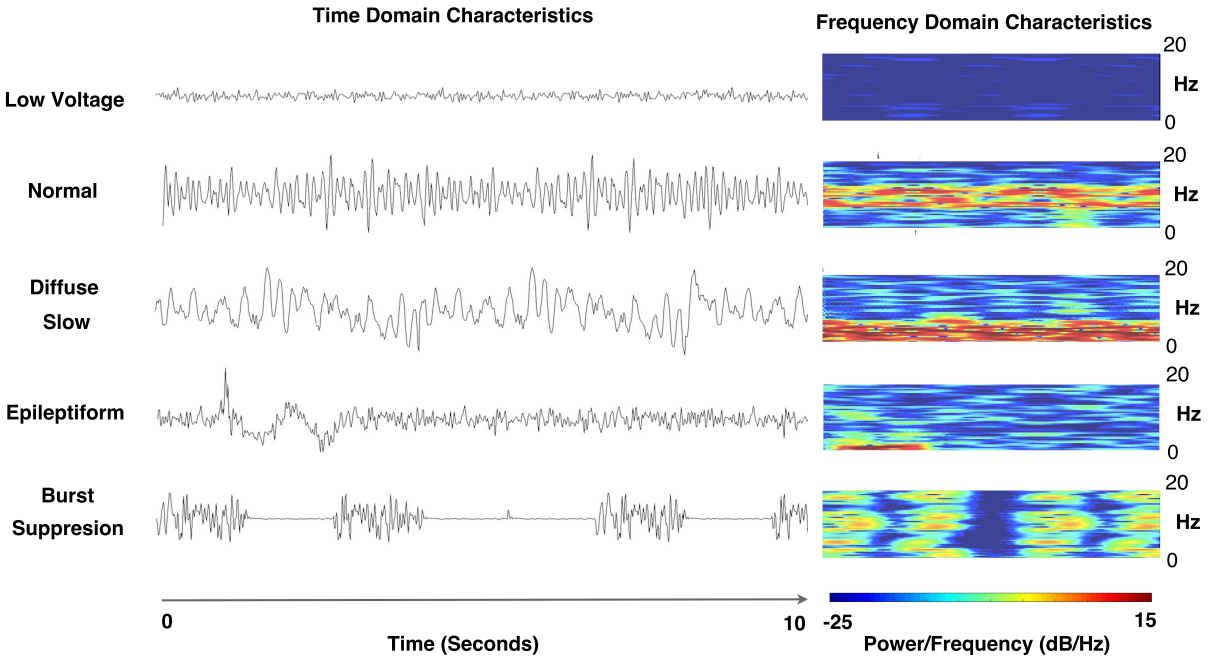

Figure 3.4: **Category.** An overview of our selected categories of interest.

predicting poor outcomes while continuous (normal) activity within 12 hours of resuscitation predicts positive outcomes [105]. These findings have been validated more recently by other investigators [106].

3.4.3 Slowing

Slowing refers to a shift in the spectral content of the EEG towards zero and is defined as a continuous EEG pattern with a dominant frequency less than 8Hz [105]. In general, the slower and more persistent the rhythm, the more severe the underlying structural damage [107].

3.4.4 Generalized Periodic Discharges, and Seizures:

Multiple authors have found an association between generalized periodic discharges and negative outcomes [108, 109]. Electrographic status epilepticus (ESE) occurs in 10-35% of PAC patients and is generally considered an indicator of poor prognosis [32]. Following the advent of therapeutic hypothermia, the relationship between ESE and outcome became less certain [13]. Patients who undergo ESE are often treated with anti-epileptic drugs but there is skepticism that suppression of ESE improves outcomes [62, 63]. Of those that survive, few recover beyond a vegetative state [110].

In 2015, Cronberg *et al.* argued that status epilepticus was not an established feature of poor prognosis and called for more aggressive treatment of PAC patients with this condition [111]. Within the same year, a response to this this letter was authored by Rossetti *et al.*, arguing that status epilepticus was in fact an established feature of poor prognosis [112]. Epileptic discharges with low power, or those with a continuous background are associated with better outcomes [32], although there remains uncertainty [106].

3.4.5 Burst Suppression:

Burst suppression (BS) may be characteristic of cerebral circulatory arrest survivors in critical care[113]. As the name implies, BS is a pattern of EEG characterized by intermittent periods of diffuse voltage (suppression), alternating with higher voltage EEG activity (bursts) [49]. While the duration of suppressions can vary, the active periods are reported as being

more stable, usually lasting between 1-2 seconds [114]. Activity should exceed 20 μVolts to be considered a burst [105].

Neurologically, burst suppression is thought to arise from a desynchronysation between pacemaker neurons in the reticular thalamic nucleus, and the cortex [115]. More specifically, the bursts in activity seen in the EEG reflect the activity of cortical neurons [115], while the periods of inactivity are thought to result from the refractory period of these cortical neurons [114]. Cortical neurons are known to play a part in the generation of slow waves, which contribute the most to the amplitude of the EEG measurement. Hence, the higher the amplitude of a burst event, the more likely that cortical neurons are undamaged. The frequency content of bursts may be indicative of damage to the cortex, and shorter durations of suppressions may be associated with more rapid stabilization of cortical neuron function.

Hence, BS has potential value in the diagnosis and prognostication of neurological outcomes. Some techniques to quantify burst suppression include the use of BS duration, amplitude, and event frequency. Studies using rats have found that an EEG which is flat during suppression periods, with infrequent bursts, or low amplitude bursts are all indicators of poor neurological outcomes[38]. Studies in humans report burst-suppression as a predictor of poor outcome [116], but the predictive value of burst-suppression depends on when it is seen and its characteristics. Hofmeijer et al. reported that burst suppression with identical bursts is even more strongly predictive of poor outcomes [117]. Burst suppression 24 hours after return of spontaneous circulation was also found to be a strong predictor of poor outcomes in humans [106].

Westhall *et al.* recently evaluated the efficacy of qualitatively derived EEG features (by specialists) for the prediction of dichotomous outcomes (CPC>2 vs CPC $<$ 3) in 103 patients following rewarming. The authors found that some malignant EEG features (suppression/burst-suppression) predicted negative outcomes better than others (nonreactive background/rhythmic patterns) and that a joint consideration of these features was most effective for outcome prediction. Similar results were reported by Monteiro *et al.* on a study of 92 patients. The authors found that diffuse slowing and burst suppression of EEG were associated with negative outcomes while epileptiform activity [118] was a less clear predictor.

"Do you have an enemy? *Make one.*"

–A Scientist

Chapter 4

State of the Art Review

Synopsis

A review of 26 papers reveals that prior PAC prognostication methodologies benefited from features that captured diverse aspects of EEG activity. The most successful attempts in the literature have accounted for the temporal sensitivity of EEG-outcome relationship to improve performance. Given the current state-of-the-art, there are opportunities to develop models using additional features of complexity, connectivity, and category, while also accounting for the time-sensitive nature of the EEG-outcome relationship.

4.1 Existing Methodologies

We now present a structured review of existing studies that have attempted to prognosticate outcomes in the PAC population. We conducted a *Google Scholar* and *PubMed* search using the following terms: 'EEG', 'postanoxic coma', 'prognostication', 'prediction', and 'cerebral performance category.' Our inclusion criterion for the review was a reported quantitative measure of performance: Area Under the Receiver Operating Curve (AUC), sensitivity, and/or specificity. A total of 26 papers were found to be directly relevant to this work. The total review covered 2,785 subjects, with a mean of 107 (+/- 76) subjects per study. In Table 4.1, we outline the population size, features used, and predictive performance achieved for the surveyed studies that met our inclusion criteria.

Many of the studies in Table 4.1 performed ROC analysis using a single variable. In all surveyed studies, a positive outcome was defined as a cerebral performance category (CPC) greater than or equal to 2. Logistic regression was frequently employed for modeling. We also found studies that used decision trees [3], random forest [2], Naive Bayes [119] and heuristics [1]. Studies that utilized laboratory measures (including neuron specific enolase (NSE) and S100 calcium-binding protein) performed well at 72 hours following resuscitation. The best reported methodology [19] utilized a multi-modal approach at 48 hours following the arrest, combining both laboratory and EEG measures.

The surveyed studies with the best performance within 24 hours of resuscitation all employed EEG. We found that the highest performing papers that utilized EEG for prognostication were performed by Cloostermans *et al.*[1, 2]. The papers outline a real-time quantitative EEG measure, The Cerebral Recovery Index (CRI), which was developed by the investigators and is run at the bedside to predict outcomes for post-anoxic patients in the ICU. The CRI uses a combination of 5 quantitative features: Inter-channel Coherence in the delta band, signal regularity (a measure defined by authors that detects burst suppression), the alpha to delta band power ratio, signal entropy, and the standard deviation. In the earlier instantiation of the method [1], the authors determined the exact parameters of their model using expert knowledge. In the latest version of the CRI model, a random forest classifier is used to determine model parameters [2]

Our review of the literature provides two essential conclusions that will motivate our analysis in later chapters.

4.1.1 Conclusion 1: EEG prognostication methodologies benefit from features that capture diverse aspects of EEG.

While many of the EEG-based prognostication techniques in the surveyed literature rely on features that describe EEG category (qualitative descriptors of signal character and behavior, e.g. burst suppression), a relatively smaller set of studies used features that describe EEG complexity (measures of signal regularity or predictability, e.g. entropy) and/or EEG connectivity (measures of cross-channel interactions, e.g. coherence). Methodologies that used a greater variety of these features also tended to produce the best prognostic performance.

4.1.2 Conclusion 2: Accounting for the temporal sensitivity of EEG features may improve performance.

EEG is well suited for rapid ($<$ 24 hours) prognostication of PAC outcomes compared to other approaches [47, 139, 140]. However, the prognostic relevance of EEG features changes over the course of patient recovery [124, 136]. To address this, many of the surveyed EEG based prognostication methodologies extract features at set time intervals (e.g. 12, 24, 48, 72 hours) but often do not consider how the changes in the EEG over time may be predictive of outcome. This creates an opportunity for modeling approaches that explicitly incorporate the temporal properties of the data.

Ref.	**Sample Size**	**All Features**	**Prediction Time** (hours)	**Outcome Time** (months)	**Performance** (AUC)	**Sensitivity / Specificity**
[120]	102	NSE	2	6	0.51	6/90
[120]	89	S-100B	2	6	0.67	3/90
[121]	123	NSE	24	0	0.70	26/90
[122]	62	BIS	24	0	0.78	63/86
[25]	50	PLR	24	3	0.79	67/91
[123]	60	BS Ratio	-	0	0.80	73/91
[105]	56	EEG Category	12, 24	6	-	40/100
[124]	90	Age, NSE, OHCA	24,48	6	-	44/100
[125]	72	GCS	24	0	0.81	61/100
[126]	56	NSE	72	6	0.81	70/100
[127]	18	Electrodermal Activity	48	0	0.82	-
[128]	35	EEG Category	72	0	0.84	-
[129]	35	NSE	24	0	0.85	30/100
[130]	109	Brainstem Reflexes Myoclonus, SSEP EEG Reactivity	72	3-6	0.86	62/100
[119]	79	EEG: Relative Delta Power, Cross correlation, Mutual Information Transfer Entropy	48	0	0.88	-
[131]	340	Age, Gender SAPS II Score Rhythm Laboratory Exams	0	-	0.88	97/32
[132]	230	NSE, CPR, ROSC, MTH	72	0	0.88-0.94	80-86/89-90
[133]	134	Rhythm, ROSC, Clinical Exams, EEG Reactivity, NSE, SSEP,		3	0.89	-/100
[134]	277	EEG Category Motor response Age	72	6	0.90	-
[135]	83	BIS	5	0	0.91	86/94
[136]*	60	Clinical Exams SEP, EEG Category	12, 24, 48, 72	12	-	82/100
[137]	130	NSE	72	6	0.92	85/100
[138]	177	Age, Gender, NSE, CPR, ROSC, TH, Location, Rhythm	96	6	0.93	63/100
[1]	109	EEG: Alpha/Delta Ratio Coherence in Delta Regularity Shanon's Entropy Standard Deviation	18	6	0.94	64/100
[2]	283	EEG: Alpha/Delta Ratio Coherence in Delta Regularity Shanon's Entropy Standard Deviation	12, 24	6	0.90,0.94	56,65/100,94
[19]	75	S-100B, BIS, NSE	48	6	0.95	65/100

Table 4.1: **Review of Prognostic Literature.** NSE: Neuron Specific Endolase, BSE: Bispectral Index, GCS: Glasgow Coma Score, PLR: Pupilary Light Response, BS: Burst Suppression, EEG: Electroencephalogram, S100B: S100 calcium-binding protein B, SSEP: Somatosensory Evoked Cerebral Potentials, OHCA: Out-of-hospital Cardiac Arrest, SAPS: Simplified Acute Physiology Score, CPR: Cardiopulmonary Resuscitation, ROSC: Return of Spontaneous Circulation, MTH: Mild Therapeutic Hypothermia, TH: Mild Therapeutic Hypothermia. *Outcome was GOS, not CPC.

"The plural of anecdote is **not** data."

–A role model

Chapter 5

Data

Synopsis

We collected a multi-center dataset containing 950 adult patients diagnosed with cardiac arrest. 165 patients were excluded due to corrupted EEG data or missing outcomes. After exclusion, the data contained 785 adult subjects diagnosed with in-hospital or out-of-hospital cardiac arrest from October 2009 to April 2016 in five university-affiliated hospitals. Collectively, the 784 subjects provided over 35,000 hours of continuously recorded EEG data. The primary outcome was defined as the best neurological function achieved up to 6-months after initial cardiac arrest, according to the Glasgow-Pittsburgh Cerebral Performance categories (CPC) scale, based on electronic medical record review. Patient outcomes were distributed as follows: CPC 1, 249 patients; CPC 2, 51 patients; CPC 3 patients, 22 patients; CPC 4, 15 patients; CPC 5, 448 patients. When dichotomized, 300 subjects has a 'good' functional outcome, and 485 patients had a 'bad' functional outcome.

5.1 The Challenges of the Data

Building a multi-center database of EEGs, relevant covariates, and functional outcomes was a non-trivial engineering effort, but an important part of this thesis. The format of the collected EEG data was highly heterogeneous both within, and across institutions. A given subject's EEG data was typically represented by a set of one or more European Data Format (EDF) files, but the EEG channel names and sampling rates were different from file to file. Furthermore, the start- and end- times of the files, as indicated by their headers, was not consistently reliable, and some signals were inverted. The untidy nature of the data required us to generate software procedures which (A) mapped and aligned various channel terminologies (B) re-sampled and merged the disparate EDF files so that each subject had a single EEG data file, sampled at 100Hz, containing up-to 72 hours of data, (C) map EDF file names to patient medical record numbers at their corresponding institutions and (D) map patient medical record numbers to pertinent covariates and outcomes.

5.2 Patient Characteristics

At the time of this thesis, the multi-center data contained 950 adult patients diagnosed with cardiac arrest. 165 patients were excluded due to either corrupted data from frontal EEG electrodes, or missing outcomes. After exclusion, the data contained 785 adult subjects diagnosed with in-hospital or out-of-hospital cardiac arrest from October 2009 to April 2016 in five university-affiliated hospitals. In Table 5.2, below, we provide an overview of the EEG data procured for all analysis in the thesis.

Institution	**Collection** dates	**Files** (.EDF)	**Size** (TB)	**Subjects** total	**Subjects** post-exclusion
Massachusetts General Hospital, Boston, MA	12/11 - 04/16	1,528	0.834	178	156
Brigham and Womens Hospital, Boston, MA	07/12 - 11/15	1,365	0.723	154	108
Universtiy of Twente, Netherlands	05/10 - 06/15	2,995	4.5	385	351
Yale University Hospital, Newhaven, CT	08/13 - 08/15	883	0.858	123	97
Beth Israel Deaconess, Boston, MA	10/09 - 10/15	1263	0.380	110	73
Totals	10/09 - 04/16	8,034	7.29	950	784

Table 5.1: **An Overview of Data Collected**. TB: Terrabytes. EDF: European Data Format

All subjects were comatose after return of spontaneous circulation and received targeted temperature management (TTM) with a goal temperature of 32-34°C. The TTM protocol

prescribed hypothermia at 32-34°C for 24 hours via external cooling pads. The hypothermia phase was followed by gradual rewarming to 37°C, at a rate of 0.25-0.5°C per hour. At the Massachusetts General Hospital, neuromuscular blockade was maintained throughout hypothermia and rewarming phases. Neuromuscular blockade was used as needed to manage shivering at other participating institutions. Analgesia and sedation were provided with midazolam (0.1 mg/kg/h), propofol (25–80 mcg/kg/h), or fentanyl (25–200 mcg/h) infusions, titrated at the discretion of the treating physicians. Data collection was performed under independent Institutional Review Board (IRB) approvals overseeing each institution. Patients received continuous EEG monitoring, which was initiated as early as possible during TTM, and maintained for 24-72 hours, unless the subject regained consciousness, had life-sustaining therapies withdrawn, or died. Recordings were collected from twenty-one electrodes, positioned on the scalp in accordance with the international 10–20 system.

5.3 Outcome Definition

The primary outcome was defined as the best neurological function achieved up to 6-months after initial cardiac arrest, according to the Glasgow-Pittsburgh Cerebral Performance categories (CPC) scale, based on electronic medical record review. When available, outcomes were also collected at patient discharge from the ICU, and 3 months following discharge. CPC values which were not specified by the care providers were estimated by reading through available patient discharge summaries, and related care notes. Patient outcomes were distributed as follows: CPC 1, 249 patients; CPC 2, 51 patients; CPC 3 patients, 22 patients; CPC 4, 15 patients; CPC 5, 448 patients. CPC score were dichotomized prior to analysis into 'good' and 'bad' functional outcome groups where 'good' functional outcome was defined as a CPC score of 1 or 2 and 'bad' functional outcome was defined as a CPC of 3 to 5.

5.4 Clinical Covariate Collection

When available, six clinical descriptors of the patient population were also collected: age, rhythm at arrest, time until return of spontaneous circulation (ROSC), sex, location of arrest (in hospital or out of hospital), and cause of arrest (cardiac, pulmonary, anesthesia or

neurological). For the purpose of all analyses, rhythm at arrest was dichotomized as shockable (ventricular fibrillation or ventricular tachycardia) or non-shockable (asystole, pulseless electrical activity, and unknown).

In Table 5.2 we show the clinical characteristics of the population, partitioned by their CPC Score. The average age of patients was 60 years, and 68% of the patients were male. Clinical covariates (other than age and gender) were missing for some patients: 46 (5.8%) patients were missing an arrest rhythm, 375 (48%) were missing the time to return of spontaneous circulation (ROSC), 316 (40%) were missing the cause of arrest, and 249 (32%) were missing the location of arrest. The average time to ROSC (when known) was 23 minutes. Of those subjects not missing clinical covariates, 50% had a shockable arrest rhythm, 40% had a pulmonary origin of arrest, and 58% had an arrest outside the hospital. The distributions of the collected clinical covariates, partitioned by outcome, are shown in Figure 5.1.

	CPC 1 (n=249)	**CPC 2** (n=51)	**CPC 3** (n=22)	**CPC 4** (n=15)	**CPC 5** (n=448)
Age (Years)	57(15)	56(12)	69(10)	50(20)	63(16)
Gender (% Male)	68	75	59	47	68
ROSC (Mins)	19.47(17)	13.7(11)	10.8(7)	17(5)	25.7(21)
Rhythm at Arrest (%)					
Shockable	69	68	41	33	34
Unshockable	25	14	35	67	61
Other/Unknown	6	18	24	0	5
Cause of Arrest (%)					
Pulmonary	50	47	36	20	34
Anesthesia	3.6	3.9	4.6	6.7	8.5
Neurologic	9.6	15.7	13.6	0	13.8
Other/Unknown (%)	36.8	33.4	45.8	73.3	43.7
Arrest Location					
In Hospital (%)	9	12	23	7	11
Out of Hospital (%)	63	67	41	40	56
Unknown (%)	28	21	36	53	33

Table 5.2: **Clinical Characteristics.** Mean and Standard Deviation of demographic features, partitioned by patient CPC at 6 months after ICU discharge. Note that for predictive modeling, CPC 1 and 2 were considered 'good' outcomes while CPC 3-5 were considered 'bad' outcomes.

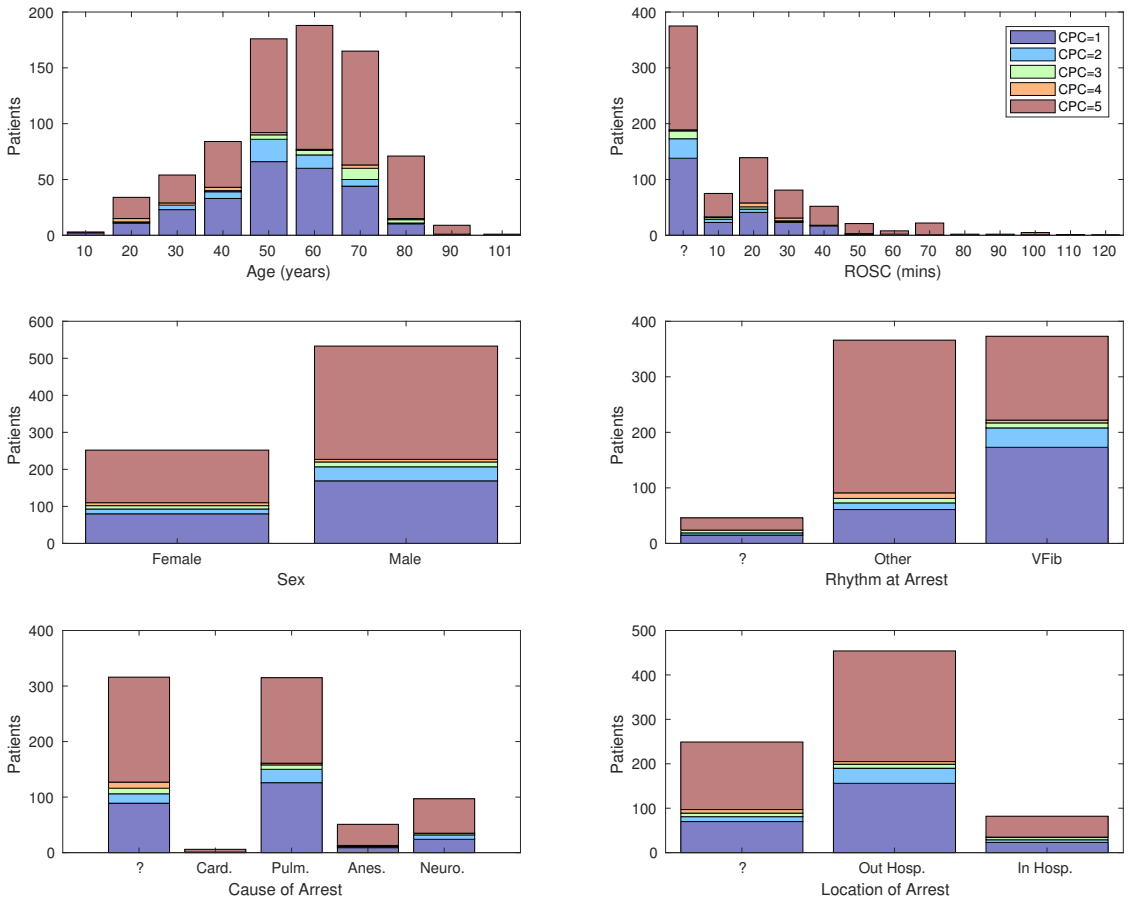

Figure 5.1: **Clinical Covariate Distributions.** Distributions of our clinical covariates, partitioned by subject CPC score. Bars denoted with the '?' symbol indicate missing data.

5.5 EEG Data Collection

Collectively, the 785 subjects provided over 35,000 hours of continuously recorded EEG data. In Figure 5.2-A, we show the total number of subjects with EEG data continuously recorded since the time of EEG initiation; All 785 patients had at least one hour of EEG recorded (leftmost point), but only 174 patients had 72 hours of EEG recored (rightmost point). In Figure 5.2-B, we show a histogram of EEG withdrawal characteristics; the decision to withdraw EEG was periodic, peaking every approximately 24 hours.

In Figure 5.2-C, we represent the characteristics of our EEG data *after* aligning waveforms with respect to the estimated time of cardiac arrest. Each bar represents the data in a time-

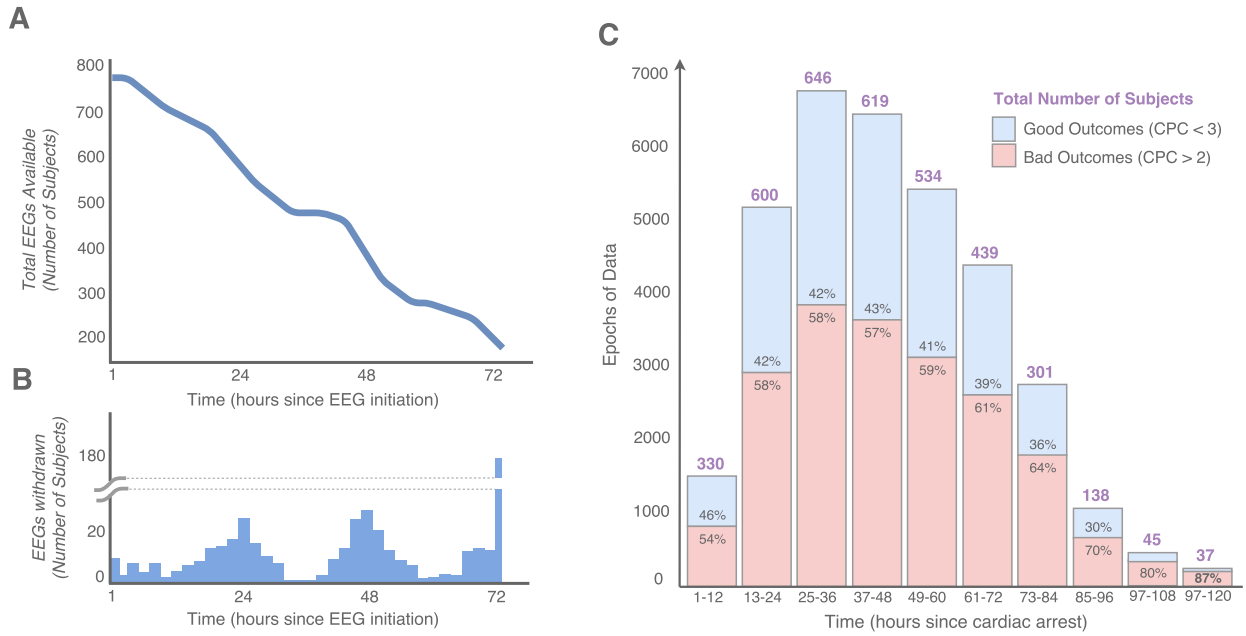

Figure 5.2: **EEG Data Characteristics.** (A) The total number of subjects with EEG data recorded with respect to the time of EEG initiation. (B) A histogram of EEG withdraw decisions. (C) The characteristics of the EEG data after alignment with respect to the time of cardiac arrest. Each bar represents the number of 5-min EEG epochs available for a 12 hour time-interval. Bars are partitioned by the number of 'Good' (blue) and 'bad' (red) outcomes. At the top of each bar, we show the total of unique subjects that contributed the epochs within each time-interval.

interval (x-axis) since cardiac arrest. The height of each bar represents the total number of 5-minute EEG epochs (y-axis) available in the time-interval. On top of each bar, we designate the total number of unique subjects that contributed data epochs to the time-interval (e.g. 330 patients contributed epochs in hour 1-12, 600 patients contributed data in hour 13-24, and so on). Bars are also annotated with the total proportion of 'good' (bottom of blue bars) and 'bad' (top of red bars) outcomes they contain; less than half of the patients (42%) had EEG recordings within 12 hours of arrest, 82% of patients had EEG data monitoring 24 hours after the arrest and only 6% had EEG monitoring 96 hours after arrest. Additionally, 46% of patients in coma between 1-12 hours had a 'Good' outcome compared to only 20% of those in coma for 97-108 hours.

5.6 EEG Preprocessing Approach

In Figure 5.3 we provides an illustrative description of our pre-processing approach. Preprocessing proceeded with the following steps:

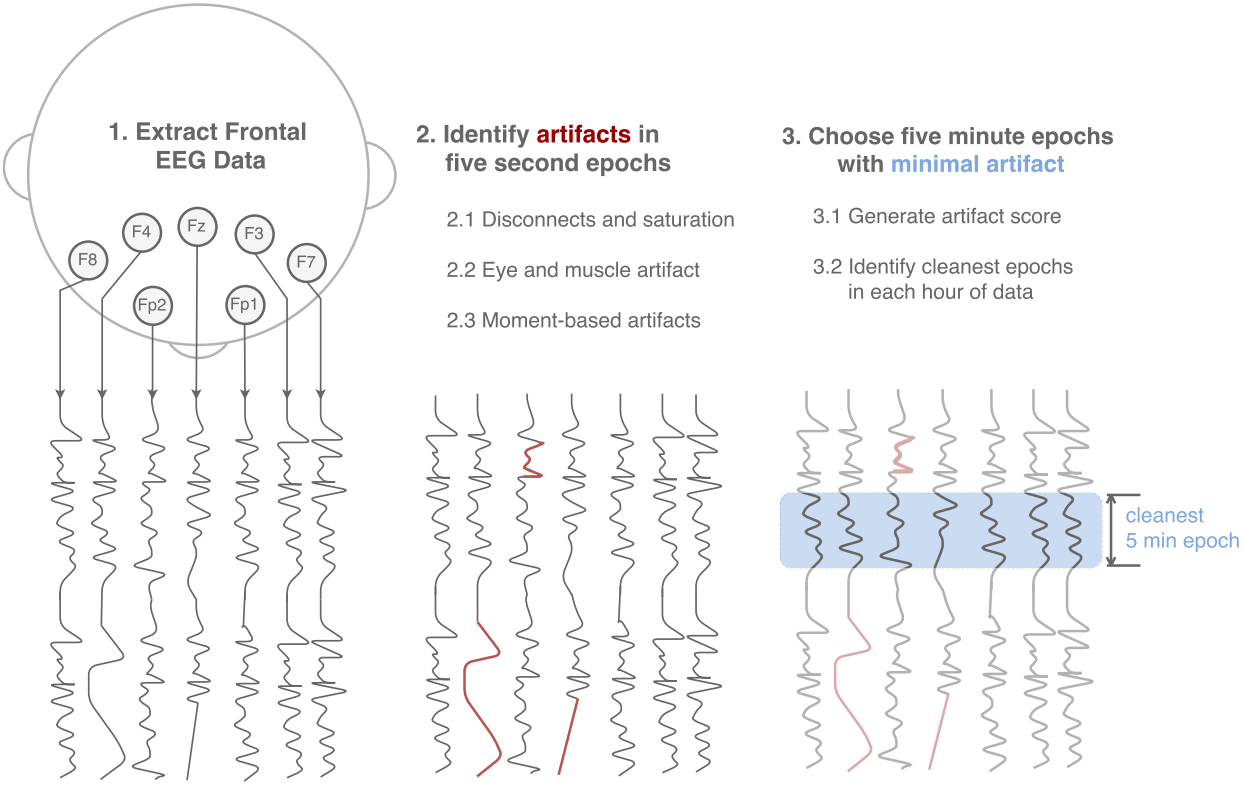

Figure 5.3: **Data Pre-processing Approach**. We extract and standardize data from frontal EEG electrodes. Following extraction, we utilize an unsupervised algorithm to annotate the presence of artifact in 5-second epochs. Finally, for each hour of available data, we select the five-minute epochs of EEG with minimal artifact. The selected epochs are used for feature extraction.

5.6.1 Data Standardization

The raw EEG data we collected were heterogeneous in terms of channel naming conventions, sampling rates, and the continuity of the recorded waveforms. To standardize the data, we re-named all channels to reflect the 10-20 naming convention, filtered (0.5-50Hz) and re-sampled all signals to 100Hz, and merged disjoint EEG snapshots into single files. Patient-level waveforms were then curtailed to contain no more than 72 hours of continuous data with respect to the time of EEG initiation.

5.6.2 Basic Cleaning

Following standardization of data, the EEG waveforms were processed to remove corrupt channels using the approach developed by Bigdely *et al.* [141]. After removing corrupt channels, EEGs were re-referenced using the average montage, the baseline was removed and the data were epoched into 5-second contiguous intervals.

5.6.3 Artifact Detection

For each channel, we identified any epochs that contained channel disconnections, saturations, eye artifacts, muscle artifacts, or statistical abnormalities. We define each of these conditions more specifically below:

- **Saturation**: voltage was above 1000 μV
- **Disconnect**: variance was less than 0.001 μV
- **Eye Artifact**: +/- 50 db from the population mean in 0-2Hz band [142]
- **Muscle Artifact**: +25 or -100db from the mean in the 20-40Hz band [142]
- **Statistical Moment Abnormalities**: signal variance, skew or kurtosis were three standard deviations or more from the population mean

The computer code (Matlab, 2017b) used to pre-process the EEG waveforms is freely available in an on-line repository [4].

[4] https://github.com/deskool/ComaPrognosticanUsingEEG

"To bin is a sin."

–A Professor

Chapter 6

Study 1: An Elastic Feature Approach

Synopsis

Electroencephelogrpahy (EEG) predicts neurological recovery following cardiac arrest, but recent work has shown that prognostic implications of some key EEG features (e.g. voltage and continuity) are dynamic. We explore whether accounting for this time-dependence can allow for better prognostic predictions. Using data from four major academic medical centers in the United States, we analyzed over 12,000 hours of continuous EEG data, extracting over 50 features of complexity, category, and connectivity from the continuous EEG recordings of 438 postanoxic coma patients. We modeled associations between EEG features and dichotomized neurological outcome in 12-hour intervals using sequential Elastic-Net. We compared a predictive model utilizing the time-varying features to a model using time-invariant features and two approaches from the literature. Performance was quantified using the area under the receiver operator curve (AUC) and statistical calibration. We found that a model utilizing time-dependent features outperformed one trained with time-invariant features (overall AUC=0.83 versus AUC=0.79) and the best performing model in the literature (AUC=0.83 versus 0.68). The time-sensitive model was also the best-calibrated of the tested approaches. The statistical association between quantitative EEG features and neurological outcome in postanoxic coma changes over time, and accounting for these changes improved prognostication.

6.1 Introduction

Specific electroencephalogram (EEG) patterns are associated with eventual recovery from coma due to hypoxic-ischemic encephalopathy (HIE) after cardiac arrest. However, existing EEG review practices rely on subjective visual analysis, which does not readily translate into reproducible quantitative predictions of neurologic outcome. Quantitative EEG (qEEG) methods are capable of detecting early signs of neurological recovery while coma persists. Hence, qEEG holds promise for reliably identifying patients likely to benefit from continued intensive supportive care. In this study, we utilize machine-learning methods to develop a model that predicts good long-term neurologic outcome by integrating clinical data with several existing and novel qEEG features. Our work in this chapter goes beyond existing results by using a more comprehensive set of qEEG features, leveraging temporal trends (both locally, and globally) and rigorously characterizing the performance of the model for both decision support, and severity-of-illness scoring.

6.2 Methods

6.2.1 Data and Outcome

Data for this study included 12,397 hours of continuous EEG data of 438 PAC patients collected from the Massachusetts General Hospital (MGH, n=147), the Brigham and Women's Hospital (BWH, n=105), Yale School of Medicine (YSM, n=107) and the Beth Israel Deaconess Medical Center (BIDMC, n=79) between 2010 to 2016. Our outcome of interest was the lowest recorded Cerebral Performance Category (CPC) of the patients measured between discharge and six months after discharge, dichotomized such that CPC < 3 was coded as 1, and CPC $>= 3$ was coded as 0. Good outcome status (CPC of 1-2) or death (CPC 5) was achieved by the time of hospital discharge for 91.6% (401/438) patients. The remaining 8.5% (37/438) had a CPC of 3 or 4 at discharge. Among these, 14 ultimately improved to a best CPC score by 6-months of 1 or 2, five remained with CPC of 3-4, and 10 died. For the remaining 8 cases (1.8% of 438), 6-month outcomes could not be determined by chart review. For these the discharge CPC score was carried forward and taken as the

	'Good' Outcome CPC 1-2	'Bad' Outcome CPC 3-5
Sample size	134	304
Age in years (Mean)	54.2	62
Time to ROSC in minutes (Mean)	19	27
Shockable rhythm (n [%])	58 [43.3%]	53 [17.4%]
Male (n [%])	82 [61.2%]	206 [67.8%]

Table 6.1: **Demographic Features of Population.** Selected clinical features for model development partitioned by functional outcome. ROSC: return of spontaneous circulation, PEA: pulseless electrical activity, Shockable rhythm: ventricular fibrillation or ventricular tachycardia. The precise time between arrest and initiation of EEG (not shown here) was unknown in 45.4% of the cases.

final neurologic outcome.

In Table 6.1 we provide the mean and standard deviation of the collected clinical features for the subjects, partitioned by their CPC outcome class. The mean age was 59.6 years and 30.6% of subjects had good outcomes (CPC 1-2) at 6-months. A Pearson's linear correlation between the rate of poor outcome, and time was statistically insignificant ($p = 0.76$). All clinical covariates missing in less than 25% of patients were selected for use in our model. These features included patient age, sex (with male coded as 1), and an *estimated* time between cardiac arrest and return of spontaneous circulation (ROSC).

6.2.2 EEG Features

EEG Data was pre-processed to identify artifacts using the approach described in Chapter 6. Following artifact detection, for each hour of available EEG data, the 5-minute EEG epoch with least artifact present was retained for feature extraction. We extracted single- and multi- channel features for use in our model. Single channel features were computed on the following frontal electrodes: Fp1, Fp2, F7, F3, Fz, F4, and F8. Multi-channel features were computed on all unique pairings of the selected frontal electrodes [5]. The extracted single channel features measured the complexity and category of EEG activity while multi-channel features measured cross-channel connectivity. We extracted a total of 57 EEG features, which are described in Chapter 4 of this thesis.

[5]Channel pairs: F3-F4, Fp2-Fz, Fp1-F7, F8-Fz, F7-Fz, F7-F8, F4-Fz, F4-F7, F3-Fz, Fp2-F8, Fp2-F7, Fp2-F3, Fp1-Fz, Fp1-F8, Fp1-F4, Fp1-F3, Fp1-Fp2, F4-F8, F3-F8, F3-F7, Fp2-F4

6.2.3 Post-Processing of EEG

Following feature extraction, the resulting data tensor was in $\mathbb{R}^{s\times c\times t\times f}$, where s denotes the number of subjects, c denotes the number of channels (and channel pairs), t denotes time (in hours) and f denotes the number of features. To reduce the dimensionality of our feature tensor, we utilized the mean value of the extracted features across all channels resulting in a data tensor in $\mathbb{R}^{s\times t\times f}$. This simplification of the data is advantageous given that the effects of hypoxic ischemic encephalopathy (HIE) are non-localized (affecting the entire brain) so the average value of the features across channels is more representative of injury than a measure from any single channel.

Next we reshape the tensor, collapsing the data across time and resulting in a $\mathbb{R}^{st\times f}$ matrix. We then removed data points with missing clinical features or outcomes, m_{st}, resulting in a data matrix in $\mathbb{R}^{(st-m_{st})\times f}$. This matrix was then z-scored with respect to the features and transformed back into a 3-dimensional tensor in $\mathbb{R}^{s'\times t'\times f}$ where s' and t' are the updated number of subjects and times after removing missing data.

Following preprocessing, The data was partitioned into a set of k contiguous tensors: $B_0, B_1, \ldots B_k$, where each element B_i is in $\mathbb{R}^{s\times h\times f}$, and h denotes the number of hours of data in each element. In our case f=56 features, h=12 hours, and s=441 subjects. B_0 contains the first 12 hours of our subject's 57 features, B_1 contains hours 13-24 of our subject's EEG features and so on. Next, $B_{0:k}$ is used to construct the dataset used for our analysis: $C_0, C_1, \ldots C_k$, where each element C_i is a temporal partition of our data, composed of $B_{0:i}$, collapsed across time. That is, C_i is in $R^{sh\times f\dot{(}i+1)}$. C_0 contains features from the first 12 hours of EEG monitoring, C_1 contains features from the first 24 hours of EEG monitoring, and so on.

6.2.4 Modeling Approach: Logistic Regression with 'Elastic' Memories

EEG activity is known to change as a function of temporal features (e.g. time since arrest) and state-related features (e.g. trajectory towards recovery). For this reason, the prognostic importance of EEG features may become more or less important over time. For example, isoelectric EEG early may not be as predictive of poor outcome shortly after arrest as it is

three days after arrest. We accounted for this evolution in the feature-outcome relationship by training a contiguous sequence of logistic regression models, one for every 12 hours of EEG data. All subjects with EEG that was discontinued before a given time interval were excluded for the model trained within that interval, thus each model was trained to predict 6-month neurologic outcome for patients still surviving up to a given epoch. Each model in the sequence included the feature information from the current time interval, as well as feature information from any preceding time intervals (e.g. the model at 36 hours contains feature information from 1-12 hours, 13-24 hours and 25-36 hours). This approach allowed models later in the sequence to consider both past and present feature information when making predictions.

One clear challenge of our sequential modeling approach is that the feature/data ratio quickly becomes statistically prohibitive as additional features are added, especially for models later in the sequence. To address this challenge, each model in the sequence is regularized using Elastic Net, a technique that forces models to discard features that are inessential for the prediction task (Zou and Hastie, 2005). In Fig 6.1, we graphically illustrate our modeling approach. Henceforth, we refer to the Elastic Net regularized sequential logistic regression model as the time-sensitive model.

6.2.5 Baseline Models

We compared the time-sensitive model to four baseline approaches: (A) A clinical baseline: a logistic regression using age, gender, ROSC and initial rhythm; (B) a time-insensitive model: a logistic regression model using features selected by Elastic Net across all time intervals; (C) the 2013 Cerebral Recovery Index (CRI): a heuristic approach using alpha/delta bandpower, standard deviation, coherence in delta band, Shannon's entropy, and regularity with parameter values selected by the investigators and (D) a Random Forest: a Random forest classifier with features and settings inspired by the 2017 CRI [2]. The CRI methods were selected as literature baselines given their impressive reported performance. The 2017 CRI used the five features from the 2013 study, in addition to four features that characterized burst-suppression activity. The authors reported that these additional features provided modest improvements in model performance, compared to a Random Forest classifier using

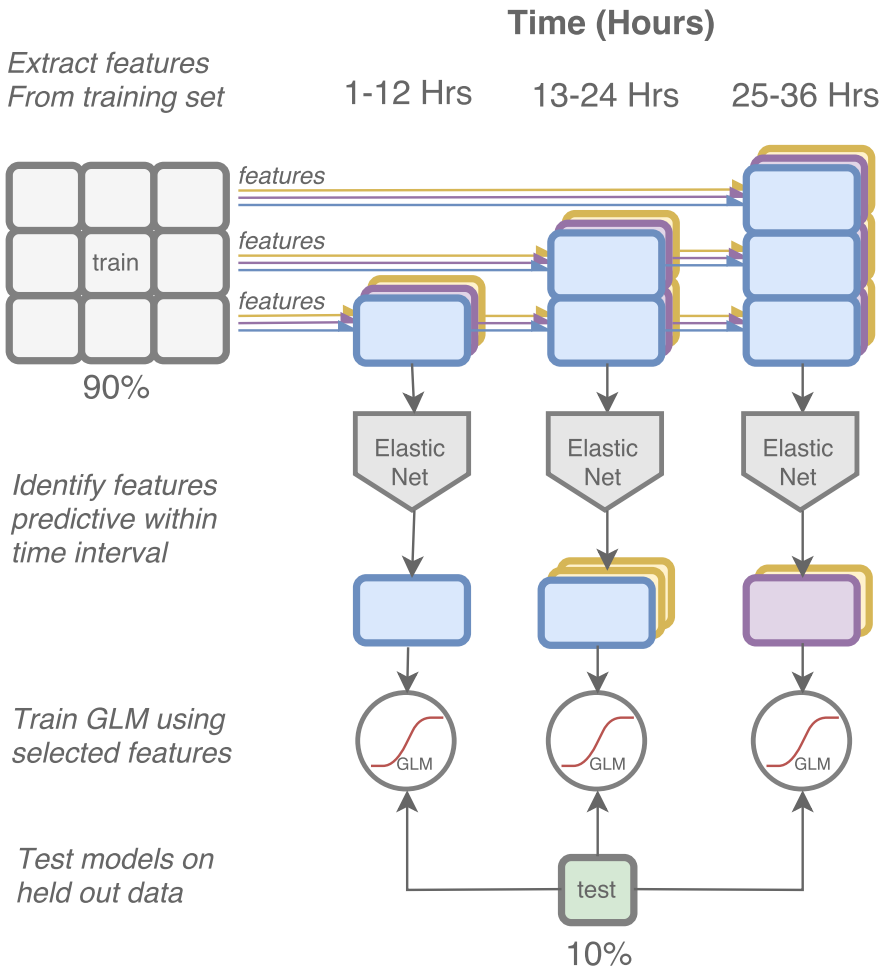

Figure 6.1: **Modeling Approach**. Features (depicted as colored rectangles) are extracted from contiguous temporal partitions of our training data. All features within, and preceding a given temporal partition are-provided to an Elastic Net, which identifies the subset of features within the temporal partition most predictive of patient outcomes. The selected features, and training data, are then used to generate a final logit-linked generalized linear model - the performance of which is evaluated on a held out test set (green box).

the original five features. For this reason, we used the five CRI features from the 2013 model when implementing the Random Forest classifier.

6.3 Performance Characterization Metrics

We evaluated model performance using EEG data available for each consecutive 12-hour interval, up to 72 hours. Performance metrics included Area Under the Receiver Operator Curve (AUC), used to evaluate the model's ability to make binary predictions, and the

sensitivity and specificity for specific operating points of interest on the ROC curve; and statistical calibration, which measures how well the observed proportions of good outcomes match predicted probabilities. Note that, because our modeling target is to predict "good" neurologic outcomes, for sensitivity statistics we computer the proportion of subjects whom the model predicts will have a good outcome who do in fact go on to have a good outcome.

Calibration is particularly important for our model when used as a risk-score. Unlike classification metrics (AUC and accuracy), calibration measures the reliability of the model's probabilistic outputs (e.g. when the model says there is 20% chance of a positive outcome, do 20% of patients actually wake up?). A well-calibrated model is particularly important for clinical decision making, where different clinical experts and/or family members may have different thresholds of acceptable misclassification risk.

To provide a single summary statistic for each model's performance that was independent of time, we also defined a single "overall" AUC for each model. To obtain the overall AUC, we used all available data for each subject, and computed a single probability at the time EEG monitoring for that patient ended. These probabilities thus represent the best prediction that could have been made for a given patient with a given model. For example, for a patient with 24 hours of EEG data, a probability of good outcome would be computed at the 24th hour, while a patient with 72 hours of EEG data would have a prediction computed at 72 hours. These probabilities were then used to construct an ROC curve, and we reported the area under it, as well as other metrics.

6.3.1 Model Validation Approach

All models were validated using 10-fold cross validation. That is, we partitioned available data into ten folds. In each fold, 90% of the data was used to identify parameters for models and the remaining 10% of the data were used to evaluate model performance on subjects that were never seen in the corresponding training sets. The subjects within each testing fold were unique. We used the average performance of models across the unseen testing sets in the ten folds when comparing performance. In Figure 6.2, we graphically illustrate the validation approach to aid in comprehension. All analyses were performed using MATLAB 2016a.

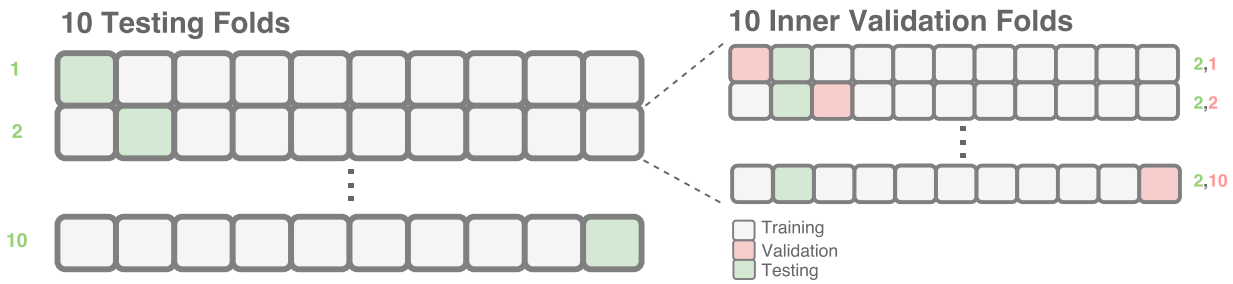

Figure 6.2: **Cross-validation Approach.** Nested cross-validation approach used to identify hyper-parameters within each training fold.

6.3.2 Elastic net hyper-parameters

Elastic Net requires investigators to set two hyper-parameters, and , that influence how "strict" the approach will be when selecting features. We determined the optimal value of these hyper-parameters empirically using the following approach: we partitioned the training set in each fold into 10 inner validation folds (see Figure 6.2). In each inner validation fold, a set of logistic regression models was fitted using various Elastic-Net hyper-parameter settings, and the deviance (a goodness-of-fit measure) was calculated. The hyper-parameter settings that produced the minimum average deviance across all 10 inner validation folds were selected as optimal. A final logistic regression model was fit to all of the training data in each fold, using the optimal Elastic Net hyper-parameter values.

6.4 Results

6.4.1 A Simple Univariate Analysis

Here we present the results of a simple, exploratory analysis of the data. In Table 6.2, we illustrate the statistically significant ($p < 0.05$) univariate Spearman correlations between our extracted feature values and raw CPC scores at various time intervals following EEG initiation. A negative correlation indicates that the feature was associated with better outcomes. We observe here that several features exhibited univariate associations with the outcome, irrespective of time (top half of Table 6.2). For instance, Shanon's entropy was consistently associated with better outcomes, while EEG Voltage beneath 5 μV, and Phase Lag Index were consistently associated with worse outcomes. Other features were associated with pa-

Time Interval (Hours)	**1-12**	**13-24**	**25-36**	**37-48**	**49-60**	**61-72**
Time-Insensitive Correlations						
Complexity Features						
Shanon Entropy	-0.09	-0.18	-0.11	-0.15	-0.31	-0.26
Ceptrum Coefficient (2)	-0.10	-0.14	-0.20	-0.14	-0.16	-0.23
Hjorth Mobility	0.09	0.09	0.14	0.09	0.21	0.15
False Nearest Neighbor	0.18	0.11	0.19	0.22	0.27	0.16
Category Features						
Low Voltage EEG (<5 uV)	0.14	0.20	0.14	0.18	0.34	0.30
Number of Epileptiform Spikes	0.21	0.23	0.26	0.24	0.26	0.20
Connectivity Features						
Coherence in delta band	0.21	0.21	0.21	0.29	0.18	0.19
Phase Lag Index	0.20	0.27	0.25	0.24	0.20	0.17
Time-Sensitive Correlations						
Complexity Features						
ARMA Coefficient (1)	-0.13	-0.15	-0.15	-	-	-
Lyapunov Exponent	-	-	-	-	0.13	0.20
Category Features						
Regularity	-0.15	-0.24	-0.19	-0.15	-	-
Mean Supression Length	-	0.19	0.10	0.16	0.26	0.21
Low Voltage EEG (<10 uV)	-	-	-	0.09	0.27	0.26
α/δ Band Ratio	-	-	-	-	0.13	0.14
Normal EEG (%)	-	-	-	-	0.16	0.16
Diffuse Slowing (%)	-	-	-	-	-0.16	-0.16
Connectivity Features						
Cross correlation Magnitude	-	-	0.17	0.28	0.30	0.23

Table 6.2: **qEEG-Outcome Correlation**. The univariate Spearman correlation between our features, and discrete CPC values at a given interval. Here, a negative correlation coefficient implies an association between the features and a good outcome. We only display features which which exhibited statistically significant correlations ($p < 0.001$). Features are organized by the type of activity they capture (complexity, category and connectivity), as well as their temporal sensitivity (top half of the table shows time-insensitive features).

tient CPC at particular points in time (Figure 6.3, right), while other features consistently discriminate outcome (Figure 6.3, left). Signal Regularity (a measure of burst-supression described by [1]), was associated with better outcomes in the first 48 hours, while signal cross correlation was associated with negative outcomes after the first 24 hours.

Figure 6.3 accentuates the results from Table 6.2 by comparing three examples of features whose relationship with the most probable outcome changes over time against three features whose relationship with the most probable outcome over time is stable. The Figure

partitions the patient data by outcome class, and illustrate the mean and standard error of three *temporally-sensitive* features and three *temporally-insensitive* features. Here again we observe that certain features distinguish the outcome classes only at particular points in time. This result, in addition to the observations from Table 6.2, provides motivation for a modeling approach that accounts the changing prognostic significance of EEG features when predicting patient outcomes. Importantly, Figure 6.3 provides evidence that alignment of waveforms with respect to the time of EEG initiation does not eliminate the presence of temporally sensitive content in our selected features over time. That is, our alignment strategy did not eliminate the separability of the outcome classes, nor did it eliminate the changes in temporal sensitivity of features.

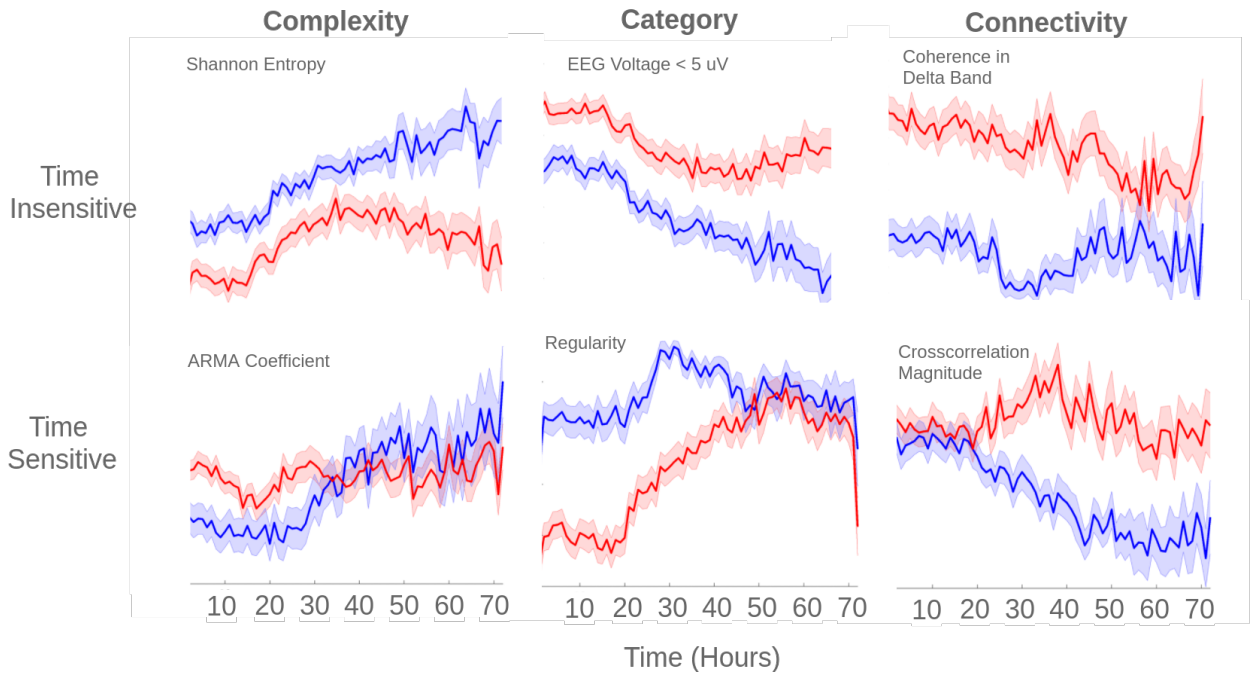

Figure 6.3: **Features Over Time.** An illustration of the temporal evolution in time-sensitive features (three figures on the right) and time-insensitive features (three figures on the left). The positive outcomes group is shown in blue while the negative outcome group is shown in red. The solid line represents the mean of the population while the shaded area represents standard error of the mean. ARMA: Auto regressive moving average.

6.4.2 Performance Characterization

We evaluate the ability of the models in two ways: First, to make accurate binary predictions (good vs. poor outcome), and second, to function as a risk scoring system by predicting the probability of a good outcome. In Figure 6.4, we compare the time-dependent AUC of the time-sensitive model to the baseline approaches within each 12-hour time interval. Note that, the model for each time point is trained only EEG data from patients who have survived up to that time, and the model predictions likewise pertain only to those who remain alive. Here we see that the time-sensitive approach performed best, exhibiting consistent improvements in performance with increased observation time (from AUC of 0.71 at 12 hours to 0.79 at 72 hours).

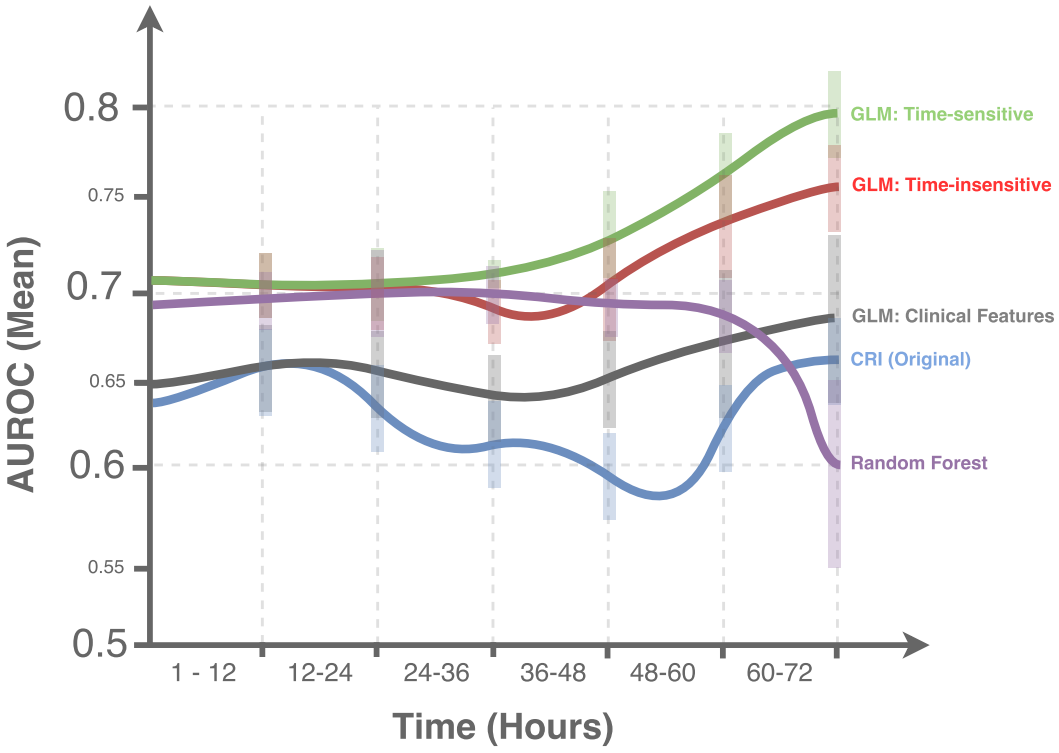

Figure 6.4: **Performance Over Time**. A comparison of our model's ten-fold Area Under the Receiver Operator Curve performance on the held out testing sets over time compared to the baseline approaches. Our model is shown in green, the temporally insensitive baseline using the features from the top half of table 6.2 is show in red, a model using four clinical features (age, gender, ROSC, and arrest rhythm) is shown in gray, a prominent heuristic model from the literature, the CRI [1], is shown in blue and a random forest model based on the more recent work of Cloostermans *et al.* is shown in purple [2].

Additional detail on the performance of the models across the ten folds is shown in Table 6.3. Note that the performance of the clinical baseline model changes slightly across intervals. This is because we are making predictions only for subjects with available EEG data, and the number of such subjects changes in each interval.

With regard to overall area under the receiver operating curve, (see methods), the time-sensitive model made more accurate predictions of 6-month functional outcome (Mean = 0.83 across the 10 testing folds) compared to the time-insensitive model (Mean = 0.79), the Random Forest (Mean = 0.74), the original CRI (Mean = 0.69), and the clinical baseline model (Mean = 0.68). The improvement in model prediction performance was statistically significant according to a t-test ($p < 0.05$). The model's sensitivity was 18%, 42%, and 57% for specificity thresholds of 99%, 95%, and 90% respectively. The model's specificity was 13%, 36%, and 51% for sensitivity thresholds of 99%, 95%, and 90% respectively.

In Figure 6.5, we compare the statistical calibration of the time-sensitive model against the original CRI, Random Forest, and time-insensitive baselines. The time-sensitive model was the best calibrated of the tested approaches. The calibration of all models deteriorated after 48 hours due to decreased number of available training examples (many subjects had termination of EEG monitoring after that time point).

Time Interval (Hours)	**1-12**	**13-24**	**25-36**	**37-48**	**49-60**	**61-72**	**Overall**
Number of Subjects (1/0)	134/304	112/239	88/153	71/111	45/81	21/64	438
Hours of EEG	4272	3178	1987	1487	872	601	12397
Logit GLM	0.71	0.70	0.72	0.72	0.74	0.79	0.83
time-sensitive, EEG + clinical	(0.05)	(0.06)	(0.07)	(0.02)	(0.09)	(0.08)	(0.08)
Logit GLM	0.71	0.70	0.70	0.68	0.73	0.75	0.79
time-insenstive, EEG + clinical	(0.05)	(0.07)	(0.07)	(0.05)	(0.11)	(0.14)	(0.07)
Logit GLM	0.68	0.66	0.67	0.67	0.70	0.65	0.75
time-insensitive, EEG	(0.05)	(0.07)	(0.09)	(0.11)	(0.10)	(0.25)	(0.05)
Logit GLM	0.65	0.66	0.65	0.64	0.66	0.68	0.74
time-insensitive, clinical	(0.07)	(0.08)	(0.07)	(0.10)	(0.20)	(0.26)	(0.13)
Cerebral Recovery Index	0.68	0.69	0.71	0.67	0.67	0.60	0.68
Random Forrest	(0.06)	(0.12)	(0.08)	(0.06)	(0.11)	(0.19)	(0.05)
Cerebral Recovery Index	0.64	0.66	0.62	0.63	0.57	0.66	0.69
Heuristic	(0.07)	(0.08)	(0.08)	(0.05)	(0.06)	(0.09)	(0.07)

Table 6.3: **Results.** A comparison of our model's ten-fold Area Under the Receiver Operator Curve performance μ (σ) on the held out testing sets over time compared to the baseline approaches.

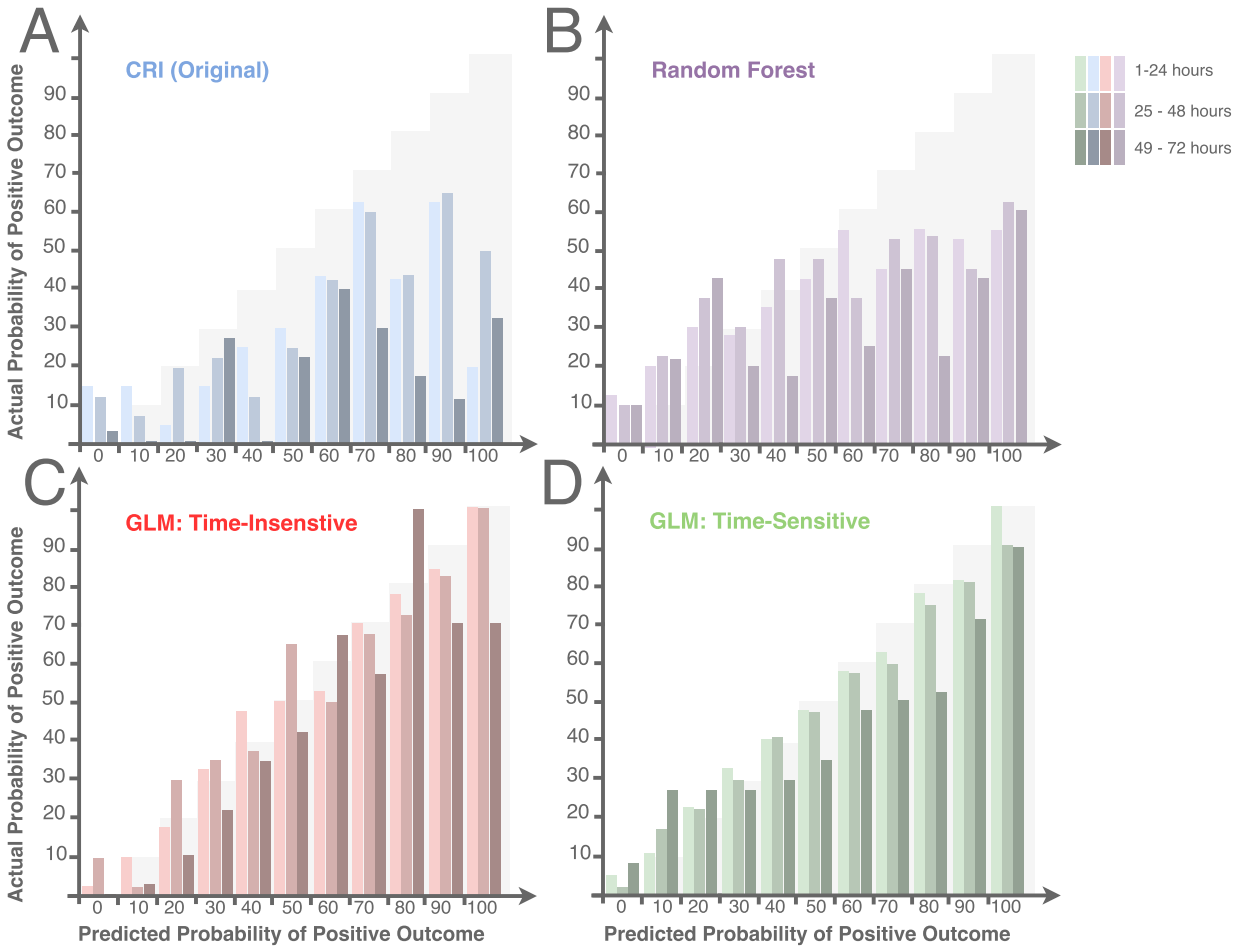

Figure 6.5: **Model Calibration.** A comparison of our model's calibration approach to the two baseline approaches over three days of patient treatment. Ideal calibration is achieved when colors bars perfectly overlap the gray shading in the background of each image. The calibration of a particular prediction level is shown as a bar. The shading of the bar reflects time (lighter is earlier, darker is later). (A) The calibration of a heuristic model proposed by [3], (B) the calibration of a random forest model proposed by [2] (C) the calibration of a GLM model using the temporally-insensitive features from 6.2, (D) The temporally-sensitive GLM model.

6.4.3 Final Model Using All Data

In Table 6.4 we illustrate the features selected by elastic net, and coefficient values of the temporally-sensitive model after training on the entire dataset. We remind the reader that features were z-scored prior to analysis. Positive coefficients indicate a positive association between features and good outcomes. Coefficients with larger magnitudes have stronger predictive effects.

The association of QEEG features with functional outcome varied across time. Not

Feature	Feature Time (Hour)	Coef. 1-12 Hrs	Coef. 13-24 Hrs	Coef. 25-36 Hrs	Coef. 37-48 Hrs	Coef. 49-60 Hrs	Coef. 61-72 Hrs
Intercept		-0.85	-0.71	-0.51	-0.48	-0.89	-1.19
Complexity Features							
Tsalis Entropy (q=2)	1-12	0.1	0.08				
Cepstrum	13-24			0.02		0.13	
	13-24						0.2
Fractal Dimention	1-12					-0.01	-0.09
	13-24						-0.08
Hjorth Complexity	1-12		-0.04	-0.09	-0.03		
	49-60					0.03	0.1
False Nearest Neighbor	1-12	-0.1	-0.12	-0.03			
	13-24					-0.03	
	25-36			-0.06	-0.01	-0.03	-0.03
	61-72						-0.01
ARMA Coefficient	1-12	0.07					
Category Features							
Regularity	1-12	0.16	0.22	0.29	0.25	0.04	
	13-24		0.07	0.29	0.39	0.43	0.31
Number of Epileptiform Spikes	1-12	-0.2	-0.11	-0.14	-0.12	-0.25	
	13-24		-0.2	-0.03	-0.01	-0.02	
	25-36			-0.08	-0.05		
Number of Sharp Waves	37-48						-0.06
Connectivity Features							
Coherence in Delta	1-12	-0.1	-0.01				
	13-24		-0.01				
	25-36				-0.11		
Phase Lag Index	1-12	-0.28	-0.38	-0.31	-0.31	-0.21	-0.09
	13-24		-0.04	-0.21	-0.16	-0.29	-0.4
	25-36				-0.01	-0.18	-0.17
	49-60						-0.02
Cross-Correlation Magnitude	49-60					-0.03	-0.05
Clinical Features							
Gender							0.59

Table 6.4: **Model Coefficients.** A summary of our coefficients from the sequential GLM with LASSO selected features. We show both the mean and 95% confidence interval of the selected features. A positive coefficient indicates association with a good outcome. Features generated earlier in time can be used by models later in time. For example, the model at hour 25-36 had access to all features at time blocks 0-12, 13-24 and 25-36, and choose those that were deemed most informative. The left-most column describes the feature and what time frame it was extracted from (0-12 hours, 13-24 hours, etc.). The other columns of the table represents the value of the coefficients at a particular time-frame (0-12 hours, 13-24 hours etc.)

all types of EEG complexity were positively associated with outcomes. Chaos theoretic metrics including Fractal Dimension and False Nearest Neighbor were associated with poor outcomes while information theoretic measures such as Entropy and Cepstrum coefficients were associated with good outcomes. As expected, the association between outcome and EEG category was dependent on the particular category: epileptiform activity was associated with poor outcomes while regularity (a measure of background continuity) was associated with good outcomes. Increased connectivity between channels was consistently associated with poor outcomes.

Certain features were more predictive in early time intervals (Tsalis entropy, autoregressive-moving average, and coherence in the delta band) while other features were more predictive in later time intervals (number of sharp waves, fractal dimension, and cross correlation magnitude). Phase lag index was the only feature selected across all time intervals. For all features, values at specific time intervals remained useful for predictions in the future. A subject's Hjorth complexity in the first 12 hours of EEG for instance, was useful for outcome predictions at hours 13-48 hours. For some features, the value at multiple time intervals was useful for predictions. For instance, regularity at both 1-12 hours and 13-24 hours were predictive of outcomes at hours 13-60. Most QEEG features were consistently associated with either good or poor outcome prediction, however Hjorth complexity and cross correlation magnitude were associated with prediction of both good and poor outcome depending on the timing of the prediction.

6.5 Discussion

6.5.1 Core Contribution

The technical contribution of this study lies in the methodological approach we developed. Our modeling framework was designed to reflect the decision process of actual care providers, considering information across multiple points in time when predicting outcomes. Importantly, we showed that the qEEG model which retains "memories" of previously encountered features, outperformed existing state-of-the-art approaches which used only the present values. The time-sensitive model had the best classification performance (overall = 0.83) and

statistical calibration of all tested approaches. The test characteristics support the use of our model as one that makes well-calibrated predictions. That is, the observed proportions of good outcomes matched predicted probabilities of good outcomes very well. The superior calibration of our time-sensitive model provides strong motivation for its use as a risk score. In Figure 6.6 we provide short EEG snapshots and spectrograms from the Cz channel of five exemplary patients, each with a different CPC outcomes. On the top-left corner of each image we superimposed the probability of survival according to our temporally sensitive model. Here we see that the temporally-sensitive model is able to correctly scale the probability of survival in accordance with the neurological outcome of the patients.

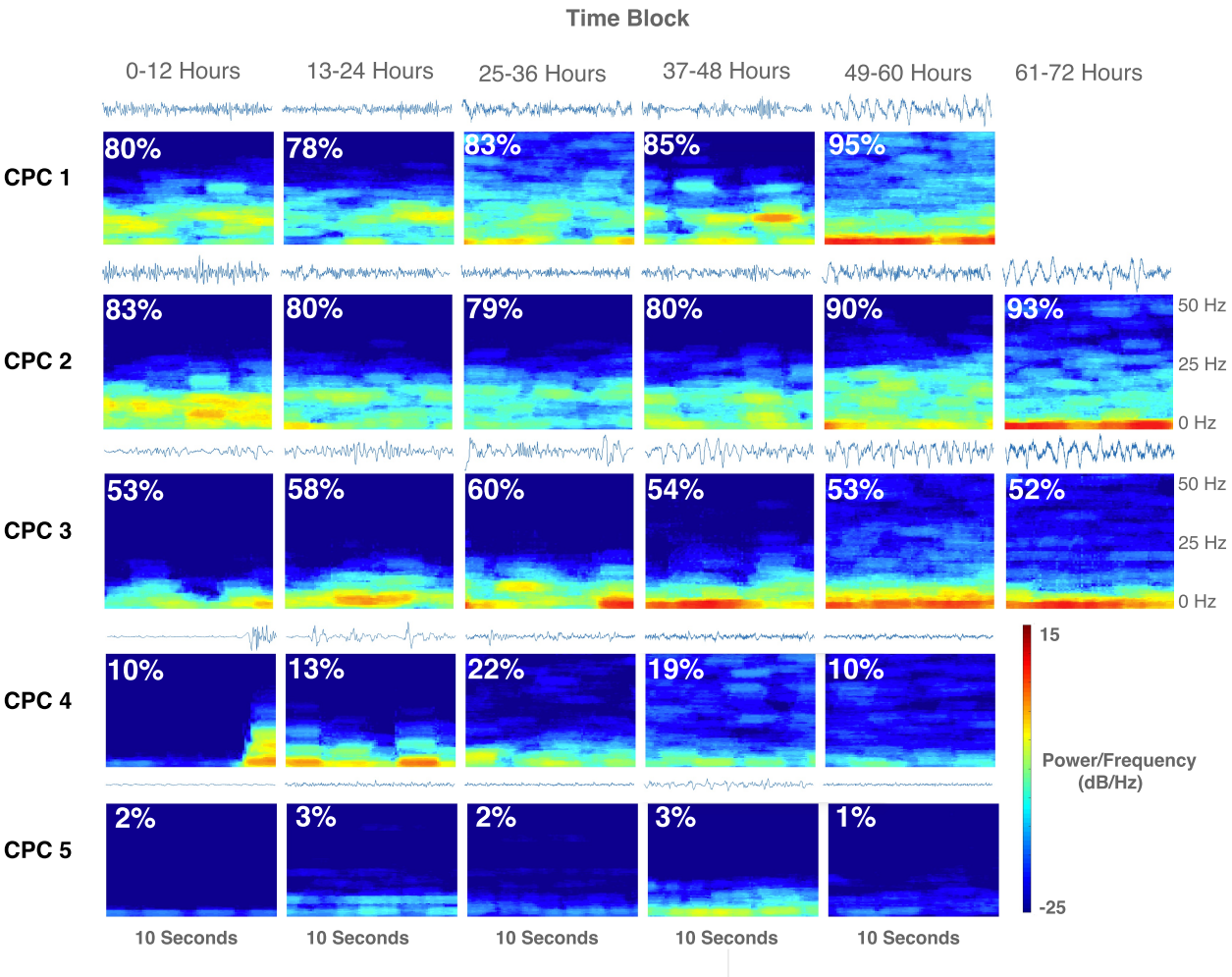

Figure 6.6: **Performance on Exemplary Patients.** Each row illustrates a sequence of multi-taper spectrograms and EEG time-series generated from 10 seconds of Cz electrode activity, across multiple time blocks. In the top-left corner of each spectrogram, we show the prediction of our time-sensitive model for the particular segment of EEG displayed.

The contributions of this work extend beyond model performance metrics. Many of the modeling choices in this paper were made in an effort to allow for ease of deployment. For instance, our model requires only a single 5-minute snapshot of EEG every 12 hours to prognosticate patient outcomes. This has clear advantages above approaches that require continuous knowledge of the EEG for assessment. One of the main strengths of this study was its multicenter nature, involving four institutions and more than 400 subjects. This size and heterogeneity of our dataset make our model more likely to be generalizable as it incorporates heterogeneities of subjects and medical management practices present in the different institutions. The model also employed a set of 57 QEEG features, several of which had not been applied for HIE prognostication before. Finally, we note that the coefficients of our model provide information that could aid in clinical prognostication, even if the model itself is not being directly used.

6.5.2 Performance

The overall AUC of the time-sensitive classifier was 0.83, which is 0.04 higher than the time insensitive approach and 0.13 better than the literature baseline. Our model demonstrates that EEG provides valuable prognostic information early after cardiac arrest, and that temporal trends can be used to further improve predictions. The time-sensitive model performance continued to improve as more data became available: from an AUC of 0.71 at 12 hours to an AUC of 0.79 at 72 hours), highlighting the incremental prognostic value of continuous EEG monitoring beyond 24 hours.

Several EEG patterns have been established as surrogate markers of brain injury or potential for recovery using conventional visual review and QEEG methods, however understanding the longitudinal changes of the brain's response to an initial insult across time using QEEG has not been as well studied. Our time-sensitive model had comparable sensitivity for good outcome prediction when compared to literature utilizing visual EEG review or QEEG methods - Sensitivity of 47% at a specificity of 95% at 72 hours.

The time-sensitive model also exhibited the best calibration of the three tested approaches by a small margin. Although the overall difference in calibration between the temporally sensitive (green) and temporally intensive (red) models is minor, the difference in the direc-

tion of mis-calibration is important. The temporally sensitive model is the more "optimistic" than the temporally insensitive model, tending to predict higher probabilities of positive outcomes than is true while the temporally insensitive model is more "pessimistic", often underestimating the probability of positive outcomes. This distinction between the models is important because an optimistic model may lead to fewer unnecessary terminations of patients than a pessimistic model would, even if their error-rate was similar.

6.5.3 Features Selected

Several of the features selected by our final model reflect prior findings in the literature. We found that information theoretic measures of EEG complexity and regularity were predictive of good functional outcomes while features measuring epileptiform discharge frequency were associated with poor outcomes. Two EEG complexity features (Cepstrum and Tsalis entropy) and EEG regularity contributed to the predictions with EEG data recorded in the first 24 hours of monitoring. These findings substantiate reports in the literature that specific EEG signatures observed during the first 24 hours after cardiac arrest have strong predictive value despite the presence of hypothermia and sedative use. Our final model also contained three connectivity features that were associated with poor outcome. Limited data is available in the literature describing the role of functional connectivity for outcome prediction in HIE. Several other QEEG features that were available to our model, but not selected, have been previously reported as useful in HIE prognostication. These features include measures of spectral content and burst-suppression This apparent difference from previous literature is likely due to our feature selection method (Elastic Net), which when faced with multiple informative but highly correlated features chooses one among them. Thus, features that have significant predictive value when considered individually, are not necessarily retained in the final multivariate prognostication model.

6.5.4 Limitations

This study has several limitations. Our method focused on EEG analysis exclusively, thus we cannot offer a quantitative statement regarding the role of this approach in multi-modal prognostication that integrates data from serial neurological exams, somatosensory evoked

potentials, neuroimaging, serum biomarkers, and visual EEG review. Multiple (often non-overlapping) reviewers assigned CPC scores. It is possible that the metric of these reviewers when scoring outcomes varied slightly. Such variations in judgment may have inadvertently introduced noise into outcome classification for those subjects at the border region between good and poor outcomes (CPC 3). We believe these effects were limited as only 4 subjects had a CPC score of 3 by 6 months, and the CPC outcome scale is relatively coarse-grained. The rate and reasons for withdrawal of life-sustaining therapies was not available, thus we cannot exclude the possibility that some poor outcomes resulted from self-fulfilling prophecies. Outcome data beyond hospital discharge was missing for some patients with CPC 3 or 4 at discharge; however, this problem was limited to 1.8% (8/438) of patients. A sensitivity analysis of our model excluding these cases showed no significant change in outcome prediction performance, therefore we do not believe that the lack of long-term follow up in this small subpopulation significantly affected our results.

In our analysis, the time-sensitive method outperformed a Random Forest classifier and the original CRI model utilizing five QEEG features (0.79 vs. 0.66 and 0.60 at 72 hours, respectively). We note that the performance of our Random Forest implementation was inferior to that reported by Tjepkema-Cloostermans et al. (0.90 at 24 hours). The discrepancy in performance may result from: (1) the heterogeneity of our data, which came from four different academic centers compared to the author's two-center study, (2) our decision to align data with respect to the start of EEG recording rather than the time of cardiac arrest, (3) use of a more extensive validation strategy (10-fold versus 1-fold), and (4) our decision to use only five out of nine QEEG features utilized in their Random Forest model.

The duration of EEG monitoring was not uniform across all subjects evaluated. This is a typical feature of retrospective studies in this patient population and is well illustrated by the decline in prediction calibration after 48 hours. For pragmatic reasons (to facilitate training machine learning models), we aligned EEG data based on time of initiation of EEG monitoring rather than on time of ROSC. Even though this caveat might affect the interpretability of our findings of how QEEG features change across time from initial injury, methodologically it resolved the problem of sparse and uneven data early after initiation of EEG monitoring. This adjustment in data alignment facilitated the deployment of a

sequential model that requires feeding features forward from one time interval to another, which would be compromised if large sections of data from the first 12 hours were unavailable. This limitation is partly mitigated by the fact that precise time of ROSC is often unknown, approximated, or misreported by third parties. Nevertheless, to evaluate the effect of data alignment timing in the model, we performed a secondary analysis in which the probabilistic output of the sequential method used the reported time of ROSC. This approach did not improve model classification performance when compared to the analysis aligned with time of EEG monitoring initiation (data not shown).

In the present analysis, we only utilized EEG data from a limited set of seven frontotemporal electrodes. The decision to limit the EEG data input aimed to reduce the dimensionality of data. This approach was supported by a recent observation that visual review of EEG from a limited set of channels had comparable sensitivity and specificity to analysis utilizing a full-set of electrodes in the comatose cardiac arrest population (Pati et al., 2017; You et al., 2017). Future studies should consider evaluating the performance of QEEG methods with limited montages as this approach might simplify and decrease costs of long term EEG monitoring in this population without significant compromise in outcome prediction performance.

Healthcare providers participating in decision making regarding life-support continuation were not blinded to EEG results, therefore we cannot account for the effect of self-fulfilling prophecies on the outcomes observed. Lastly, our time-sensitive model included measures of background continuity and spike detection, however the determination of EEG reactivity, seizure burden, and medical management changes were not incorporated to the present model. Integration of temperature, sedation levels, and medication use in predictive algorithms are fundamental to determine the clinical value of QEEG monitoring at the bedside and might improve outcome prediction accuracy.

In most studies of this population, EEG waveforms are temporally aligned with respect to the time of cardiac arrest, prior to feature extraction and analysis. The alignment of waveforms with respect to arrest is important for both physiological interpretation of model coefficients, and appropriate modeling of feature trajectories. In this study, we decided to align waveforms with respect to the time of EEG initiation instead of the time of arrest.

As we will explain in greater detail below, this decision was made to advance the primary objective of the paper: to develop a model that is both maximally applicable and maximally predictive. Alignment of waveforms with respect to arrest decreases model applicability because, for many patients, the time of arrest is either a rough estimate made by third parties or worse, is simply unknown. Hence, a predictive model that requires the time of arrest is less generally applicable than one which does not. Additionally, the alignment of waveforms with respect to the time of arrest may harm model performance because of large variance in the time between the arrest, and the initiation of EEG. These gaps in knowledge translate into sparse data structures that limits the diversity and sophistication of modeling approaches. Our sequential model, for instance, requires the ability to feed features forward from one time interval to another, which would not be possible if the first 12 hours of data following arrest was unavailable (as it often is). In brief, we acknowledge that the lack of alignment is a source of noise, but we also feel that this source of noise is acceptable given the gains in model applicability and general performance from a classification perspective. It is certainly the case that the time of arrest is important, and it is for this reason that we performed a secondary analysis where we attempted to updated the probabilistic output of the sequential model, given a known time of arrest. This analysis resulted in a statistically insignificant improvement in model classification performance. This does not mean that the time of arrest is uninformative, but rather, that the estimated values are unreliable.

6.5.5 Conclusion

In this study, we utilized a large, multi-center cohort of subjects with HIE to train an adaptive sequential prognostication model for the prediction of good and poor functional outcomes. The QEEG model we developed was time-sensitive, selecting specific feature values at specific points in time that were most predictive of outcome. The time-sensitive model had better classification performance and statistical calibration compared to several state-of-the-art baseline approaches. These results demonstrate that the statistical association between quantitative EEG features and neurological outcome in HIE changes over time, and accounting for these changes improves the prognostication performance of predictive models.

"Proximity is the mother of opportunity."

– *A Student*

Chapter 7

Study 2: A Bayesian Approach

Synopsis

The relationship between quantitative electroencephalography (qEEG) features and coma outcomes after cardiac arrest changes over time. Existing approaches account for the evolution of this relationship over long time periods but the dependence between the EEG-outcome in shorter time-frames is neglected. In this study, we present a time-varying logistic regression model to prognosticate coma outcomes that accounts for the short-term dependence of model coefficients, while also allowing for longer-term independence. Using a multi-center collection of 785 patients, we utilized a Dynamic Bayesian Network (DBN) to optimally combine (1) maximum-likelihood coefficients generated at contiguous time intervals with (2) coefficient estimates from a model of coefficient dynamics. We demonstrated that a logistic regression model using the DBN coefficients had better generalized performance (AUC = 0.86, TPR = 0.49 at FPR = 0.05) than a model using the maximum likelihood coefficient measures alone (AUC = 0.84, TPR = 0.46 at FPR = 0.05), and a Random Forest approach inspired by the literature (AUC = 0.82, TPR = 0.44 at FPR = 0.05). We also demonstrated that a model using the *optimal coefficients* consistently outperformed the baselines over time; the best performance was observed at 24 hours (AUC = 0.88), with a 0.06 absolute improvement over the Random Forest approach, and 0.13 absolute improvement over the clinical baseline. Lastly, our model was the best calibrated of the tested approaches, highlighting its utility as both a risk-scoring and decision support tool. Importantly, the coefficients of our model are interpretable, providing clinical insights that can guide care practice and prognostication, independent of formal deployment.

7.1 Introduction

A major challenge of qEEG prognostication is the time-sensitive nature of biomarkers: the same EEG activity, observed at different time-points following resuscitation, may have differing prognostic significance. Existing approaches account for the time-sensitive nature of the EEG-outcome relationship by re-estimating model coefficients in contiguous time-intervals [2]. While re-estimation of coefficients improves prognostic performance, it also (incorrectly) assumes that the EEG-outcome relationship in a given time-interval is statically independent from the relationship seen in its neighboring time-intervals. In reality, the EEG-outcome relationship is likely to change gradually, at rates determined by the underlying physiological process of recovery (or deterioration). For example, if burst-suppression predicts poor outcomes 12 hours after arrest, it may have similar (albeit, not identical) prognostic significance at hours 11 and 13.

In this paper we demonstrate a PAC prognostication methodology that accounts for both the evolution of the EEG-outcome relationship over long periods of time, as well as the dependence between the EEG-outcome relationship in short periods of time. Specifically, we deploy a Dynamic Bayesian Network (DBN) popularly used in navigation systems: the Kalman filter (KF) [143]. Since its description over half a century ago, the Kalman filter (and its many adaptations) has been applied to an impressive range of problems in the domain of EEG processing including brain-machine interface algorithms [144], artifact detection [145] and source localization [146].

For navigation systems, the KF combines imperfect satellite measurements with a vehicle's last known position and velocity to estimate its most likely current position. Similarly, we use the KF to estimate the optimal value of evolving logistic regression model coefficients, given their previous values, and imperfect maximum likelihood estimates from a given time interval. We demonstrate that our approach provides superior classification performance compared to state-of-the-art methods. We also assess the utility of our approach for more nuanced clinical use, such as risk-scoring, by reporting statistical calibration [147].

7.2 Methods

7.2.1 Data and Outcome

We utilized the continuous EEG recordings, and available clinical covariates of the 785 adult subjects diagnosed with in-hospital or out-of-hospital cardiac arrest described in Chapter 6. For this paper, we used the first 72 hours of recorded data from the seven frontal-most electrodes, exclusively [6]. When available, six clinical descriptors of the patient population were also collected: age, rhythm at arrest, time until return of spontaneous circulation (ROSC), sex, location of arrest (in hospital or out of hospital), and cause of arrest (cardiac, pulmonary, anesthesia or neurological). Rhythm at arrest was dichotomized as shockable (ventricular fibrillation or ventricular tachycardia) or non-shockable (asystole, pulseless electrical activity, and unknown). Missing clinical covariates were imputed using the K-nearest-neighbors algorithm [148].

The outcome of interest was the neurological status of patients six months after the arrest as indicated by the Glasgow-Pittsburg Cerebral Performance Categories scale (CPC). The CPC of patients were dichotomized into 'good' and 'bad' functional outcome groups where 'good' defined a CPC of 1 or 2, and 'bad' defined a CPC between 3 through 5 (inclusive). The 'bad' outcome class was used as the prediction target (coded as '1') for all modeling.

7.2.2 EEG Preprocessing and Quantitative Descriptors

EEE Waveforms were preprocessed to identify 5-sec artifact ridden segments using the approach described in Chapter 6. Then, starting from the time of EEG initiation, every 5-minute contiguous epoch of EEG was assigned an artifact score that reflected the total number of 5-second artifact ridden segments within it. For each hour of available EEG data, the 5-minute EEG epoch with the *least* artifact was retained for the extraction of quantitative EEG descriptors. If more than one segment satisfied this criteria, we selected the 5-minute epoch closest to the middle of the hour long interval. Hence, each patient had one 5-minute EEG epoch representing every 1 hour of continuously monitored EEG.

[6] Fp1, Fp2, F7, F3, Fz, F4, and F8

In Table 7.1 we show the 10 descriptors of EEG activity chosen for this study: three descriptors quantify the complexity of the waveforms (*Shannon's entropy*, *Hjorth complexity*, and *false nearest neighbor*), three descriptors quantify the connectivity between channel pairs (*coherence-delta*, *Phase Lag Index*, *cross-correlation magnitude*), and four descriptors quantify the presence of clinically relevant EEG signals categories (*standard deviation*, *regularity* , *diffuse slowing*, *spikes*).

Single channel descriptors of complexity and category were computed on seven frontotemporal electrodes: Fp1, Fp2, F3, F4, F7, F8, and Fz. Multi-channel descriptors of connectivity were computed on the 21 unique pairs of the frontotemporal electrodes. For analysis, descriptors were aggregated across channels into their average value. The aggregated descriptors were then z-scored across patients (scaled to have a mean of zero, and a variance of one). The computer code (Matlab, 2017b) used to extract our signal descriptors is freely available in an on-line repository [149].

Signal Feature	**Ref.**	**Brief Description**
Complexity		
Shannon Entropy	[76]	signal randomness
Hjorth Complexity	[81]	rate of change in μ signal frequency
False Nearest Neighbor	[82]	signal continuity and smoothness
Category		
Standard Deviation	[84]	average difference between the signal's value and mean value
Regularity (burst-suppression)	[1]	signal stationarity / spectral consistency
Diffuse Slowing	[85]	indicator of peak power spectral density less than 8Hz
Spikes	[85]	amplitude exceeds μ by 3 σ, lasting 70 ms or less
Connectivity		
Coherence - delta	[1]	correlation in δ power
Phase Lag Index	[89]	association between the instantaneous phase
Cross-correlation Magnitude	[90]	maximum correlation

Table 7.1: **Descriptions of EEG signal features.** Three features of EEG signal complexity, four features of clinically-relevant EEG signal categories, and three features that quantify the statistical dependence between EEG channel pairs

7.2.3 Bayesian Logistic Regression via a Kalman Filter

The prognostic significance of EEG descriptors are known to change with time [2]. To account for these changes, we propose a logistic regression model with time-varying coefficients:

$$logit(y^j) = \beta_0^j + \beta_1^j x_1 + ... + \beta_n^j x_n$$

where $x_{1:n}$ represent our n EEG signal descriptors, y^j represents patient at time-interval j and $\beta_{i:n}^j$ are coefficients that describes the relationship between the descriptors and the outcome at time interval j. We formally define a *time interval* as a uniform, contiguous period between cardiac arrest, and 72 hours after EEG initiation:

$$[(k(j-1)+1) : kj] \quad j,k \in \mathbb{N}$$

where k defines the time-interval length in hours. If $k = 12$, for example, the time interval associated with $j = 1$ are hours 1-12, and the time interval associated with $j = 2$ are hours 13-24.

To identify the time-varying coefficients of our model, we first generate coefficient values within each time-interval, independently. Specifically, an independent logistic regression model is trained for each time interval, using only the subset of patients having EEG recorded *within the interval*. We will refer to these as *coefficient measurements*, $\tilde{\beta}_i^j$, which have a corresponding *coefficient measurement error*, R_i^j (the standard error of the *coefficient estimates*).

The *coefficient measurements* will reflect changes in the feature-outcome relationship over time-intervals, but incorrectly assumes that the relationship is independent in short intervals of time. In reality, the EEG-outcome relationship in neighboring time-intervals are associated. To relate the *coefficient measurements* across time intervals, we specify a model of coefficient dynamics. Specifically, we assume that the current value of coefficients is dependent on their prior values, and a *coefficient velocity* (the rate of coefficient change across time-intervals):

$$f(\beta_i^{j-1}) = \beta_i^{j-1} + \Delta(\beta_i^{j-1}) = \bar{\beta}_i^j$$

Here $\Delta(.)$ is the velocity function, and $\bar{\beta}_i^j$ are the *coefficient estimates* generated by the model of system dynamics. Of course, the *coefficient estimates* are prone to error because the descriptor-outcome relationship is not fully explained by the dynamics model. Hence, for each *coefficient estimate*, we also specify a measure of *coefficient estimate error*, $\bar{P_i^j}$:

$$\bar{P_i^j} = P_i^{j-1} + Q$$

where P_i^{j-1} represents the *coefficient error estimate* from a previous time-interval, and Q is the *process noise*: a measure of expected coefficient variance across time intervals .

For practical use, we may initialize the model with *a priori values* of the *coefficient estimates* and *coefficient errors*, reflecting clinical intuitions about the nature of the feature-outcome relationship. Finally, a simple Dynamic Bayesian Network (the kalman filter) may be used to recursively determine the *optimal coefficient estimates* at each time interval as an error-weighted combination of *coefficient measurements*, and *coefficient estimates*:

$$\beta_i^j = \bar{\beta}_i^j + K_i^j[\tilde{\beta}_i^j - \bar{\beta}_i^j], \quad K_i^j \in [0,1]$$

The Kalman gain parameter, K_i^j, updates coefficient estimates based on the ratio of *coefficient estimate error* and the *coefficient measurement error*:

$$K_{(i,j)} = \frac{\bar{P}_i^j}{\bar{P}_i^j + R_i^j}$$

If the *coefficient measurement error* is large at time interval j, the KF will rely more strongly on *coefficient estimates* when generating the *optimal coefficient estimates*. Conversely, if the *coefficient measurement error* is small, the KF will rely more strongly on the *coefficient measurements*. The uncertainty of the *optimal coefficient estimates* are retained for use in coefficient estimation at the next time interval:

$$P_i^j = (1 - K_i^j)\bar{P}_i^{j-1}$$

In Figure 7.1 we illustrate how the Kalman filter constructs the *optimal coefficient esti-*

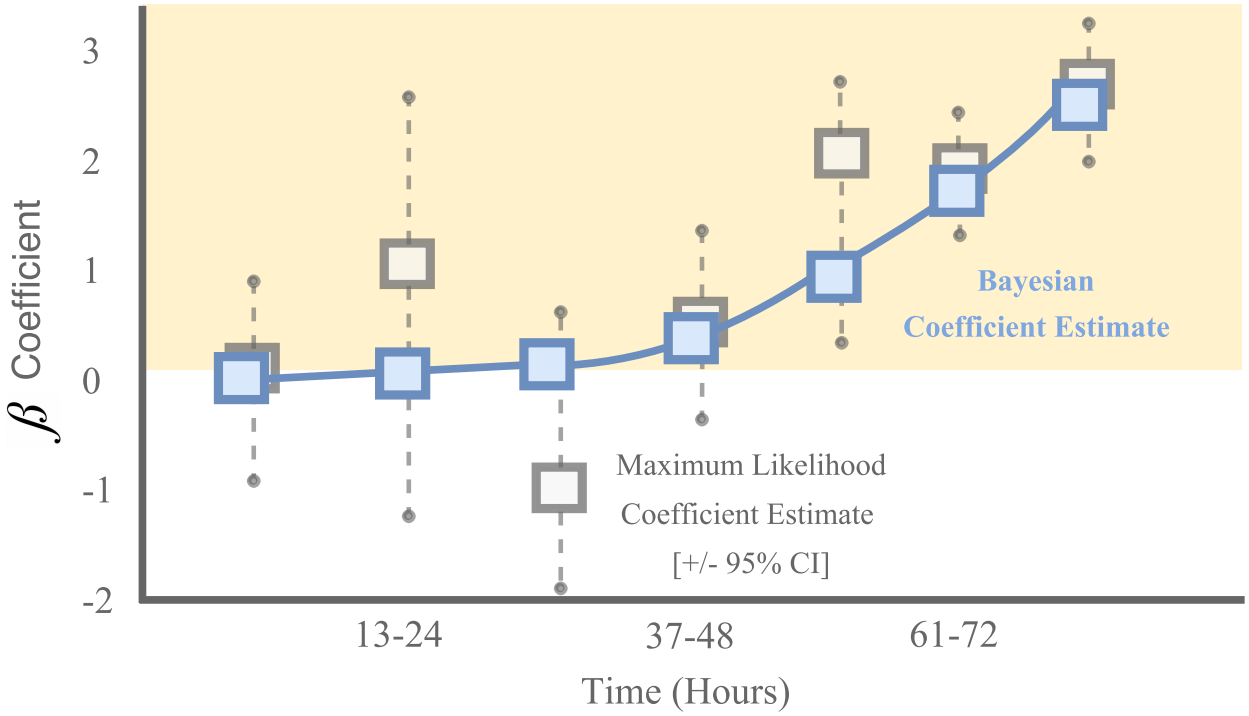

Figure 7.1: **Kalman Filtered Regression.** The Kalman filter constructs an optimal, Bayesian coefficient estimate (blue boxes), using coefficient measurement independently estimated in each time interval (gray boxes), and estimates of the coefficient measurement's standard errors (dashed lines).

mates (Blue boxes), using coefficient measurements for each time interval (gray boxes) and coefficient measurement error.

7.2.4 On-line, Artifact Adjusted Predictions:

As mentioned in the preprocessing approach, our algorithm is trained to prognosticate dichotomous patient outcomes using the *cleanest* 5-min segment of EEG available within each hour of recording. Implementation of this algorithm on a live data stream would require the ability to continuously update the prediction of patient outcomes, conditioned on the aforementioned artifact score. This problem is analogous to the problem of identifying *optimal model coefficients*, and can be addressed once again through the use of a KF. In this case, we specify a model of *outcome probability dynamics*. Specifically, we assume that the outcome probability at a given time is influenced by the outcome probability in the previous time point, and the outcome probability velocity (observed rate of change in the outcome

probability across time). We may use this model to generate an *outcome probability estimate*. Treating the probabilities generated by our model as *outcome probability measurements*, and the EEG artifact score (described in the preprocessing approach) as the *measurement error*, we may use the Kalman filter to generate a *optimal outcome probability estimate* over time that is robust to artifacts.

7.2.5 Model Interpretation

Our time-sensitive model coefficients may be interpreted through their odds-ratios. However, because the descriptors were z-scored prior to modeling, the odds-ratio will reflect the effects of a 1σ change in the descriptor, on the probability of a 'bad' outcome. In the results and discussion sections, we will refer to coefficient values less than $|0.4|$ as 'weak' associations, coefficients with values between $|0.4|$ and $|0.8|$ will be referred to as 'moderate' association, and coefficients with values exceeding $|0.8|$ will be refereed to as strong associations.

7.2.6 Model Comparison

The Proposed Approach: Bayesian logistic regression We utilized a Bayesian Logistic Regression model with five clinical descriptors (time to ROSC, age, gender, VFib, arrest Location), and the ten EEG descriptors shown in Table 7.1: *Shannon entropy*, *Hjorth complexity*, *false nearest neighbor*, *standard deviation*, *regularity*, *diffuse slowing*, *spikes*, *coherence - delta*, *Phase Lag Index*, and *cross-correlation magnitude*.

Baseline 1: Sequential logistic regression A unique logistic regression model was trained to classify patient outcomes in each time interval using all features from Table 7.1. This baseline represents a modeling approach where the feature-outcome relationship in neighboring time intervals is independent.

Baseline 2: Random Forest We implemented a Random Forest algorithm with features and settings inspired by the qEEG literature[1, 2]. Specifically, we used five previously described descriptors [1] (alpha/delta band-power, *standard deviation*, *coherence - delta*, *Shannon entropy*, and signal *regularity*), and parameterized a Random Forest algorithm in accordance with the 2017 Cerebral Recovery Index [2]. Like the sequential regression model, the coefficients of the Random Forest were re-estimated at each time-interval.

Baseline 3: Logistic Regression using clinical descriptors A logistic regression model was trained using clinical descriptors, exclusively: age, gender, time to ROSC, rhythm at arrest, and location of arrest. This baseline serves as a methodological "sanity check", allowing us to assess the difference in performance between the relatively complex quantitative EEG-methods and a simple clinical model.

7.2.7 Model Validation and Assessment

All models were validated using leave-one-out-cross-validation (LOOCV): for each individual patient, a separate model was trained using the data from all *other* patients. LOOCV is a rigorous form of model validation, providing performance characterization that is not sensitive to the statistical artifacts of data partitioning (e.g. splitting data in 70% training, 30% testing sets) [150].

We assessed the prognostication performance of all models using the Area Under the Receiver Operator Curve (AUC). We also report the true positive rates (TPR; ability to predict 'bad' outcomes) of the model, at various false positive rate (FPR; ability to predict 'good' outcomes) thresholds of 1% and 5%. Lastly, we report the statistical calibration of the models: the association between predicted probabilities of 'bad' outcomes, and the actual probabilities of 'Bad' outcomes, as observed in our data.

We compared the performance of our Bayesian Logistic Regression model against the baseline approaches for three time-interval lengths: k =6 hours, k =12 hours, and k =24 hours. For each of the three time-interval lengths, we also assessed the performance of the models for various resolutions of classification including: individual EEG epochs (one prediction per patient epoch), individual time-intervals (one prediction per patient time-interval), and individual patients (one prediction per patient). In the results and discussion section we show the performance of the model for a time-interval length of 12 hours, with a classification resolution at the time-interval level. Because many of the clinical covariates were subject to missing data, we also compared the performance of our proposed approach to the baseline models after excluding the clinical covariates from all models.

7.3 Results

7.3.1 Clinical Characteristics

In Table 7.2 we show the clinical characteristics of the population, partitioned by their dichotomized outcome classifications ('good' and 'bad'). Clinical covariates (other than age and gender) were missing for some patients. For the purpose of statistical modeling, all missing clinical covariates were imputed using the K-nearest-neighbors algorithm [148].

As discussed in the Chapter 3, our EEG descriptors capture aspects of signal complexity, category and connectivity. Figure 7.2 shows that descriptors of signal complexity are associated with outcomes differently: *Shannon entropy* is consistently associated with positive outcomes (blue line above red line), *False nearest neighbor* is consistently associated with negative outcomes (red line above blue line) while *Hjorth complexity's* association with outcome depends on time (red line crosses blue line). The association between some EEG descriptors of clinically relevant categories (*standard deviation*, *regularity*, *diffuse slowing*,

	'Good' Outcomes	**'Bad' Outcomes**
Sample size	300	485
Age in Years ($\mu[\sigma]$)	57 [14]	63 [16]
ROSC in Mins ($\mu[\sigma]$)	19 [16]	25 [20]
Sex (% Male)	68	75
Rhythm at Arrest (%)		
Vfib	69	34
Other	24	60
Unknown	7	6
Cause of Arrest (%)		
Pulmonary	50	34
Anesthesia	4	8
Neurologic	11	13
Other/Unknown (%)	35	45
Arrest Location (%)		
In Hospital	10	11
Out of Hospital	63	54
Unknown	27	35

Table 7.2: **Clinical Characteristics.** Characteristics of the collected demographic features for patients dichotomized into 'Good' (CPC 1 and 2), and 'Bad' (CPC 3 to 5) outcome categories, 6 months after ICU discharge. μ: mean, σ: standard deviation.

spikes) and outcome were influenced by time: *regularity*, for instance, is positively associated with outcome when observed between 12 and 72 hours after arrest, *diffuse slowing* is associated with positive outcomes between 48 and 96 hours following arrest, but *spikes* are consistently associated with 'bad' outcomes. EEG descriptors of connectivity (*coherence - delta*, *Phase Lag Index*, *cross-correlation magnitude*) were consistently associated with negative outcomes (red lines above blue), but their significance was also time dependent. *Cross-correlation magnitude* was associated with 'Bad' outcomes when observed between 48 and 96 hours after arrest while *coherence - delta* was associated with outcomes between 12 and 72 hours after arrest.

When considered jointly, the results shown in figures 5.2 and 7.2 provide strong motivation for our Bayesian approach to outcome prediction; it is clear from Figure 7.2 that the optimal value of a descriptor threshold to distinguish patient outcomes depends on time, but the optimal value in one time interval is also related to the optimal value in neighboring time-intervals. For time-intervals with lower patient data-densities , linking coefficient estimates across time should improve performance.

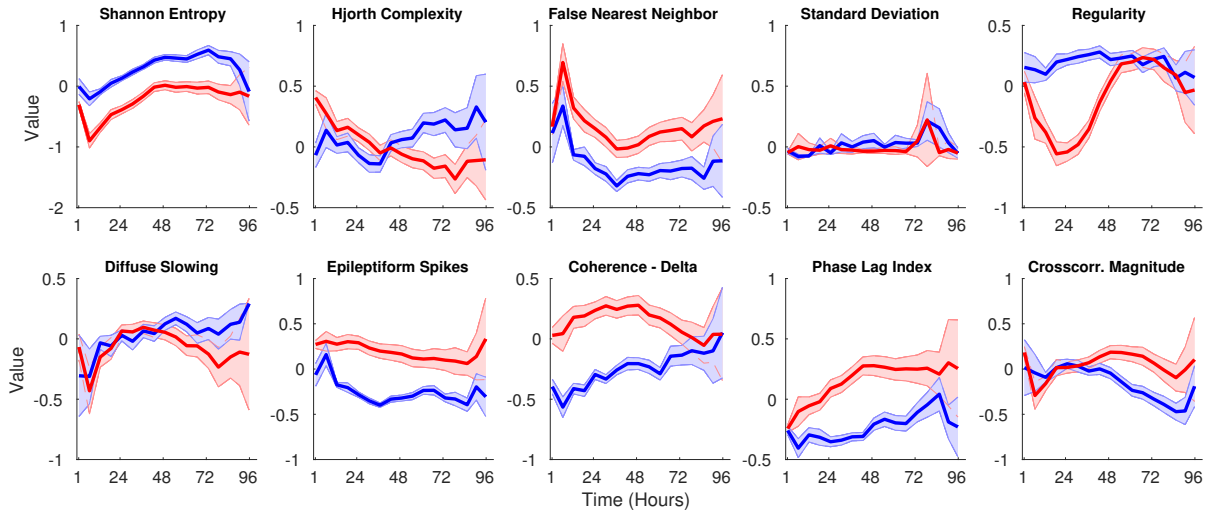

Figure 7.2: **Features Over Time**. The mean and standard errors of our 10 quantitative EEG descriptors shown in Table 7.1. We display the value of the descriptors over time for patients with 'good' (blue) and 'bad' (red) outcomes.

7.3.2 Main Results

In Table 7.3, we compare the LOOCV test-set performance of the Bayesian logistic regression model (with and without clinical descriptors) to the baseline approaches using a time-interval length of 12 hours. The proposed approach (row one) was found to have the best overall performance, while the model using only clinical descriptors (row six) was found to have the worst performance. Relative to the Random Forest baseline (row four), the Bayesian model (row one) exhibited a 0.04 absolute improvement in test-set AUC and a 8% improvement in TPR for FPR <= 1%. Even after excluding clinical features from the model, the Bayesian approach (row three) continued to outperform the Random Forest baseline (row four), albeit, much less strongly: 0.01 absolute improvement in AUC and a 4% improvement in TPR for FPR <= 1%. Importantly, the Bayesian approach consistently outperformed the sequential regression using *the same features* (compare row one to row two, and row three to row five). This improvement provides evidence that the feature-outcome relationship changes over time, but that the relationship in a given time-interval contains useful information about the relationship in neighboring time-intervals. In Table 7.4, we provide additional comparisons of the Bayesian approach to the baselines using different time-interval lengths (6, 12 and 24 hours) and different classification endpoints (epoch-level, interval-level, and patient-level). *In all cases*, the Bayesian approach was found to out-perform the baselines.

In Figure 7.3-A, we compare the classification performance (AUC) of the models over time. More specifically, the performance of the model at a given time-interval (x-axis) concerns patients with EEG recorded within that time-interval, *exclusively*. Note that the

Model	**Features**	**Temporal** Assumptions	**AUC**	**TPR** FPR=0.05	**TPR** FPR=0.01
Logistic Regression	5 Clinical, 10 EEG	Kalman Filter	0.86	0.49	0.29
Logistic Regression	5 Clinical, 10 EEG	Independent	0.84	0.46	0.26
Logistic Regression	10 EEG	Kalman Filter	0.83	0.47	0.25
Random Forest	5 EEG [2]	Independent	0.82	0.44	0.21
Logistic Regression	10 EEG	Independent	0.81	0.44	0.19
Logistic Regression	5 Clinical	None	0.73	0.26	0.11

Table 7.3: **Cumulative model performance**. AUC: Area under the receiver operator curve. TPR: True Positive Rate (the correct identification of 'Bad' outcomes). FPR: False positive rate: (the erroneous prediction of 'Good' outcomes, as 'Bad' outcomes)

	Calibration Error			AUC			TPR| FPR = 0.05		
Interval	Epoch	Interval	Patient	Epoch	Interval	Patient	Epoch	Interval	Patient
	Logistic Regression (Dynamic Bayesian Network, All Features)								
6 Hr	0.52	1.08	0.43	0.84	0.84	0.66	0.47	0.48	0.13
12 Hr	0.50	0.03	0.73	0.84	0.86	0.66	0.46	0.49	0.18
24 Hr	0.53	0.08	0.12	0.84	0.85	0.76	0.47	0.49	0.28
	Logistic Regression (Sequential, Independent, All Features)								
6 Hr	0.10	0.12	0.56	0.83	0.83	0.63	0.46	0.47	0.10
12 Hr	0.07	0.06	0.70	0.83	0.84	0.64	0.46	0.46	0.12
24 Hr	0.06	0.12	0.37	0.83	0.84	0.75	0.46	0.48	0.27
	Random Forest (EEG Features via Cloostermans et al, 2017)								
6 Hr	0.14	0.02	0.34	0.74	0.80	0.66	0.26	0.41	0.03
12 Hr	0.08	0.03	0.21	0.77	0.82	0.66	0.35	0.44	0.13
24 Hr	0.06	0.01	0.11	0.77	0.82	0.71	0.36	0.45	0.11
	Logistic Regression (Clinical Features)								
6 Hr	0.83	0.83	1.08	0.74	0.74	0.63	0.28	0.25	0.39
12 Hr	0.83	0.83	0.96	0.74	0.73	0.63	0.28	0.26	0.37
24 Hr	0.83	0.83	0.23	0.74	0.74	0.68	0.28	0.27	0.32

Table 7.4: **Performance given interval size.** Comparison of model performance calibration using 6, 12, and 24 hour time interval sizes. Performance is compared for various classification endpoint: individual EEG epochs (one prediction per patient epoch), individual time-intervals (one prediction per patient time-interval), and individual patients (one prediction per patient).

Bayesian approach (blue line) had superior classification performance consistently across time. Also note that the performance of the Bayesian approach was best 24 hours after cardiac arrest, providing an AUC of 0.88 and a TPR of 30% for FPR $< 1\%$. Lastly, observe that the classification performance of the Bayesian regression varied less over time compared to the baseline approaches.

In Figure 7.3-B, we compare the calibration of the Bayesian approach (blue line) to the sequential logistic regression (red line) and Random Forest baselines (yellow line). While all three approaches were well-calibrated (close to 45 degrees), the calibration of the Bayesian approach was best; the average difference between the predicted- and actual probability of 'bad' outcomes was smallest. The excellent calibration of the model validates its use as a risk-scoring method at the bedside. The direction of mis-calibration for the Bayesian approach implies that the model may assign a higher probability of 'Good' outcome than is always true (straying into yellow region). This direction of mis-calibration is superior to the sequential regression and Random Forest, which tend to assign a higher probability of 'bad' outcome than is true (straying into gray region).

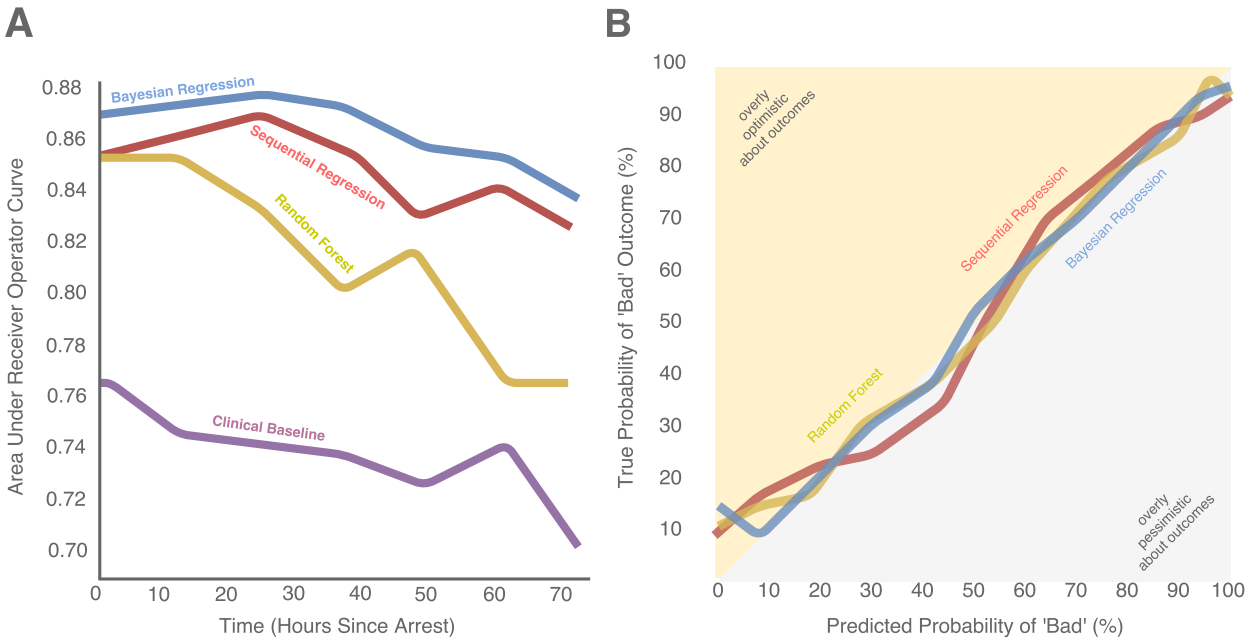

Figure 7.3: **Temporal Performance and Calibration.** The mean and standard errors of our 25 quantitative EEG descriptors over time for patients with 'good' (blue) and 'bad' (red) outcomes.

7.3.3 Model Coefficients

In Figure 7.4, we show the *coefficients measures* (gray circles), *coefficient measurement errors* (error bars), and *optimal coefficient estimates* (blue line) of our Bayesian logistic regression model trained using all available data. The first two rows of the figure show the coefficients of the 10 qEEG descriptors from Table 7.1, while the third row shows the coefficients of the clinical descriptors from Table 7.2. For most of the descriptors, there are only minor differences between the *optimal coefficient estimates*, and the *coefficient measures* (e.g. *Hjorth complexity*, *regularity*, gender). For some descriptors however, the differences between the *optimal coefficient estimates* and the *coefficient measurements* are more noticeable (e.g. *standard deviation*, *spikes*, *Shannon entropy*). As shown in Table 7.3 and Figure 7.3 these differences (even if they are minor) have important implications for the calibration, and classification performance of the Bayesian approach.

The prognostic significance of the all qEEG features changed over time. The *cross-correlation magnitude* features had close to zero prognostic power between 12 and 48 hours after arrest, but was associated with 'Bad' outcomes thereafter. Patients with *spike* activity

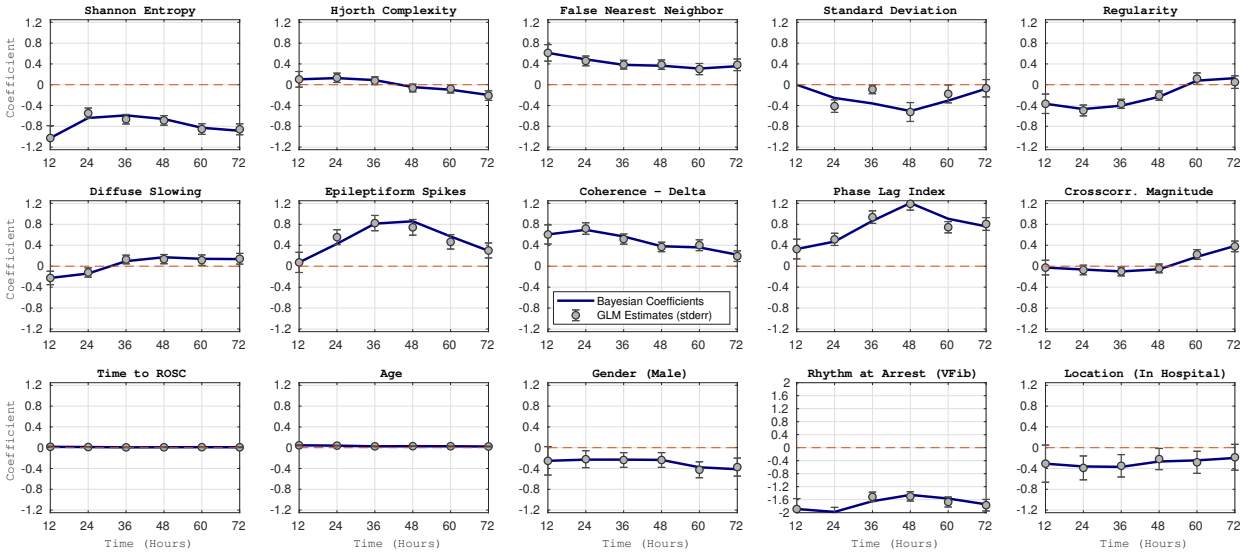

Figure 7.4: **Bayesian Model Coefficients.**

one standard deviation above the mean of the population at 36 hours were 2.2 times as likely to have 'Bad' outcomes but patients with the same level of *spike* activity at 12 hours were only 1.1 times as likely to have a 'Bad' outcome. The prognostic power of *coherence - delta* gradually decreased over time.

The prognostic polarity of some features also changed over time. *Hjorth complexity* was associated with 'Bad' outcomes 12-36 hours after arrest, and 'Good' outcomes thereafter. *Diffuse slowing* was associated with 'Good' functional outcomes before 24 hours, and 'Bad' outcomes thereafter. *Regularity* was associated with 'Good' outcomes between 12-60 hours after arrest, and 'Bad' outcomes thereafter.

The strongest qEEG predictor of 'Bad' outcomes was the *Phase Lag Index* 48 hours after arrest. Patients with *Phase Lag Index* values one standard deviation above the population mean were nearly 3.3 times as likely to have a 'Bad' outcome at 48 hours. The strongest predictor of a 'Good' outcome was the *Shannon Entropy* 12 hours after arrest. Patients with *Shannon entropy* values one standard deviation above the population mean were nearly 2.9 as likely to have a 'Good' outcome.

The clinical descriptors also had an important prognostic significance for the model. Male gender, arrest rhythm and in-hospital arrest location were all associated with 'Good'

outcomes. The most powerful clinical descriptor was a shockable arrest rhythm, which was associated with 'Good' outcomes.

7.4 Discussion

7.4.1 General Conclusion:

The relationship between qEEG features and coma outcome changes with time (Figure 7.2). Existing approaches account for the evolution of this relationship over long time periods but the dependence between the EEG-outcome in shorter time-frames is neglected[2]. In this study we presented a time-varying logistic regression model to prognosticate coma outcomes after cardiac arrest that accounts for the short-term dependence of model coefficients, while also allowing for longer-term independence. Using a retrospective multi-center collection of 785 patients, we deployed a dynamic Bayesian network to combine (1) maximum-likelihood *coefficient measures* generated at contiguous time intervals with (2) *coefficient estimates* from a model of coefficient dynamics to generate *optimal coefficients* over time. We demonstrated that a model using the *optimal coefficients* had better cumulative performance (AUC = 0.86, TPR = 0.49 at FPR = 0.05) than a model using the *coefficient measures* alone (AUC = 0.84, TPR = 0.46 at FPR = 0.05), and the Random Forest approach (AUC = 0.82, TPR = 0.44 at FPR = 0.05). The Bayesian model consistently outperformed the baselines over time; the best performance was AUC=0.88 at 24 hours after the arrest, a 0.06 absolute improvement over the Random Forest approach, and 0.13 absolute improvement over the clinical baseline. The Bayesian approach was also the best calibrated of the tested approaches, highlighting its utility as both a risk-scoring and decision support tool. Importantly, the coefficients of our model are interpretable, providing clinical insights that can guide care practice and prognostication, independent of formal model deployment.

The core contribution of this work lies in the proposed methodological framework. As illustrated in Figure 7.1, the strength of our method lies in its recursive identification of the optimal coefficient estimates. The method is Bayesian in the sense it uses *a priori estimates* of the coefficients, a model of coefficient dynamics, and independent coefficient measurements within each time-interval to generate optimal *a posteriori* model coefficients. More

specifically, the method adaptively combines the *coefficient estimates* from the dynamics model, and the *coefficient measures* from sequential regression, based on the ratio of their errors. These estimates are linked across time by setting the optimal coefficient estimate at a given time-interval as the *a priori value* for the next.

Beyond the novelty of our methodology, this study distinguishes itself from previous studies in this area by (1) taking a comprehensive approach to performance characterization (2) using a large, multi-center dataset (3) releasing computer code to facilitate reproducibility. We will discuss each of these contributions in more detail, below.

7.4.2 Model Validation:

It is common practice for studies in the medical informatics community to rely on single-fold validation (e.g. using 70% of collected data for model training and reserving 30% for model testing). For computationally demanding approaches, such as Deep Neural Networks, this validation practice is defensible because model development may require weeks of computation. However, for models with faster computation times, cross validation is essential to ensure that results are generalizable, and not influenced by the statistical characteristics of the training/testing data partitions [150]. Our study used a comprehensive form of model validation (leave-one-subject out) to compare the performance of our proposed approach, against the three baseline methods.

Importantly, our efforts to rigorously validate of our results went beyond LOOCV: while the main results of our study used a time-interval length of 12 hours, we also evaluated the performance of the model: across time, for varying time-interval lengths (1, 3, 6, and 24 hours), for varying classification endpoints (5-min EEG epochs, entire patient), and using different feature combinations (with and without clinical features). Each of these additional analyses were themselves LOOCV validated, and in all cases the performance of the Bayesian approach was equal to, or better than the selected baselines. This level of validation, required us to compute 7,995,225 models, but provides confidence that the results are valid, and generalizable.

7.4.3 Performance Metrics:

Existing models from the literature tend to assess their performance using classification metrics: AUC, and TPR at or beneath an FPR threshold. The first of these metrics (AUC) provides information on the generalized classification performance of the models. For practical purposes, the second classification metric (TPR at FPR threshold) is more useful, characterizing model performance when tolerance for prematurely withdrawing care, is very low (beneath 1% or 5%). As shown in Table 7.3 and Figure 7.3-B, our model provided the best classification performance of the tested approaches according to both these commonly used metrics.

While classification metrics are an important (and established) way to assess model performance in this domain, it is not the only type of performance metric to consider for the deployment of models at the bedside. Tools that generate decisions (via classification) are subject to stricter regulatory requirements compared to risk-scoring tools. Both types of tools (decision support and risk-scoring) may influence care practice, but decision-support tools do so by formally making decisions, while risk-scoring tools do so by informing care providers, who then make decisions. Indeed, risk-scoring tools are an important intermediary step before the deployment of models for decision support. Hence, we assessed the performance of our model as a risk-scoring tool by reporting its statistical calibration: the correspondence between the predicted and actual probability of outcome. In Figure 7.3-B we observed that the Bayesian model was the best calibrated of the tested approaches. For example, 40% of the patients that were assigned a risk-score of 40% by the Bayesian model, had a 'Bad' outcome. In contrast, only 30% of the patients that were assigned a risk-score of 40% by the Sequential regression model, has a 'Bad' outcome. We found that both the sequential regression, and Random Forest approaches were more likely to over-estimate the probability of a 'Bad' outcome than the Bayesian regression. Hence, our model was both the best calibrated of the tested approaches, and when errors in calibration occurred, our model was more optimistic about patient outcomes than the baseline approaches.

7.4.4 Model Coefficients

In Table 7.1, we outlined ten qEEG descriptors that reflected various aspects of EEG signal complexity, category, and multi-channel connectivity. These ten features were selected from a super-set of 52 qEEG features available in a public code repository [149]. The particular subset were selected to mitigate descriptor collinearity, which may negatively impact coefficient stability and interpretability.

The univariate association of the ten descriptors with coma outcome were observed to evolve over time (7.2), and the coefficients of our final Bayesian model (7.1), also reflected this evolution. *Shannon entropy*, and *standard deviation* were consistently associated with 'Good' outcomes, while *false nearest neighbor*, *spikes*, *coherence - delta*, and *phase-lag index* were consistently associated with 'Bad' outcomes. Here we will briefly discuss the implication of each descriptor's polarity, magnitude, and changes over time.

1. The *Shannon entropy* descriptor is a measure of a signal's statistical uncertainty; it was strongly associated with 'Good' outcomes 12 hours after arrest, and moderately associated with 'Good' outcomes thereafter. Higher values may imply more complex cortical activity, reflecting the survival of the synchronized neuron networks responsible for the diverse oscillations that drive complex EEG activity.

2. The *Hjorth complexity* descriptor is an estimate of EEG signal bandwidth; it was weakly associated with 'Bad' outcomes before 48 hours, and weakly associated with 'Good' outcomes thereafter. Higher values imply greater variability in spectral activity (it evaluates to zero for a pure sign wave). This implies that monorhythmic activity (e.g. alpha- or spindle- coma) are less indicative of a 'Bad' outcome in the first 48 hours after cardiac arrest than thereafter. We are uncertain why the descriptor changes polarity over time, but it may be related to as-of-yet unknown physiological processes associated with early versus later-stage recovery, or the timing of treatment protocols.

3. The *false nearest neighbor* descriptor is a measure of signal continuity and smoothness; it was weakly to moderately associated with 'Bad outcomes'. Higher values imply that the amplitude of an EEG signal at time t is less associated with the amplitude of the

signal at time *t-1*. Hence, the descriptor is sensitive to higher frequency activity and spiking, which may reflect epileptic seizures, or other interictal patterns.

4. *Standard deviation* is a measure of signal power; it was weakly to moderately associated with 'Good' outcomes. Higher values imply more signal power. These findings reflect previous observations in the qEEG literature [1, 151], and mesh with physiological intuitions. Higher *standard deviation* values reflect greater ionic current flows within the cortical areas of the brain [37], implying cellular activity.

5. *Regularity* is a measure of signal stationarity; it was weakly associated with 'Good' outcomes. Higher values imply greater signal stationarity. This feature was used previously to detect burst-suppression activity [1] in PAC patients. It is known that burst-suppression has differing prognostic significance depending on the time of its observation, and our results corroborate these findings [152, 153].

6. *Diffuse slowing* is an indicator of peak spectral power below 8Hz; it was weakly associated with good outcomes in the first 24 hours after arrest, and weakly associated with 'Bad' outcomes thereafter. Like *Hjorth complexity*, it is unclear why the feature changes polarity over time, but it is has been previously reported that the prognostic significance of diffuse slowing changes with respect to the time of arrest [105]. We speculate that it may result from physiological processes associated with early, versus later-stage recovery, or the timing of unknown treatment procedures.

7. The *spikes* descriptor identifies spiking activity in the EEG; it was weakly to strongly associated with 'Bad' outcomes, depending on the precise time of its observation. Higher values imply a larger number of *spikes*. *Spikes* may be indicative of neuronal network damage, resulting in spontaneous or erratic behavior. These findings reflect previous observations from the literature that *spikes* are associated with poor functional outcomes [154–157, 157–159].

8. *Coherence -delta* is a measure of correlation between the spectral power in the 0-8Hz band of two EEG channels; it was weakly to moderately associated with 'Bad' functional outcomes, as observed previously in the literature [1]. Higher values imply

greater similarity in the 0-8Hz spectral power of two channels, implying less diverse (and healthy) spatio-temporal activity.

9. *Phase Lag Index* is a measure of the instantaneous phase between two signals; it was weakly to strongly associated with 'Bad' outcomes depending on the time of its observation. Higher values imply greater synchrony in the instantaneous phase of EEG channels. Like *coherence-delta*, *Phase Lag Index* may characterize the amount of healthy spatio-temporal variation in the EEG.

10. *Cross-correlation magnitude* is a multi-channel measure of the correlation between one signal and another signal (subject to various time-delays); it was weakly associated with 'Bad' outcomes between 48 and 72 hours after the arrest. Higher values imply greater correlation between two signals across channels, once again, indicating less spatio-temporal variance in brain activity.

7.4.5 Limitations:

Our work has several limitations, and opportunities for extension, that are typical of observational studies in critical care. The prognostic performance of all models decreased over time (Figure 7.3). The changes in model performance may be the consequence of differing data volumes over time (See Figure 5.2) and the inability of the models to formally retain knowledge of their prior outcome predictions, when assessing outcomes latter in time. The later of these issues was beyond the scope of this study, but may be addressed in future work.

We observed large differences in performance between what the literature-inspired Random Forest baselines reported in their initial work (AUC=0.94,[2]), and what we observed of their performance on our study data (AUC =0.82). These differences are likely to result from our use of (1) a more heterogeneous, multi-center dataset, (2) our inability to perfectly replicate their model coefficients and (3) our validation strategy, which is less prone to over-fitting than the single-fold validation approach used by the original authors [2].

Another limitation of this work is the gaps in knowledge of care practice including sedative use and temperature management. Without knowledge of these factors, it is impossible to

know if (or to what extent) changes in our model coefficients were driven by underlying physiological processes associated with recovery/decline, or if changes are the consequence of stereotypic care practices (e.g. patient rewarming).

There are other limitations of this work that result from incomplete data; we were missing information on the intent of providers when life-support was continued or withdrawn. Of course, all observationally-trained algorithms suffer from the reality that they are primarily prognosticating clinician behavior, and only secondarily predicting patient outcomes. This can be problematic if algorithms are trained on data where patients were prematurely withdrawn from life-support, but we have no reason to believe this was a problem in our data.

Indeed, observational analyses are not without weakness, but the alternative (Randomized Control Trials) are subject to their own limitations including: (1) small sample sizes that may not be representative of actual health care settings and (2) ethical standards that limit the inclusion of patients unable to give consent. Importantly, the conclusions of large observational trials and randomized controlled trials (RCTs) are often comparable [160]. With health data being collected at a historically unprecedented scale (exabytes) and resolution (up to 500Hz), the importance of rigorous observational studies, and the advanced algorithmic techniques they facilitate, will only continue to grow.

"People are attracted to good ideas like sharks are to chickens!"

–A Lawyer

Chapter 8

Study 3: A Deep Learning Approach

Synopsis

In this chapter, we demonstrate that deep neural networks can provide an end-to-end solution for the PAC prognostication problem that does not require artifact detection, is simple to implement, and has performance that is on par or exceeding state-of-the-art feature-based models. Using 724 of the patients described in Chapter 5, we extracted a set of topographic-images that characterized the scalp-level energy in four clinically relevant EEG bands ($\delta, \theta, \alpha, \beta$). Several deep network topologies were applied to the images to automatically learn spatio-temporal features that were predictive of outcomes. We compared the performance of the deep network models to several feature-based approaches (SVMs, Discriminant Analysis, Decision Trees, KNN, Logistic Regression) using leave-one-institution out cross validation. The deep network models had the highest overall classification performance of the tested modeling frameworks (AUROC = 0.81), with TPR values of 0.24, 0.34 and 0.48 at FPR thresholds of 0%, 1% and 5% respectively. Overall, the performance of the 3D CNN exceeded that of best performing feature based approach (a RUS boosted random forest).

8.1 Introduction

In Chapters 6 and 7, we investigated the use of multi-feature models for the prognostication of PAC outcomes. At the heart of these two chapters was a traditional approach to machine learning: feature engineering, feature selection, and models for classification. We observed in the last two chapters that this approach can provide performance above the state-of-the-art, while also maintaining model interpretability. The weakness of the traditional approach however, is that feature engineering requires *substantial* engineering effort, and furthermore, it provides no guarantees that important features will not be overlooked by the investigators (See Figure 8.1, for an example).

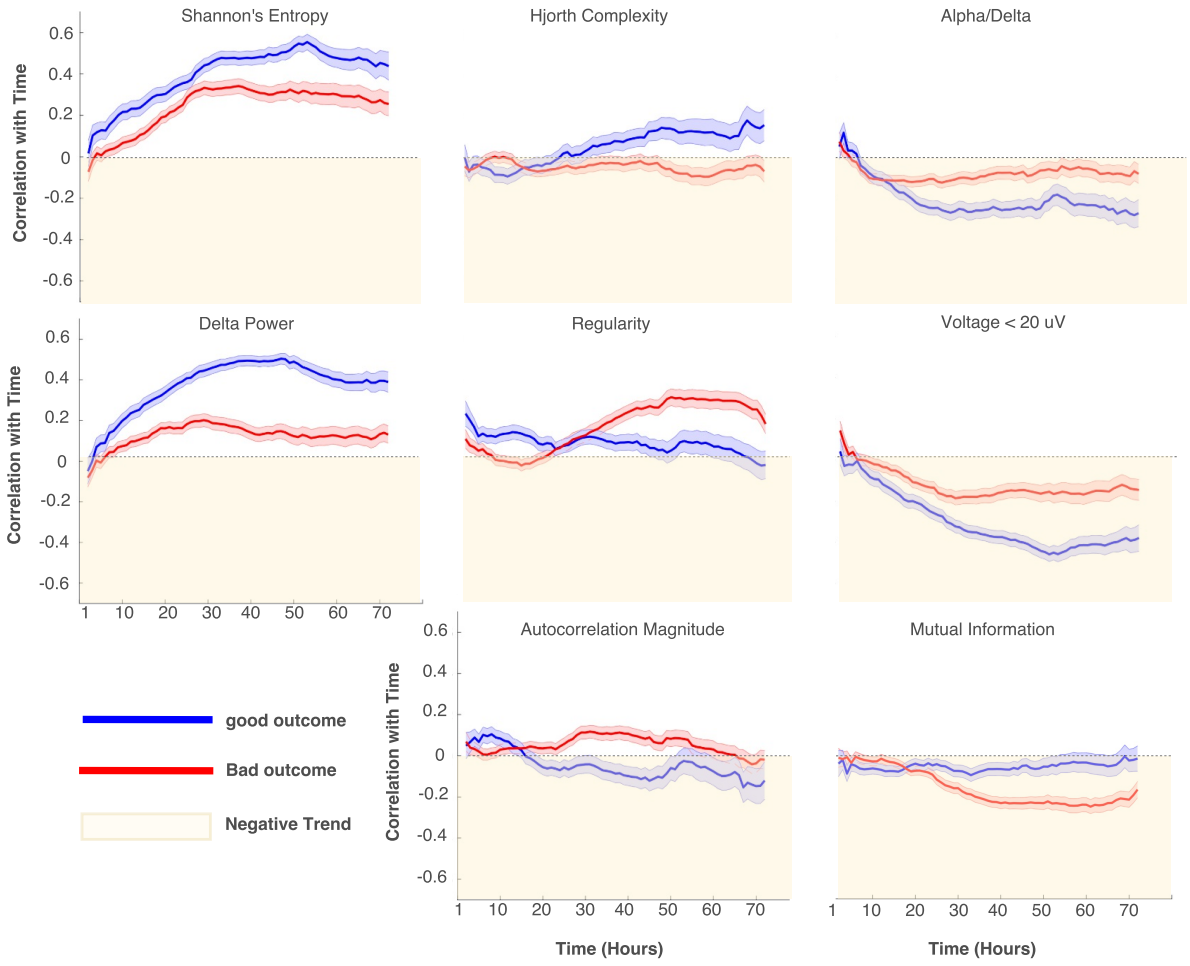

Figure 8.1: **Motivation for 'Deep' Approaches**. The correlation coefficients of several features with time, may also contain useful information.

In recent years, neural networks with multiple hidden layers (aka 'deep' networks) have gained significant attention in the machine learning community (and beyond) for their demonstrated performance in solving difficult classification problems without the requirement of feature engineering. The speech community was the first to leverage the power of these networks for practical purposes, and demonstrated tremendous increases in the efficacy of speech recognition when using 'deep' techniques compared to the then state-of-the-art (Gaussian Mixture Models) [161]. Today, deep techniques have moved beyond speech, and pervade practically every application area imaginable.

8.1.1 The Varieties of Deep Learning

Deep networks come in variety of forms. The earliest and simplest of these forms is the classical feed-forward neural network (FFNN) formulated half a century ago, and inspired by information processing in biological systems. The classical FFNN is composed of simple nonlinear kernels which, when used in combination, can approximate incredibly complex phenomenon. The core unit of the neural network is appropriately called a neuron, which performs a linearly weighted combination of its input features to generate an output:

$$y^{(1)}(x, w) = \sigma\left(\sum_{j=1}^{M} w_j f_j(x)\right)$$

where $\sigma(.)$ typically represents a non-linear activation function (e.g. sigmoid), and $f_j(x)$ represents some function of the raw data and M represents the number of input features. Neural networks gain their modeling flexibility by using the output of multiple neurons as the input to other neurons. For instance

$$y^{(2)}(x, w) = \sigma\left(\sum_{j=1}^{D} w_j y_j^{(1)}(x, w)\right)$$

where D represents the number of neuron outputs in a first layer, $y^{(1)}$ which were passed as inputs into a neuron in a second layer, $y^{(2)}$. Intuitively, this layering can be thought of as a series of functional transforms providing, at each level, an additional level of non-linear abstraction. Hence, FFNNs are *universal approximators*; that is, they have the ability to approximate any continuous function, provided that the network has the right number and

configuration of hidden units.

Our goal in this thesis is the classification of spatio-temporal time-series, but classical feed-forward network do not (on their own) account for either the spatial or the temporal dependencies in such data. One neural network formulation which accounts for temporal dependencies are Recurrent Neural Networks (RNNs). RNNs are nonlinear dynamical systems with an ability to map time-series to outcomes. Unlike FFNNs, they have have one layer dedicated to each point in a time-series that considers both the temporal input, as well as the neuronal activations from the previous temporal layer (See Figure 8.2). One weakness of classical RNNs is that the network treats all points in time as equally relevant for the generation of an outcome. But this may not be true for many classification problems (As we saw in chapters 6 and 7, feature values may matter only at particular points in time.). Long-Short Term Recurrent Neural Networks (LSTM) overcome this limitation of classical RNNs by weighting certain segments of data more strongly when generating outcomes.

While LSTM's are capable of accounting for temporal dynamics, they do not (on their own) capture spatial dependencies in the data. For this, there are convolutional neural networks (CNNs) [162]. In FFNNs, every neuron in one layer is densely connected with every neuron in the next layer; in contrast, CNNs selectively connect neurons in one layer

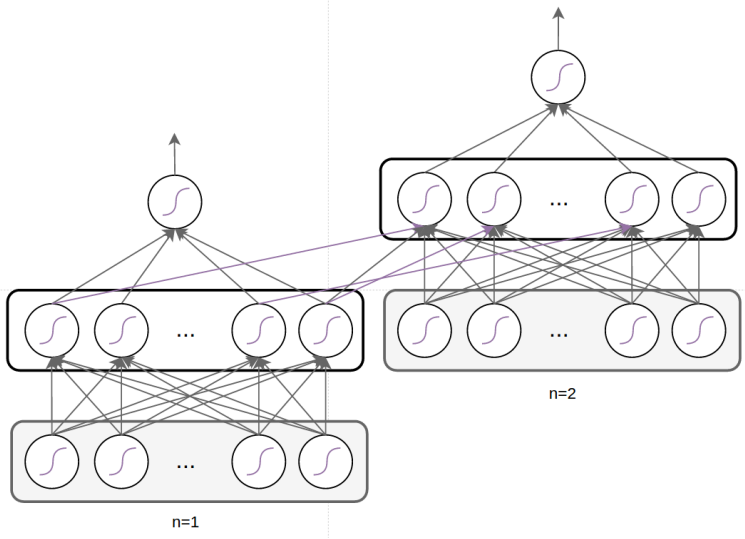

Figure 8.2: **Recurrent Neural Networks**. RNNs have one layer dedicated to each point in a time-series that considers both the temporal input, as well as the neuronal activations from the previous temporal layer

to neurons in the next layer *as a function of the spatial properties of their inputs*. This property of CNNs helps them naturally model both spatial (2D CNNs) and spatio-temporal (3D CNNs) relationships in data.

An advantage of neural network paradigms is that general model components (FFNNs, CNNs, LSTMs) may be combined to accomplish more specialized, or complex tasks. To account for both the spatial and temporal dependencies of data, one may use a Long-term Recurrect Convolutional Network (LRCN) [163]: One (or more) 2-dimensional CNN layers connected to an LSTM layer, connected to a final classification layer.

8.1.2 Deep Leaning For EEGs

Both 3D CNNs and LRCNs, are well suited for EEG analysis as they can account for both the spatial and temporal dynamics of the data automatically when classifying an outcome. Indeed, deep neural networks have already been applied for EEG prediction tasks in other contexts. Bashivan *et al.* performed a small study of 13 subjects where EEG was utilized to predict the number of symbols visually presented to a subject in a working memory experiment [164]. The authors reported that a LRCN network outperformed other methodologies (SVM, Logistic Regression, Random Forest) by as much as 8% (for the SVM). To our knowledge, deep learning has not been used in any published studies of the PAC coma population to date. Hence, in this study, we will explore the utility of 3D CNNs and LRCN deep networks for the prognosis of PAC patients.

8.2 Data and Outcome

This study utilized the continuous EEG recordings of 785 patients described in Chapter 5 after excluding an additional 61 patients due to EEG abnormalities that prohibited topographic feature extraction. The differences between the clinical characteristics of the subjects after exclusion of the additional 61 patients was statistically insignificant.

We used any available EEG data recorded within 72 hours of the time of cardiac arrest. As in previous chapters, the outcome of interest was the neurological status of patients six months after the arrest, as indicated by the CPC scale. Outcomes were dichotomized into

'good' (CPC of 1 or 2) and 'bad' (CPC between 3 through 5) functional outcome groups. 265 subjects had 'good' functional outcomes, and 459 had 'bad' functional outcomes. The 'bad' outcome class was coded as 1. The clinical characteristics of the subjects were *not* included as features for any of the analyses performed in this chapter.

8.2.1 Preprocessing Approach

For each EEG channel, we computed a windowed Welch's power spectral density. The window size was 5 minutes, with a stride of 1 minute (80% overlap). The power densities were used to estimate the spectral energy within four clinically relevant EEG bands: δ (0-3Hz), θ (4-7Hz), α (8-15Hz), β (16-31Hz). For each energy band, in each 5-minute window, channel data were combined to yield an $8x8$ gray-scale topographic image, describing the spatial distribution of the band energy on the scalp. Topographical images were extracted for any data falling within the first 72 hours after the time of cardiac arrest; missing data, and periods where EEG was withdrawn were represented as zeros. Hence our *initial dataset* was a 5D tensor with dimensions:

$$(724\,patients) \times (72\,hr \cdot 60\,\frac{windows}{hr}) \times (8\,pixels) \times\ (8\,pixels) \times (4\,freq.bands)$$

Topographic intensity was normalized such that 0 corresponded to an energy spectral density of -10dB, while 1 corresponded to an energy spectral density of 20 dB. Our decision to represent patient data as a time-series of topological energy plots was inspired by the practices of the speech community, who transform audio waveforms into physiologically relevant spectral bands prior to applying deep learning [165].

Although 724 is a large number of patients in the world of PAC prognostication, it is far too small to effectively train deep networks. To address this challenge, each patient's data was transformed into a set of 600 pseudo-patients by randomly re-selecting topographical images, in each hour of patient data. This provided a *final dataset* of size:

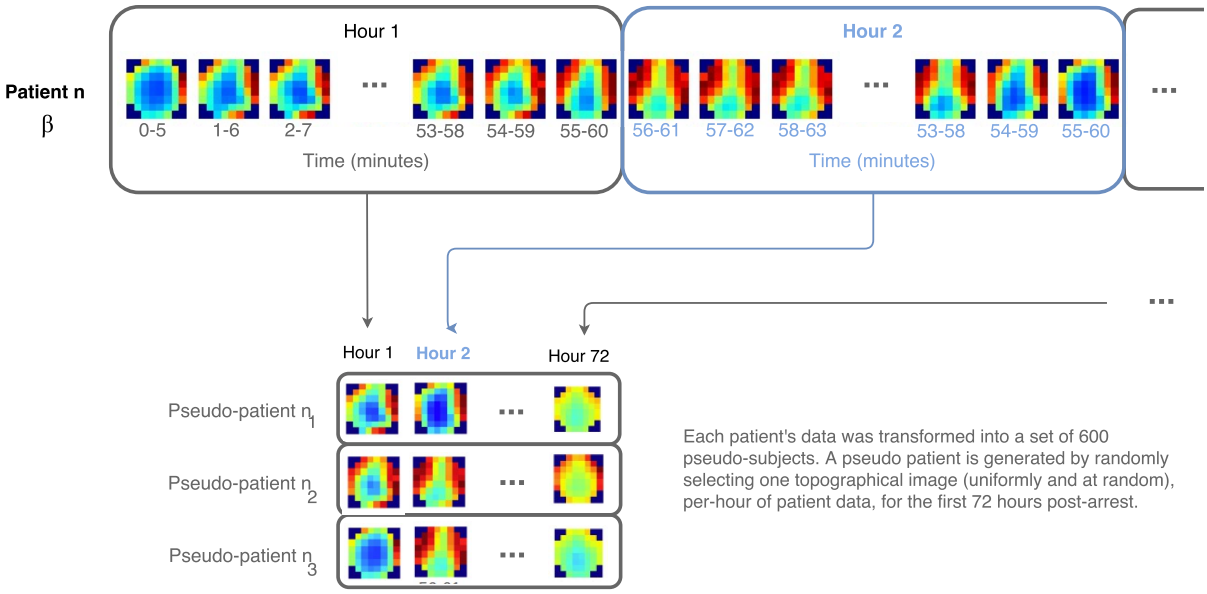

Figure 8.3: **Pseudo-patient Generation**.

$$(724\,patients \cdot 600\,\frac{psudopatients}{patient}) \times (72\,hr) \times (8\,pixels) \times (8\,pixels) \times (4\,bands)$$

The approach used to generate pseudo-patients is illustrated in Figure 8.3. The preprocessing approach is anticipated to help the deep networks learn representations of the data that will better generalize to unseen patients.

8.3 Methods

We evaluated the performance of 100 2D CNN models, 100 3D CNN models and 100 LRCN models for the prediction of coma outcomes using the pre-processed data.

The 2D and 3D CNN models were constructed as: an input layer [7], followed by one or more convolutional layer stacks (convolution layer, batch normalization layer, activation layer, max-pooling layer, dropout layer), followed by a flattening layer, followed by one or more feed-forward layer stacks (Dense layer, dropout layer), followed by an output layer.

The LRCN models were were constructed as: an input layer, followed by one or more 2D convolutional layer stacks (convolution layer, batch normalization layer, activation layer,

[7]For 2D Networks, 3D data were converted first into a 2D representation

max-pooling layer, dropout layer), followed by a flattening layer, followed by a dropout layer, followed by an LSTM layer[8], followed by an output layer.

All models were trained using Keras 2.0.9 with a Tensorflow 1.5.0 backend. All network weights were optimized using stochastic gradient descent on 250 epochs of the data, trained in batches that were 1% of the epoch size. The loss function of all neural network models was *binary cross-entropy*. The learning rate was reduced by a factor of five for three epochs if model performance plateaued for five epochs (less than 0.01 change in loss). For batch normalization layers, default Keras parameters were used. For all networks, the final layer consisted of a single neuron with sigmoid activation.

The hyper-parameters for each neural network architecture were selected uniformly and at random from a set of parameter-specific options shown in Table 8.1. From the 300 trained models, We retained the 2D CNN, 3D CNN and LRCN models with the best performance for comparison against baseline approaches.

8.3.1 Baselines

We compared the classification performance of the neural network models against the following modeling approaches: Decision Trees, Discriminant Analysis, Logistic Regression, Support Vector Machines, k-Nearest Neighbors (k-NN) and Ensemble Learning. All baseline methods used the ten qEEG features from Chapter 7, extracted within 12 hours intervals ($10\,features \times 6\,intervals = 60\,features$). Features included three descriptors of complexity (*Shannon's entropy*, *Hjorth complexity*, and *false nearest neighbor*), three descriptors of connectivity (*coherence-delta*, *Phase Lag Index*, *cross-correlation magnitude*), and four descriptors of clinically relevant EEG signals categories (*standard deviation*, *regularity* , *diffuse slowing*, *spikes*)

8.3.2 Performance Metrics

The performance of models were compared using several metrics: the area under the receiver operator curve (AUROC); true positive rate (TPR) at false positive rate thresholds of 0%,1%, and 5%; Accuracy; and Brier Score (a measure of statistical calibration). In Table

[8]Bias was disabled for the LSTM layer.

Category	Hyper-Parameter	Options Explored
General		
	Batch size	1%
	Number of Epochs	125
Initialization		
	Initial Network Weights	LeCun Normal, He Normal, Glorot Normal
	Initial Network Bias	Zeros
Optimizer		
	Optimizer	Stochastic Gradient Descent
	Learning Rate	[1E-1: 1E-1: 1E-9]
	Clipnorm	1
	Clipvalue	0.5
	Momentum	0.9
	Decay	1E-6
	Nesterov Momentum	True
3D CNN		
	Number of Convolutional layers	[4, 3, 2, 1]
	(per layer) CNN filters	[4, 8, 16, 32, 64, 128]
	(per layer) CNN kernal size	[(4,4,4), (3,3,3), (2,2,2)]
	(per layer) CNN strides	[(2,2,2), (1,1,1), (1,1,1)]
	(per layer) CNN batch normalization	[True,False]
	(per layer) CNN activation	ReLU
	(per layer) CNN MaxPooling Pool Size	[(2,2,2), None]
	(per layer) CNN dropouts	[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, None]
	Number of FFNN layers	[2, 1]
	(per layer) FFNN neurons	[5, 10, 20, 40, 60, 120]
	(per layer) FFNN activation	[ReLu, Tanh]
	(per layer) FFNN dropouts	[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, None]
2DCNN		
	Number of Convolutional layers	[4, 3, 2, 1]
	(per layer) CNN filters	[4, 8, 16, 32, 64, 128]
	(per layer) CNN kernal size	[(4,4), (3,3), (2,2)]
	(per layer) CNN strides	[(2,2), (1,1)]
	(per layer) CNN batch normalization	[True, False]
	(per layer) CNN activation	ReLU
	(per layer) CNN MaxPooling Pool Size	[(2,2), None]
	(per layer) CNN dropouts	[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, None]
	Number of FFNN layers	[3, 2, 1]
	(per layer) FFNN neurons	[4, 8, 16, 32, 64, 128]
	(per layer) FFNN activation	ReLU
	(per layer) FFNN dropouts	[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, None]
LRCN		
	Number of 2D Convolutional layers	[3, 2, 1]
	(per layer) CNN filters	[4, 8, 16, 32, 64, 128]
	(per layer) CNN strides	[(2,2), (1,1)]
	(per layer) CNN kernal size	[(3,3), (2,2)]
	(per layer) CNN batch normalization	[True, False]
	(per layer) CNN activation	ReLU
	(per layer) CNN MaxPooling Pool Size	[(2,2), None]
	(per layer) CNN MaxPooling strides	[(2,2), (1,1)]
	(per layer) CNN dropouts	[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, None]
	LSTM recucrent dropout	[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, None]
	LSTM Nodes	[2, 4, 8, 16, 32]
	LSTM Activation	[Tanh, Sigmoid]
	LSTM Recurrent Activation	[Hard Sigmoid, Sigmoid]
	LSTM Recurrent Initialization	Orthogonal
	LSTM Kernal Initialization	Glorot Normal

Table 8.1: **Neural Network Hyperparamters.** We explored various parameterizations of 2D CNNs, 3D CNNs, and LRCNs. Hyper-parameters were selected uniformly, and at random from the set of listed options (third column).

Primary Metric	Acronym	Purpose
Area under the reciever-operator curve	AUROC	The area under a curve that explores the relationship between true positive rate and false positive rate for various classification thresholds
Accuracy	ACC	% of cases correctly classified
F1 Score	F1	The harmonic mean of precision and recall
Breir score	BScore	measures the accuracy of probabilistic predictions (calibration)
TPR when FPR = 0%, 1% and 5%	$TPR\|FRP_{0,1,5}$	TRP given various FPR thresholds

Table 8.2: **Performance Metrics.** Metrics used to evaluate our model performance. TPR: True Positive Rate. FPR: False Positive Rate.

8.2, we show the selected performance metrics, and provide a brief commentary on their interpretation.

8.3.3 Validation Strategy

For methods to be useful, they must robustly generalize outside of the institutions where the models were trained. To evaluate how well our models might generalize to patients beyond the five medical centers where the data was collected, we performed leave-one-institution-out cross validation. That is, the data of four institutions was used to train models that were evaluated on the held-out fifth institution. Hence, each testing fold contained data from unseen subjects, at a unique institution. We used the average performance of models across the five folds when comparing performance. 20% of the training data in each fold was used for model validation purposes, and the models with optimal validation set performance were applied to the held-out test data.

8.4 Results

In Figure 8.4, We compare the average topographic images of subjects with good and bad outcomes in each of the clinical bands $(\delta, \theta, \alpha, \beta)$, at 12 hour intervals. We see clear spatio-temporal patterns in the data across the four extracted bands that differentiate good and bad outcomes (particularly in α). The figure provides evidence that our preprocessing approach is appropriate.

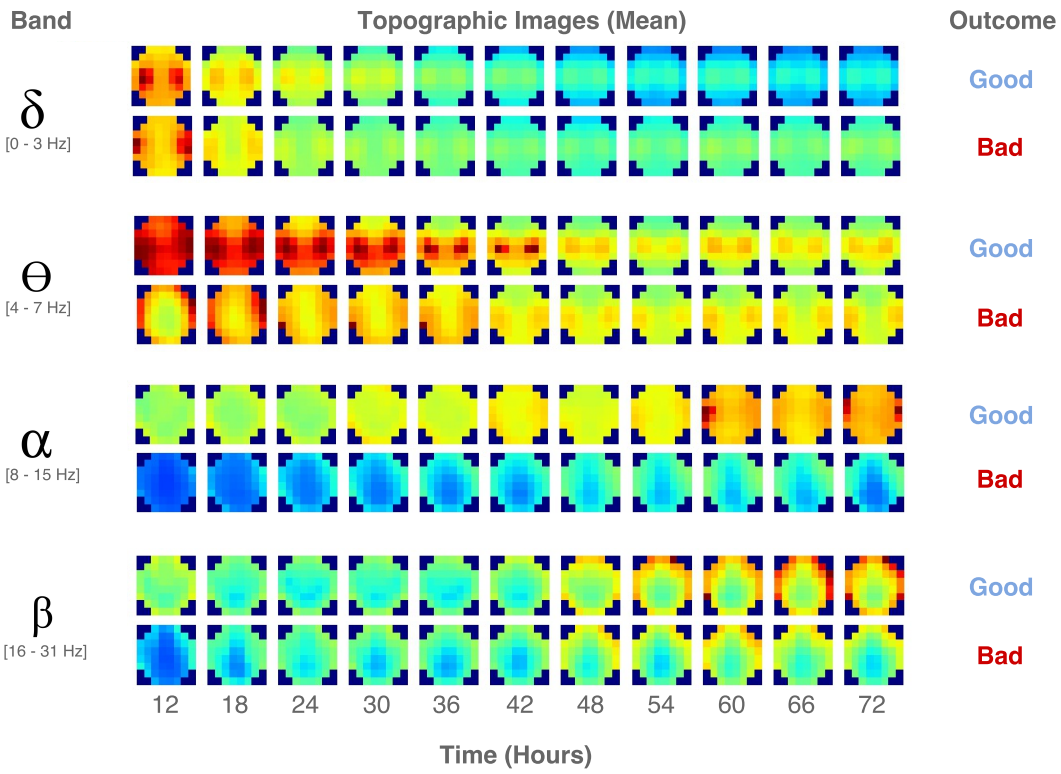

Figure 8.4: **Topographic Plots (Mean)**. The average topographic images of subjects with good and bad outcomes in each of the clinical bands $(\delta, \theta, \alpha, \beta)$ at 12 hour intervals since the time of cardiac arrest. The plots are oriented as though the eyes of the patients were facing the top of the page.

8.4.1 Main Results

In Table 8.3, we compare the performance of the deep learning methods against the baselines. Of the three tested deep learning approaches, the 3D CNN model provided the best AUROC on the held-out institutional data (0.81 for the 3D CNNs versus 0.79 for LRCNs). The 2D CNN model exhibited the lowest AUROC performance of the three approaches (0.71), which is not surprising given its inability to learn representations of the data, across time.

Incredibly, the deep networks approaches provided performance on par with (and sometimes exceeding) that of the best-performing feature-based approaches. The 3D CNNs had the highest average AUROC of the tested models (AUROC = 0.80). These results are particularly striking when we consider the fact that the networks are learning data representations, in a fully-automated way.

For use in an actual clinical environment, families and care providers are likely to be most

	AUROC	TPR| FPR=0	TPR| FPR=1%	TPR| FPR=5%	ACC	BScore
Neural Networks						
2D CNN	0.71 (0.09)	0 (0.0)	0.09 (0.06)	0.23 (0.10)	0.65 (0.08)	0.25 (0.03)
**3D CNN	0.81 (0.04)	0.24 (0.05)	0.30 (0.05)	0.45 (0.09)	0.78 (0.05)	0.39 (0.05)
LRCN	0.79 (0.08)	0.19 (0.11)	0.34 (0.15)	0.48 (0.15)	0.76 (0.07)	0.38 (0.03)
SVM						
*Linear	0.76 (0.13)	0.17 (0.18)	0.21 (0.16)	0.32 (0.19)	0.72 (0.09)	0.28 (0.09)
Quadratic	0.72 (0.14)	0.14 (0.09)	0.16 (0.09)	0.26 (0.12)	0.72 (0.07)	0.28 (0.07)
Fine Gaussian	0.75 (0.14)	0.18 (0.14)	0.2 (0.12)	0.31 (0.19)	0.72 (0.09)	0.28 (0.09)
Medium Gaussian	0.56 (0.03)	0.01 (0.02)	0.01 (0.02)	0.06 (0.04)	0.67 (0.11)	0.33 (0.11)
Coarse Gaussian	0.71 (0.13)	0.07 (0.12)	0.08 (0.12)	0.15 (0.1)	0.7 (0.1)	0.3 (0.1)
Discriminant Analysis						
Linear	0.77 (0.15)	0.21 (0.13)	0.23 (0.14)	0.34 (0.25)	0.73 (0.09)	0.27 (0.09)
*Quadratic	0.78 (0.14)	0.22 (0.13)	0.25 (0.14)	0.4 (0.22)	0.72 (0.11)	0.28 (0.11)
Cubic	0.72 (0.13)	0 (0)	0 (0)	0.15 (0.24)	0.64 (0.11)	0.36 (0.11)
Decision Trees						
Simple	0.68 (0.11)	0 (0)	0 (0)	0 (0)	0.72 (0.12)	0.28 (0.12)
Medium	0.7 (0.09)	0.05 (0.11)	0.05 (0.11)	0.11 (0.16)	0.7 (0.08)	0.3 (0.08)
Complex	0.69 (0.07)	0 (0)	0 (0)	0 (0)	0.69 (0.07)	0.31 (0.07)
*RUS Boosted	0.8 (0.1)	0.25 (0.2)	0.25 (0.2)	0.39 (0.17)	0.73 (0.07)	0.27 (0.07)
Ensemble Boosted	0.74 (0.12)	0.12 (0.17)	0.12 (0.17)	0.14 (0.19)	0.75 (0.07)	0.25 (0.07)
Ensemble Bagged	0.77 (0.1)	0.09 (0.14)	0.11 (0.13)	0.22 (0.15)	0.75 (0.06)	0.25 (0.06)
KNN						
Fine	0.61 (0.04)	0 (0)	0 (0)	0 (0)	0.65 (0.04)	0.35 (0.04)
Medium	0.72 (0.12)	0.04 (0.1)	0.04 (0.1)	0.18 (0.18)	0.71 (0.07)	0.29 (0.07)
*Coarse	0.76 (0.14)	0.23 (0.19)	0.23 (0.19)	0.28 (0.2)	0.7 (0.12)	0.3 (0.12)
Cosine	0.74 (0.15)	0.03 (0.06)	0.03 (0.06)	0.07 (0.1)	0.7 (0.12)	0.3 (0.12)
Cubic	0.69 (0.12)	0 (0)	0 (0)	0.08 (0.12)	0.69 (0.08)	0.31 (0.08)
Weighted	0.73 (0.11)	0.07 (0.1)	0.08 (0.11)	0.23 (0.19)	0.7 (0.08)	0.3 (0.08)
Logistic Regression	0.74 (0.13)	0.09 (0.09)	0.1 (0.09)	0.21 (0.18)	0.71 (0.09)	0.21 (0.07)

Table 8.3: **Comparison of Model Performance.** We show the mean (and standard deviation) of the models across the held-out institutional folds. * denotes the best performing model in-class. ** denotes best performing model, overall.

interested in the ability of our model to correctly identify bad outcomes (TPR), assuming the model is unlikely to ever mistake a good outcome for a bad one (ie. low FPR). In Table 8.3, we compare the TPR of the models for various FPR thresholds (0%, 1%, and 5%). Here again, the performance of the 3DCNN is on par with, or exceeds that of the best feature-based method (RUS Boosted Random Forest). At a FPR threshold of 0%, the performance of the 3D CNN is proximal to the RUS Boosted Random Forest (TPR = 0.24 vs 0.25), but importantly, the standard deviation in 3D CNN model performance across institutions is much lower (σ = 0.05 vs. σ = 0.20). At an FPR threshold of 1%, the mean TPR of the 3D CNN was 0.30 (0.05 absolute improvement over the best performing feature-based approach), and the mean TPR of the LRCN was 0.34 (0.09 absolute improvement over the

best performing feature-based approach). Similar improvements were observed at a FPR threshold of 5%. Importantly, the calibration of the deep learning approaches were among the worst of the tested methods, matching prior observations from the literature [166].

8.5 Discussion

In this chapter, we demonstrated that representation learning can provide an end-to-end solution for the coma prognostication problem that is on par with, or even exceeding, the performance of state-of-the-art feature-based models. What is particularly impressive about these results is the relative simplicity of the pre-processing approach for the neural networks compared to the feature-based approaches. The deep networks require four, $8x8$ topographic images that represent spatio-temporal energy in four clinically relevant bands $(\delta, \theta, \alpha, \beta)$, computed once every hour over the 72 hour period following cardiac arrest. In contrast, the feature based models required a sophisticated artifact detection pipeline (discussed in Chapter 5), the generation of computationally expensive features (some taking seconds to compute on a given 5-min segment of EEG), and the statistical consolidation of features across multiple channels before any model may be applied. Indeed, this chapter serves as yet-another testament to the elegance and impressive performance of artificial neural networks.

The pre-processing approach we selected in this study was grounded in common practices from other machine learning communities (8x8 images from popular datasets in vision [167], and spectral energy bands from speech [165]), it is possible that a different pre-processing approach might allow the networks to extract richer representations from the data, and further improve performance.

Deep networks have many user specified hyper-parameters, and exponentially more trainable parameters. The most successful 3D convolutional neural network model we trained in this study had 282,233 trainable parameters (390 times the number of patients we collected). With so many possible configurations of the hyper-parameters, it is possible (and quite likely) that with additional tuning, the performance of our models would have exceeded that of the feature based approaches.

In previous chapters, we often dedicated time to the interpretation of our results, and

made efforts to understand the implications of model coefficients beyond the context of the formal modeling frameworks at hand. Although there is ample evidence that deep learning may outperform techniques that rely on feature engineering, the ability to interpret these networks is, in most cases, lost. For a given function with an identical mapping from inputs to outputs, there will exist 2^H unique settings of parameters, where H is the number of "hidden" units in the model [168]. The precise features learned by neural networks, and how to interpret them, are ongoing and important topics of research which lay beyond the scope of this thesis.

"Everything is hard, until it's easy."

–A Friend

Chapter 9

Conclusions

Synopsis

The goal of this thesis was to develop techniques for the rapid prognostication of post-anoxic coma (PAC) patients. We conclude that quantitative EEG (qEEG) methods are capable of detecting early signs of neurological recovery while coma persists. We found that prognostication of bad outcomes (CPC > 2) may be performed robustly by spatio-temporal deep learning methods, and feature-based approaches that account for time. The performance of our methods on unseen data was 0.8 AUROC, with a true positive rate of 25% at a false positive rate of 0%. Hence, our results provide strong motivation for the deployment of quantitative approaches to coma prognostication, and highlight the immense potential of machine learning for health care problems in general.

9.1 What was accomplished

The goal of this thesis was to develop techniques for the rapid prognostication of post-anoxic coma (PAC) patients. Specific electroencephalogram (EEG) patterns are associated with eventual recovery from coma after cardiac arrest. However, existing EEG review practices rely on subjective visual analysis, which does not readily translate into reproducible quantitative predictions of neurological outcome. PAC prognostication requires analytically grounded methods that assign risks of 'good' and 'bad' neurological outcomes to reduce clinical subjectivity, and avoid poor outcomes as a result of self-fulfilling prophecies. Quantitative EEG (qEEG) methods are capable of detecting early signs of neurological recovery while coma persists. Existing qEEG methodologies are limited by small sample sizes (< 100), limited qEEG features (ten or less), and tend not to account for changes in the prognostic significance of EEG features over the course of patient recovery. We addressed each of these challenges in the thesis:

- **We collected the world largest dataset of PAC patients** from five university-affiliated hospitals between October 2009 and April 2016. The data contained over 35,000 hours of 21-channel continuous EEG recordings, a selection of clinical covariates, and an ordinal measure of neurological outcome (Glasgow-Pittsburgh Cerebral Performance categories, CPC) for 950 adult patients diagnosed with in-hospital and out-of-hospital cardiac arrests.

- **We extracted 57 quantitative EEG features** that capture three signal properties:
  - complexity: the degree of randomness in the EEG signal,
  - category: qualitative descriptors of signal characteristics or behaviors, and
  - connectivity: interactions between EEG electrodes.

- **We tested novel methods for time-sensitive classification** of patient outcomes:
  - penalized, sequential, logistic regression using 57 multi-scale features,
  - logistic regression using 10 multi-scale features, with feature dynamics constrained by a dynamic Bayesian network, and

– a variety of deep neural network architectures including: convolutional, recurrent, and feed-forward.

We found that prognostication of bad outcomes (CPC > 2) may be performed robustly by both spatio-temporal deep learning methods and feature-based approaches that account for time. The performance of our methods on unseen data was approximately 0.8 AUROC, with a true positive rate of approximately 25% at a false positive rate of 0%. Hence, our results provide strong motivation for the deployment of quantitative approaches to coma prognostication, and highlight the immense potential of machine learning for health care problem in general.

9.1.1 Future Directions

A classical dilemma of scientific investigation is the trade-off between model interpretability and model performance. This problem has become more poignant in recent years with the resurgence of neural networks, which are difficult to interpret but provide performance advantages at, or exceeding, other modeling frameworks.

The optimal balance between performance and interpretability is a function of scientific objectives. Higher performance is chosen at the expense of interpretability for problems where human accountability is less necessary, or for problems where model assessment does not require specialized knowledge (for example, the performance of a house cleaning robot or self driving car be easily assessed without knowledge of the algorithm). Conversely, higher interpretability (even at the expense of performance) is necessary for problems where accountability matters, or where model assessment requires specialized knowledge. Such is the case in the prognostication of PAC outcomes - clinicians are held accountable for care decisions, and assessment of model performance can be difficult.

Two of the three studies presented in this thesis used a linear modeling framework because of their ease of interpretation. For instance, the coefficients of the models in chapter 7 provide information that could aid in clinical prognostication, even if the model was not being used: we found that earlier measures of Shannon's entropy are more predictive of outcomes than later measures. For the sake of model reproducibility, we reported coefficients, but these

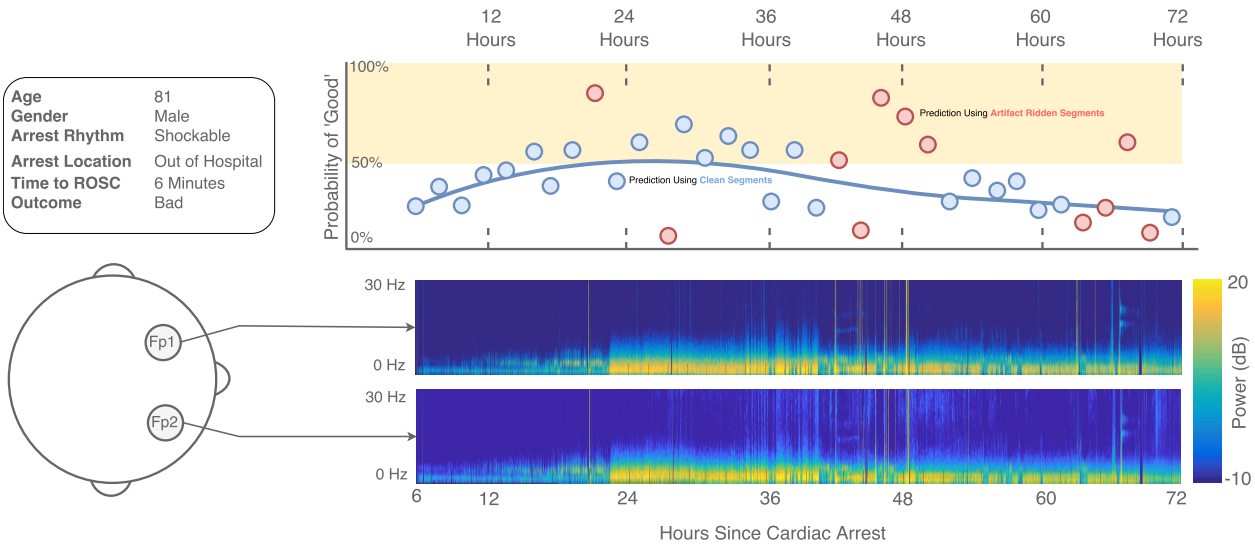

Figure 9.1: **Exemplary patient.** Performance of the model from Chapter 7 on an exemplary patient.

coefficients may also be translated into odds ratios through an exponential transform to further enhance their clinical interpretability.

Our algorithm may be used for both decision support, and risk-scoring at the bedside. In Figure 9.1 we illustrate how our method from Chapter 7 could be used for real-time prognostication using continuous EEG recordings. The figure contains information on an exemplary patient's clinical characteristics (see top left), the spectrograms from the patient's two frontal electrodes (see bottom), instantaneous predictions from our model approximately once every other hour (red and blue dots), and our proposed model's artifact adjusted prediction of outcome (blue line). We observe here that the proposed approach provides predictions that fluctuate less over time, and are more robust to estimates corrupted by artifact. We believe that a graphical user interface at the bedside, based on Figure 7.4, would be an interesting future direction of this work.

9.2 Broader Implications of the work

Machine learning (ML) and pattern recognition techniques touch practically every aspect of modern life, and and their influence will only continue to grow. From mail delivery and weather prediction, to portfolio optimization and movie recommendation, all of the

technologies that make modern life seamless are increasingly leveraging large and complex data archives to make increasingly nuanced decisions, automatically.

The theoretical advances in machine learning research stem from the work of several prominent thought leaders including Hinton, Rassmussen, Jordan, Ghahramani, Ng and others whose major thrusts of research attempt to create bigger (e.g. techniques specialized for Big Data), faster (e.g. parallelized versions of existing ML algorithms) and stronger (e.g. Deep Learning) machine learning methods.

A majority of the ML community however, are the consumers of theoretical techniques, which they apply (or modify) for their particular application areas. The most notable application areas are computer vision, natural language processing, and automated speech recognition. Each of these areas specialize in a particular modality of data (video, audio or text) and typically at a single time scale. This reality has led to relatively less work on the technical and practical challenges related to multi-modal, muti-scale data, such as health care. To further complicate matters, many of the developed techniques assume idealized conditions that are not applicable to real-world "Big" data sets (e.g., assuming minimal background noise for speaker identification, or a stationary image sizes for image recognition).

From a purely technical perspective, health-care data is interesting because it is naturally multi-modal, containing signals (such as electroencephalograms), text (such as doctors' notes), and images (such as CT scans). Health-care data also satisfy all the criteria that make data 'Big' [169]. This is important because methodologies developed for health-care may be transferable to other application areas including finance, transportation, and geophysics, to name a few.

In addition to its technical merits, there is a strong practical motivation for working on heath-care data - the industry is clearly in need of optimization. The total health expenditure in the United States is over 17% of the national Gross Domestic Product (GDP), costing a dizzying 17.5 trillion U.S. dollars annually [170]. Nowhere are the complexities and inefficiencies of health-care as obvious as the intensive care unit (ICU), which accounts for 13.4% of hospital costs and 4.1% of national health expenditures [170]. When admitted to an ICU, patients are connected to countless devices monitoring their physiology in various formats, at varying time scales (lab values may be recorded daily, while an EKG is sampled

at 200 Hz). Clinicians are then asked to incorporate this muti-modal multi-scale information in an extremely short amount of time, while accounting for noise in the data, to come up with a diagnosis and choose an optimal treatment or predict a patient's outcome. Most of this process is performed completely within the mind of the clinician, and it comes as no surprise that it is far from perfect. One recent study estimated that up to 51% [171] of ICU deaths are preventable, and there have been countless calls to develop more robust approaches to care. Fixing an entire health-care system is, of course, far too ambitious a goal for a doctoral thesis. Instead, we focused our attention on a smaller subset of critical care patients that provides a representative technical challenge of great practical importance: the rapid prognostication of post-anoxic coma outcomes.

9.3 Thoughts on EEG

9.3.1 Challenges and Opportunities of EEG

There are several practical reasons that ICUs have not adopted the regular use of EEG. It is costly, laborious, and in some cases, even impractical. The collection of surface EEG signals is made difficult by the need for a dedicated, meticulous staff that must attach sensors to patients and ensure the proper attachment of the sensors over the course of the data collection period. Even with rigorous data collection, EEG systems are notoriously prone to artifacts induced by innocuous events such as coughing, scratching the head, or eye movements. Additionally, EEG data suffer from low spatial resolution and strong cross-channel correlations.

While it is the oldest and arguably the most primitive of neuromonitoring technologies, EEG has certain advantages with regards to the aforementioned challenges. There are several existing methods and analytic toolkits that have been developed specifically to address the issues of EEG noise, and signal separation (albeit with varying levels of success) [172]. Many of the more sophisticated monitoring systems even have built-in cameras that record the physical activity of the patient and flag movements in the output time-series to help with the denoising problem.

While EEG monitoring and interpretation may be challenging, it is also the only neu-

romonitoring technology which is practical for wide-scale deployment. EEG is cheap, mobile and safe relative to other technologies. It does not require liquid helium cooled detectors (MEG), the use of a 1-ton magnet (fMRI) or radioactive isotopes (PET) to deploy. It is also significantly less bulky than most other technologies, making EEG uniquely qualified for continuous monitoring. Hence, if one can overcome the challenges inherent to the technology, EEG can serve as a valuable, and highly pragmatic tool for patients in need of neuromonitoring.

An interesting study by Abend *et al.* investigated the cost-effectiveness of EEG monitoring in critically ill-children with electrograhic status epilepticus (ESE) [173]. More specifically, the authors investigated the association between the cost of 48 hours of continuous EEG monitoring and improvements in quality-adjusted life-years gained. The authors concluded that for 48 hours of continuous EEG to be cost effective, its utilization must improve outcomes by as little as 3%. The PAC population is, of course, significantly older than the children studied by Abend *et al.*, which would make the improvement threshold for cost-effectiveness in the PAC population much higher.

9.4 Reflections on Deep Approaches

The advantage of deep networks are their ability to automatically extract pertinent representations from the data. In general, the more representative a dataset, the more reliably such networks may go about identifying meaningful representations. In the absence of unlimited data however, it is common for investigators to thoughtfully pre-process data, prior to model training. The pre-processing decisions of an investigator can have immense implications for the types of representations extracted from the data.

To emphasize the importance of pre-processing, consider the simple task of classifying images of handwritten digits in the range of 0 through 9, but drawn using red, green and blue colored pencils. In its rawest form, a single image from such data may be represented as a 3-dimensional tensor $\in \mathbb{R}^{l \times w \times c}$, where l describes the image length (in pixels), w describes image height (in pixels), and c describes image color (red, green and blue). One could use these tensors to train a neural network and, with sufficient tuning of model hyper-parameters, it is very likely that the network would perform well on a given training/validation/test

partitioning of the data. The network is likely to learn representations of the data that capture the shape of each digit. However, a problem may arise if there is a correlation between digit value and digit color. Consider if 90% of all images of the digit '6' in the data are red. One representation that the network is likely to learn is the color red. Furthermore, the network is also likely to learn that the color red is useful for predicting if a digit is a 6 or not. Of course, the color of a digit has nothing to do with its value, but the networks do not know that! They are simply looking for any set of mathematic transforms of the input data that increase their ability to optimize their performance. This problem can be solved by simply pre-processing the images to remove color information; a network cannot learn something it cannot observe.

In brief, preprocessing approaches may be understood as a *prior*, reflecting the domain knowledge of the investigator, to help ensure that learned representations will generalize beyond the dataset (i.e. removing the color of the images for the digit recognition task). So, just as conventional machine learning may be limited by the features chosen by the investigator, so too can deep learning techniques be limited by the pre-processing choices of the investigator.

9.5 Data, Science, and Data Science

With the rise in popularity of Deep Neural Networks (DNNs), statistical modeling requires greater computational resources than ever before. More computational power provides faster optimization of model hyper-parameters, which in turn provides better results, higher impact publications, media attention, and perhaps even startup fortunes. Even with the greatest of resources, advanced models can take a long time to train (hours to months, depending on the data). But in our phenomenally competitive research ecosystem, even short amounts time can make a difference between being scooped by a competitor, or being published in a premier journal or conference.

In the pursuit of better and faster results, there are practical pressures to relax model validation standards. Most researchers know and respect leave-one-out-cross validation (LOOCV) as a scientific ideal, but when models take weeks (or even months) of computation time to train, cross validation quickly becomes infeasible. We made every effort in

this thesis to grade the performance of our approaches using cross-validation approaches, often-times at the cost of the model's 'testing set performance' compared to a simple training/validation/testing split of the data.

The reality of science is this: investigators only report performance on the testing set if the results are favorable. When the results are not favorable, it is tempting to start looking for another model, or for new settings of the existing model's hyper-parameters. Next, investigators may re-evaluate the new model on the "unseen" test set, and this process is repeated until a publishable test-set performance is found. This process is *highly problematic* because it inadvertently leaks information from the testing set to the model, via the investigator. With each iteration of model development, the investigator's brain (the original neural network) gets better at choosing the model, features, (or settings of the hyper-parameters) that optimize the performance of the method on the "unseen" test-set.

This leakage of information from the testing to the training set is not done maliciously, but it is real, and problematic. It is one of the reasons that we should have a healthy skepticism of models (both our own, and those of our peers) with impressive performance that are validated using a single training/validation/testing fold [150]. Let us never forget that science is, first and foremost, about finding the truth.

References

[1] M. Tjepkema-Cloostermans, F. van Meulen, G. Meinsma, and M. van Putten, "A cerebral recovery index (cri) for early prognosis in patients after cardiac arrest," *Critical Care*, vol. 17, p. R252, 2013.

[2] M. C. Tjepkema-Cloostermans, J. Hofmeijer, A. Beishuizen, H. W. Hom, M. J. Blans, F. H. Bosch, and M. J. van Putten, "Cerebral recovery index: Reliable help for prediction of neurologic outcome after cardiac arrest.," *Critical Care Medicine*, 2017.

[3] M. Cloostermans, C. de Vos, and M. van Putten, "A novel approach for computer assisted eeg monitoring in the adult icu," *Clinical Neurophysiology*, vol. 122, no. 10, pp. 2100–2109, 2011.

[4] G. Citerio, J. Bakker, M. Bassetti, D. Benoit, M. Cecconi, J. Curtis, G. Doig, M. Herridge, S. Jaber, and M. Joannidis, "Year in review in intensive care medicine 2014: I. cardiac dysfunction and cardiac arrest, ultrasound, neurocritical care, icu-acquired weakness, nutrition, acute kidney injury, and miscellaneous," *Intensive Care Medicine*, vol. 41, no. 2, pp. 179–191, 2015.

[5] K. Kern, "Optimal treatment of patients surviving out-of-hospital cardiac arrest," *JACC: Cardiovascular Interventions*, vol. 5, no. 6, pp. 597–605, 2012.

[6] R. Deng, M. Koenig, L. Young, and X. Jia, "Early quantitative gamma-band eeg marker is associated with outcomes after cardiac arrest and targeted temperature management," *Neurocritical Care*, vol. 23, no. 2, pp. 262–273, 2015.

[7] P. R. L. I. for Health Policy, "Icu outcomes (mortality and length of stay) methods, data collection tool and data." `http://healthpolicy.ucsf.edu/content/icu-outcomes`, 2016.

[8] X. Jia, M. Koenig, R. Nickl, G. Zhen, N. Thakor, and R. Geocadin, "Early electrophysiologic markers predict functional outcome associated with temperature manipulation after cardiac arrest in rats," *Critical Care Medicine*, vol. 36, no. 6, p. 1909, 2008.

[9] A. Rossetti, M. Oddo, L. Liaudet, and P. Kaplan, "Predictors of awakening from postanoxic status epilepticus after therapeutic hypothermia," *Neurology*, vol. 72, no. 8, pp. 744–749, 2009.

[10] E. Samaniego, S. Persoon, and C. Wijman, "Prognosis after cardiac arrest and hypothermia: A new paradigm," *Current Neurology and Neuroscience Reports*, vol. 11, no. 1, pp. 111–119, 2011.

[11] E. Wijdicks, A. Hijdra, G. Young, C. Bassetti, and S. Wiebe, "Practice parameter: Prediction of outcome in comatose survivors after cardiopulmonary resuscitation (an

evidence-based review) report of the quality standards subcommittee of the american academy of neurology," *Neurology*, vol. 67, no. 2, pp. 203–210, 2006.

[12] A. Roest, B. van Bets, P. Jorens, I. Baar, J. Weyler, and R. Mercelis, "The prognostic value of the eeg in postanoxic coma," *Neurocritical Care*, vol. 10, no. 3, pp. 318–325, 2009.

[13] A. Rossetti, "Prognostic utility of electroencephalogram in acute consciousness impairment," in *Clinical Neurophysiology in Disorders of Consciousness*, pp. 55–71, Springer, 2015.

[14] N. Blondin and D. Greer, "Neurologic prognosis in cardiac arrest patients treated with therapeutic hypothermia," *The Neurologist*, vol. 17, no. 5, pp. 241–248, 2011.

[15] J. Claassen, F. Taccone, P. Horn, M. Holtkamp, N. Stocchetti, and M. Oddo, "Recommendations on the use of eeg monitoring in critically ill patients: Consensus statement from the neurointensive care section of the esicm," *Intensive Care Medicine*, vol. 39, no. 8, pp. 1337–1351, 2013.

[16] S. Herman, N. Abend, T. Bleck, K. Chapman, F. Drislane, R. Emerson, E. Gerard, C. Hahn, A. Husain, and P. Kaplan, "Consensus statement on continuous eeg in critically ill adults and children, part i: Indications," *Journal of Clinical Neurophysiology*, vol. 32, no. 2, pp. 87–95, 2015.

[17] P. Nunez and R. Srinivasan, *Electric Fields of The Brain: The Neurophysics of EEG*. Oxford University Press, 2006.

[18] Headway, "Effects of hypoxic/anoxic brain injury." `https://www.headway.org.uk/about-brain-injury/individuals/types-of-brain-injury/hypoxic-and-anoxic-brain-injury/anoxic-brain-injury-effects/`, 2018.

[19] P. Stammet, D. Wagner, G. Gilson, and Y. Devaux, "Modeling serum level of s100b and bispectral index to predict outcome after cardiac arrest," *Journal of the American College of Cardiology*, 2013.

[20] A. A. of Neurology, "Compelling statistics." `https://www.aan.com/uploadedFiles/Website_Library_Assets/Documents/6.Public_Policy/1.Stay_Informed/4.Public_Policy_Resources/compell.pdf`, 2016.

[21] A. Park and J. Boyd, "Eeg utilization in the medical/surgical icu: A single centre prospective observational study," *Intensive Care Medicine*, vol. 41, no. 10, pp. 1869–1870, 2015.

[22] C. H. Jeon, J. S. Park, J. H. Lee, H. Kim, S. C. Kim, K. H. Park, K. S. Yi, S. M. Kim, C. S. Youn, Y.-M. Kim, *et al.*, "Comparison of brain computed tomography and diffusion-weighted magnetic resonance imaging to predict early neurologic outcome before target temperature management comatose cardiac arrest survivors," *Resuscitation*, vol. 118, pp. 21–26, 2017.

[23] P. Zanatta, F. Linassi, A. P. Mazzarolo, M. Aricò, E. Bosco, M. Bendini, C. Sorbara, C. Ori, M. Carron, and B. Scarpa, "Pain-related somato sensory evoked potentials: a potential new tool to improve the prognostic prediction of coma after cardiac arrest," *Critical Care*, vol. 19, no. 1, p. 403, 2015.

[24] D. Vondrakova, A. Kruger, M. Janotka, F. Malek, V. Dudkova, P. Neuzil, and P. Ostadal, "Association of neuron-specific enolase values with outcomes in cardiac arrest survivors is dependent on the time of sample collection," *Critical Care*, vol. 21, no. 1, p. 172, 2017.

[25] T. Suys, P. Bouzat, P. Marques-Vidal, N. Sala, J. Payen, A. Rossetti, and M. Oddo, "Automated quantitative pupillometry for the prognostication of coma after cardiac arrest," *Neurocritical Care*, vol. 21, no. 2, pp. 300–308, 2014.

[26] P. Stammet, O. Collignon, C. Hassager, M. P. Wise, J. Hovdenes, A. Åneman, J. Horn, Y. Devaux, D. Erlinge, J. Kjaergaard, *et al.*, "Neuron-specific enolase as a predictor of death or poor neurological outcome after out-of-hospital cardiac arrest and targeted temperature management at 33 c and 36 c," *Journal of the American College of Cardiology*, vol. 65, no. 19, pp. 2104–2114, 2015.

[27] E. Efthymiou, R. Renzel, C. R. Baumann, R. Poryazova, and L. L. Imbach, "Predictive value of eeg in postanoxic encephalopathy: A quantitative model-based approach," *Resuscitation*, vol. 119, pp. 27–32, 2017.

[28] N. V. Thakor and S. Tong, "Advances in quantitative electroencephalogram analysis methods," *Annual Review of Biomedical Engineering*, vol. 6, pp. 453–495, 2004.

[29] D. Subha, P. Joseph, R. Acharya, and C. Lim, "Eeg signal analysis: A survey," *Journal of Medical Systems*, vol. 34, no. 2, pp. 195–212, 2010.

[30] Q. Noirhomme, R. Lehembre, Z. d. R. Lugo, D. Lesenfants, A. Luxen, S. Laureys, M. Oddo, and A. O. Rossetti, "Automated analysis of background eeg and reactivity during therapeutic hypothermia in comatose patients after cardiac arrest," *Clinical EEG and neuroscience*, vol. 45, no. 1, pp. 6–13, 2014.

[31] M. Rundgren, I. Rosén, and H. Friberg, "Amplitude-integrated eeg (aeeg) predicts outcome after cardiac arrest and induced hypothermia," *Intensive Care Medicine*, vol. 32, no. 6, pp. 836–842, 2006.

[32] B. Ruijter, M. Putten, and J. Hofmeijer, "Generalized epileptiform discharges in postanoxic encephalopathy: Quantitative characterization in relation to outcome," *Epilepsia*, 2015.

[33] R. Homan, J. Herman, and P. Purdy, "Cerebral location of international 10–20 system electrode placement," *Electroencephalography and clinical Neurophysiology*, vol. 66, no. 4, pp. 376–382, 1987.

[34] M. Nuwer, "Assessment of digital eeg, quantitative eeg, and eeg brain mapping: Report of the american academy of neurology and the american clinical neurophysiology society*," *Neurology*, vol. 49, no. 1, pp. 277–292, 1997.

[35] R. Schwartz, E. Brown, R. Lydic, and N. Schiff, "General anesthesia, sleep, and coma," *New England Journal of Medicine*, vol. 363, no. 27, pp. 2638–2650, 2010.

[36] P. Ktonas, S. Golemati, H. Tsekou, T. Paparrigopoulos, C. Soldatos, P. Xanthopoulos, V. Sakkalis, M. Zervakis, and M. Ortigueira, "Potential dementia biomarkers based on the time-varying microstructure of sleep eeg spindles," in *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pp. 2464–2467, IEEE, 2007.

[37] E. Niedermeyer and F. da Silva, *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Lippincott Williams & Wilkins, 2005.

[38] D. Zhang, X. Jia, H. Ding, D. Ye, and N. Thakor, "Application of tsallis entropy to eeg: Quantifying the presence of burst suppression after asphyxial cardiac arrest in rats," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 4, pp. 867–874, 2010.

[39] R. Geocadin, R. Ghodadra, T. Kimura, H. Lei, D. Sherman, D. Hanley, and N. Thakor, "A novel quantitative eeg injury measure of global cerebral ischemia," *Clinical Neurophysiology*, vol. 111, no. 10, pp. 1779–1787, 2000.

[40] M. Koenig, P. Kaplan, and N. Thakor, "Clinical neurophysiologic monitoring and brain injury from cardiac arrest," *Neurologic Clinics*, vol. 24, no. 1, pp. 89–106, 2006.

[41] K.-A. Hossmann and B. G. Ophoff, "Recovery of monkey brain after prolonged ischemia. i. electrophysiology and brain electrolytes," *Journal of Cerebral Blood Flow & Metabolism*, vol. 6, no. 1, pp. 15–21, 1986.

[42] H. Lukatch and M. MacIver, "Synaptic mechanisms of thiopental-induced alterations in synchronized cortical activity.," *Anesthesiology*, vol. 84, no. 6, pp. 1425–1434, 1996.

[43] G. Visser, G. Wieneke, and A. Van Huffelen, "Carotid endarterectomy monitoring: Patterns of spectral eeg changes due to carotid artery clamping," *Clinical Neurophysiology*, vol. 110, no. 2, pp. 286–294, 1999.

[44] E. Jørgensen and A. Malchow-Møller, "Natural history of global and critical brain ischaemia part i: Eeg and neurological signs during the first year after cardiopulmonary resuscitation in patients subsequently regaining consciousness," *Resuscitation*, vol. 9, no. 2, pp. 133–153, 1981.

[45] C. Sandroni, A. Cariou, F. Cavallaro, T. Cronberg, H. Friberg, C. Hoedemaekers, J. Horn, J. Nolan, A. Rossetti, and J. Soar, "Prognostication in comatose survivors of cardiac arrest: An advisory statement from the european resuscitation council and the european society of intensive care medicine," *Resuscitation*, vol. 85, no. 12, pp. 1779–1789, 2014.

[46] H. Friberg, M. Rundgren, E. Westhall, N. Nielsen, and T. Cronberg, "Continuous evaluation of neurological prognosis after cardiac arrest," *Acta Anaesthesiologica Scandinavica*, vol. 57, no. 1, pp. 6–15, 2013.

[47] M. Karapetkova, M. Koenig, and X. Jia, "Early prognostication markers in cardiac arrest patients treated with hypothermia," *European Journal of Neurology*, 2015.

[48] J. Borjigin, U. Lee, T. Liu, D. Pal, S. Huff, D. Klarr, J. Sloboda, J. Hernandez, M. Wang, and G. Mashour, "Surge of neurophysiological coherence and connectivity in the dying brain," *Proceedings of the National Academy of Sciences*, vol. 110, no. 35, pp. 14432–14437, 2013.

[49] R. Geocadin, J. Muthuswamy, D. Sherman, N. Thakor, and D. Hanley, "Early electrophysiological and histologic changes after global cerebral ischemia in rats," *Movement Disorders*, vol. 15, no. S1, pp. 14–21, 2000.

[50] M. Koenig and P. Kaplan, "Clinical applications for eps in the icu," *Journal of Clinical Neurophysiology*, vol. 32, no. 6, pp. 472–480, 2015.

[51] J. Claassen, L. Hirsch, K. Kreiter, E. Du, E. Connolly, R. Emerson, and S. Mayer, "Quantitative continuous eeg for detecting delayed cerebral ischemia in patients with poor-grade subarachnoid hemorrhage," *Clinical Neurophysiology*, vol. 115, no. 12, pp. 2699–2710, 2004.

[52] J. Leon-Carrion, J. Martin-Rodriguez, J. Damas-Lopez, J. Martin, and M. Dominguez-Morales, "Delta–alpha ratio correlates with level of recovery after neurorehabilitation in patients with acquired brain injury," *Clinical Neurophysiology*, vol. 120, no. 6, pp. 1039–1045, 2009.

[53] M. van Putten, M. Tjepkema-Cloostermans, and J. Hofmeijer, "Infraslow eeg activity modulates cortical excitability in postanoxic encephalopathy," *Journal of Neurophysiology*, vol. 113, no. 9, pp. 3256–3267, 2015.

[54] L. Colgin, "Mechanisms and functions of theta rhythms," *Annual Review of Neuroscience*, no. 0, 2013.

[55] S. Bresnahan, J. Anderson, and R. Barry, "Age-related changes in quantitative eeg in attention-deficit/hyperactivity disorder," *Biological Psychiatry*, vol. 46, no. 12, pp. 1690–1697, 1999.

[56] J. Carrier, S. Land, D. Buysse, D. J. Kupfer, and T. Monk, "The effects of age and gender on sleep eeg power spectral density in the middle years of life (ages 20–60 years old)," *Psychophysiology*, vol. 38, no. 2, pp. 232–242, 2001.

[57] P. Safar, A. Bleyaert, E. Nemoto, J. Moossy, and J. Snyder, "Resuscitation after global brain ischemia-anoxia.," *Critical Care Medicine*, vol. 6, no. 4, pp. 215–227, 1978.

[58] R. Hickey, P. Kochanek, H. Ferimer, H. Alexander, R. Garman, and S. Graham, "Induced hyperthermia exacerbates neurologic neuronal histologic damage after asphyxial cardiac arrest in rats," *Critical Care Medicine*, vol. 31, no. 2, pp. 531–535, 2003.

[59] H. Choi, N. Badjatia, and S. Mayer, "Hypothermia for acute brain injury—mechanisms and practical aspects," *Nature Reviews Neurology*, vol. 8, no. 4, pp. 214–222, 2012.

[60] T. Deboer, "Brain temperature dependent changes in the electroencephalogram power spectrum of humans and animals," *Journal of Sleep Research*, vol. 7, no. 4, pp. 254–262, 1998.

[61] V. Vanston, M. Lawhon-Triano, R. Getts, J. Prior, and R. Smego Jr, "Predictors of poor neurologic outcome in patients undergoing therapeutic hypothermia after cardiac arrest," *Southern Medical Journal*, vol. 103, no. 4, pp. 301–306, 2010.

[62] J. Hofmeijer, M. Tjepkema-Cloostermans, M. Blans, A. Beishuizen, and M. van Putten, "Unstandardized treatment of electroencephalographic status epilepticus does not improve outcome of comatose patients after cardiac arrest," *Frontiers In Neurology*, vol. 5, 2014.

[63] A. Crepeau, A. Rabinstein, J. Fugate, J. Mandrekar, E. Wijdicks, R. White, and J. Britton, "Continuous eeg in therapeutic hypothermia after cardiac arrest prognostic and clinical value," *Neurology*, vol. 80, no. 4, pp. 339–344, 2013.

[64] T. Schnider, C. Minto, S. Shafer, P. Gambus, C. Andresen, D. Goodale, and E. Youngs, "The influence of age on propofol pharmacodynamics," *Anesthesiology*, vol. 90, no. 6, pp. 1502–1516, 1999.

[65] D. Trickey, "Encephalography." `http://www.amcresidents.com/lectures/Dr_Trickey/EEG.htm`, 2004.

[66] E. Brown, P. Purdon, and C. Van Dort, "General anesthesia and altered states of arousal: A systems neuroscience analysis," *Annual Review Of Neuroscience*, vol. 34, p. 601, 2011.

[67] L. Becker, T. Aufderheide, R. Geocadin, C. Callaway, R. Lazar, M. Donnino, V. Nadkarni, B. Abella, C. Adrie, and R. Berg, "Primary outcomes for resuscitation science studies a consensus statement from the american heart association," *Circulation*, vol. 124, no. 19, pp. 2158–2177, 2011.

[68] R. Phelps, F. Dumas, C. Maynard, J. Silver, and T. Rea, "Cerebral performance category and long-term prognosis following out-of-hospital cardiac arrest," *Critical Care Medicine*, vol. 41, no. 5, pp. 1252–1257, 2013.

[69] E. Edgren, S. Kelsey, K. Sutton, and P. Safar, "The presenting ecg pattern in survivors of cardiac arrest and its relation to the subsequent long-term survival," *Acta Anaesthesiologica Scandinavica*, vol. 33, no. 4, pp. 265–271, 1989.

[70] X. Jia, M. Koenig, H. Shin, G. Zhen, C. Pardo, D. Hanley, N. Thakor, and R. Geocadin, "Improving neurological outcomes post-cardiac arrest in a rat model: Immediate hypothermia and quantitative eeg monitoring," *Resuscitation*, vol. 76, no. 3, pp. 431–442, 2008.

[71] J. Madhok, A. Maybhate, W. Xiong, M. Koenig, R. Geocadin, X. Jia, and N. Thakor, "Quantitative assessment of somatosensory-evoked potentials after cardiac arrest in rats: Prognostication of functional outcomes," *Critical Care Medicine*, vol. 38, no. 8, p. 1709, 2010.

[72] L. Katz, U. Ebmeyer, P. Safar, A. Radovsky, and R. Neumar, "Outcome model of asphyxial cardiac arrest in rats," *Journal of Cerebral Blood Flow & Metabolism*, vol. 15, no. 6, pp. 1032–1039, 1995.

[73] D. L. Sherman, A. M. Brambrink, R. N. Ichord, V. K. Dasika, R. C. Koehler, R. J. Traystman, D. F. Hanley, and N. V. Thakor, "Quantitative eeg during early recovery from hypoxic-ischemic injury in immature piglets: Burst occurrence and duration," *Clinical EEG And Neuroscience*, vol. 30, no. 4, pp. 175–183, 1999.

[74] F. Xiao, P. Safar, and A. Radovsky, "Mild protective and resuscitative hypothermia for asphyxial cardiac arrest in rats," *The American Journal of Emergency Medicine*, vol. 16, no. 1, pp. 17–25, 1998.

[75] A. Zeiner, M. Holzer, F. Sterz, W. Behringer, W. Schörkhuber, M. Müllner, M. Frass, P. Siostrzonek, K. Ratheiser, and A. Kaff, "Mild resuscitative hypothermia to improve neurological outcome after cardiac arrest a clinical feasibility trial," *Stroke*, vol. 31, no. 1, pp. 86–94, 2000.

[76] C. E. Shannon and W. Weaver, *The mathematical theory of communication*. University of Illinois press, 1998.

[77] H.-C. Shin, X. Jia, R. Nickl, R. G. Geocadin, N. V. Thakor, *et al.*, "A subband-based information measure of eeg during brain injury and recovery after cardiac arrest," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 8, pp. 1985–1990, 2008.

[78] A. V. Oppenheim and R. W. Schafer, "From frequency to quefrency: A history of the cepstrum," *IEEE signal processing Magazine*, vol. 21, no. 5, pp. 95–106, 2004.

[79] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, "Determining lyapunov exponents from a time series," *Physica D: Nonlinear Phenomena*, vol. 16, no. 3, pp. 285–317, 1985.

[80] A. Accardo, M. Affinito, M. Carrozzi, and F. Bouquet, "Use of the fractal dimension for the analysis of electroencephalographic time series," *Biological cybernetics*, vol. 77, no. 5, pp. 339–350, 1997.

[81] S.-H. Oh, Y.-R. Lee, and H.-N. Kim, "A novel eeg feature extraction method using hjorth parameter," *International Journal of Electronics and Electrical Engineering*, vol. 2, no. 2, pp. 106–110, 2014.

[82] R. Hegger and H. Kantz, "Improved false nearest neighbor method to detect determinism in time series data," *Physical Review E*, vol. 60, no. 4, p. 4970, 1999.

[83] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[84] S. M. Ross, *Introductory statistics*. Academic Press, 2017.

[85] J. M. Stern, *Atlas of EEG patterns*. Lippincott Williams & Wilkins, 2005.

[86] J. Shaw, "An introduction to the coherence function and its use in eeg signal analysis," *Journal of medical engineering & technology*, vol. 5, no. 6, pp. 279–288, 1981.

[87] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Mutual information-based selection of optimal spatial–temporal patterns for single-trial eeg-based bcis," *Pattern Recognition*, vol. 45, no. 6, pp. 2137–2144, 2012.

[88] K. J. Blinowska, R. Kuś, and M. Kamiński, "Granger causality and information flow in multivariate processes," *Physical Review E*, vol. 70, no. 5, p. 050902, 2004.

[89] C. Stam, G. Nolte, and A. Daffertshofer, "Phase lag index: Assessment of functional connectivity from multi channel eeg and meg with diminished bias from common sources," *Human Brain Mapping*, vol. 28, no. 11, pp. 1178–1193, 2007.

[90] S. M. Kay, *Fundamentals of statistical signal processing*. Prentice Hall PTR, 1993.

[91] T. Cover and J. Thomas, *Elements Of Information Theory*. John Wiley & Sons, 2012.

[92] A. Bezerianos, S. Tong, and N. Thakor, "Time-dependent entropy estimation of eeg rhythm changes following brain ischemia," *Annals of Biomedical Engineering*, vol. 31, no. 2, pp. 221–232, 2003.

[93] H. Shin, S. Tong, S. Yamashita, X. Jia, G. Geocadin, and N. Thakor, "Quantitative eeg and effect of hypothermia on brain recovery after cardiac arrest," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 6, pp. 1016–1023, 2006.

[94] D. Childers, D. Skinner, and R. Kemerait, "The cepstrum: A guide to processing," *Proceedings Of The IEEE*, vol. 65, no. 10, pp. 1428–1443, 1977.

[95] N. Güler, E. Übeyli, and İ. Güler, "Recurrent neural networks employing lyapunov exponents for eeg signals classification," *Expert Systems With Applications*, vol. 29, no. 3, pp. 506–514, 2005.

[96] E. Pereda, A. Gamundi, R. Rial, and J. González, "Non-linear behaviour of human eeg: Fractal exponent versus correlation dimension in awake and sleep stages," *Neuroscience Letters*, vol. 250, no. 2, pp. 91–94, 1998.

[97] K. Blinowska and M. Malinowski, "Non-linear and linear forecasting of the eeg time series," *Biological Cybernetics*, vol. 66, no. 2, pp. 159–165, 1991.

[98] W. Pritchard, D. Duke, and K. Krieble, "Dimensional analysis of resting human eeg ii: Surrogate-data testing indicates nonlinearity but not low-dimensional chaos," *Psychophysiology*, vol. 32, no. 5, pp. 486–491, 1995.

[99] J. Schoffelen and J. Gross, "Source connectivity analysis with meg and eeg," *Human Brain Mapping*, vol. 30, no. 6, pp. 1857–1865, 2009.

[100] P. Uhlhaas and W. Singer, "Abnormal neural oscillations and synchrony in schizophrenia," *Nature Reviews Neuroscience*, vol. 11, no. 2, pp. 100–113, 2010.

[101] J. Hipp, A. Engel, and M. Siegel, "Oscillatory synchronization in large-scale cortical networks predicts perception," *Neuron*, vol. 69, no. 2, pp. 387–396, 2011.

[102] M. Beudel, M. Tjepkema-Cloostermans, J. Boersma, and M. van Putten, "Small-world characteristics of eeg patterns in post-anoxic encephalopathy," *Frontiers in Neurology*, vol. 5, 2014.

[103] R. Vigário, J. Särelä, V. Jousmiki, M. Hämäläinen, and E. Oja, "Independent component approach to the analysis of eeg and meg recordings," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 5, pp. 589–593, 2000.

[104] D. Y. Ko, "Epileptiform discharges," *Medscape*, pp. http://emedicine.medscape.com/article/1138880–overview, 2016.

[105] M. Cloostermans, F. van Meulen, C. Eertman, H. Hom, and M. van Putten, "Continuous electroencephalography monitoring for early prediction of neurological outcome in postanoxic patients after cardiac arrest: A prospective cohort study*," *Critical Care Medicine*, vol. 40, no. 10, pp. 2867–2875, 2012.

[106] A. Sivaraju, E. Gilmore, C. Wira, A. Stevens, N. Rampal, J. Moeller, D. Greer, L. Hirsch, and N. Gaspard, "Prognostication of post-cardiac arrest coma: Early clinical and electroencephalographic predictors of outcome," *Intensive Care Medicine*, pp. 1–9, 2015.

[107] E. Donnelly and A. Blum, "Focal and generalized slowing, coma, and brain death," in *The Clinical Neurophysiology Primer*, pp. 127–140, Springer, 2007.

[108] A. Ribeiro, R. Singh, and F. Brunnhuber, "Clinical outcome of generalized periodic epileptiform discharges on first eeg in patients with hypoxic encephalopathy postcardiac arrest," *Epilepsy & Behavior*, vol. 49, pp. 268–272, 2015.

[109] M. Koutroumanidis and D. Sakellariou, "Low frequency nonevolving generalized periodic epileptiform discharges and the borderland of hypoxic nonconvulsive status epilepticus in comatose patients after cardiac arrest," *Epilepsy & Behavior*, vol. 49, pp. 255–262, 2015.

[110] J. Lucas, M. Cocchi, J. Salciccioli, J. Stanbridge, R. Geocadin, S. Herman, and M. Donnino, "Neurologic recovery after therapeutic hypothermia in patients with post-cardiac arrest myoclonus," *Resuscitation*, vol. 83, no. 2, pp. 265–269, 2012.

[111] T. Cronberg, "Should postanoxic status epilepticus be treated agressively? yes!," *Journal of clinical Neurophysiology: Official Publication of The American Electroencephalographic Society*, vol. 32, no. 6, pp. 449–451, 2015.

[112] A. Rossetti, "Should postanoxic status epilepticus be treated aggressively?—no!," *Journal of Clinical Neurophysiology*, vol. 32, no. 6, pp. 447–448, 2015.

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

[114] A. Beydoun, C. Yen, and I. Drury, "Variance of interburst intervals in burst suppression," *Electroencephalography And Clinical Neurophysiology*, vol. 79, no. 6, pp. 435–439, 1991.

[115] N. Schaul, "The fundamental neural mechanisms of electroencephalography," *Electroencephalography And Clinical Neurophysiology*, vol. 106, no. 2, pp. 101–107, 1998.

[116] M. Rundgren, E. Westhall, T. Cronberg, I. Rosén, and H. Friberg, "Continuous amplitude-integrated electroencephalogram predicts outcome in hypothermia-treated cardiac arrest patients," *Critical Care Medicine*, vol. 38, no. 9, pp. 1838–1844, 2010.

[117] J. Hofmeijer, M. Tjepkema-Cloostermans, and M. van Putten, "Burst-suppression with identical bursts: A distinct eeg pattern with poor outcome in postanoxic coma," *Clinical Neurophysiology*, vol. 125, no. 5, pp. 947–954, 2014.

[118] M. Monteiro, F. Taccone, C. Depondt, I. Lamanna, N. Gaspard, N. Ligot, N. Mavroudakis, G. Naeije, J. Vincent, and B. Legros, "The prognostic value of 48-h continuous eeg during therapeutic hypothermia after cardiac arrest," *Neurocritical Care*, pp. 1–10, 2015.

[119] F. Zubler, C. Koenig, A. Steimer, S. Jakob, K. Schindler, and H. Gast, "Prognostic and diagnostic value of eeg signal coupling measures in coma," *Clinical Neurophysiology*, 2015.

[120] M. Rundgren, T. Karlsson, N. Nielsen, T. Cronberg, P. Johnsson, and H. Friberg, "Neuron specific enolase and s-100b as predictors of outcome after cardiac arrest and induced hypothermia," *Resuscitation*, vol. 80, no. 7, pp. 784–789, 2009.

[121] T. Zellner, R. Gärtner, J. Schopohl, and M. Angstwurm, "Nse and s-100b are not sufficiently predictive of neurologic outcome after therapeutic hypothermia for cardiac arrest," *Resuscitation*, vol. 84, no. 10, pp. 1382–1386, 2013.

[122] M. Leary, D. Fried, D. Gaieski, R. Merchant, B. Fuchs, D. Kolansky, D. Edelson, and B. Abella, "Neurologic prognostication and bispectral index monitoring after resuscitation from cardiac arrest," *Resuscitation*, vol. 81, no. 9, pp. 1133–1137, 2010.

[123] Q. Yang, Y. Su, M. Hussain, W. Chen, H. Ye, D. Gao, and F. Tian, "Poor outcome prediction by burst suppression ratio in adults with post-anoxic coma without hypothermia," *Neurological Research*, vol. 36, no. 5, pp. 453–460, 2014.

[124] T. Oksanen, M. Tiainen, M. B. Skrifvars, T. Varpula, A. Kuitunen, M. Castrén, and V. Pettilä, "Predictive power of serum nse and ohca score regarding 6-month neurologic outcome after out-of-hospital ventricular fibrillation and therapeutic hypothermia," *Resuscitation*, vol. 80, no. 2, pp. 165–170, 2009.

[125] J. Schefold, C. Storm, A. Krüger, C. Ploner, and D. Hasper, "The glasgow coma score is a predictor of good outcome in cardiac arrest patients treated with therapeutic hypothermia," *Resuscitation*, vol. 80, no. 6, pp. 658–661, 2009.

[126] W. Schoerkhuber, H. Kittler, F. Sterz, W. Behringer, M. Holzer, M. Frossard, S. Spitzauer, and A. Laggner, "Time course of serum neuron-specific enolase a predictor of neurological outcome in patients resuscitated from cardiac arrest," *Stroke*, vol. 30, no. 8, pp. 1598–1603, 1999.

[127] V. Alvarez, C. Reinsberger, B. Scirica, M. O'Brien, K. Avery, G. Henderson, and J. Lee, "Continuous electrodermal activity as a potential novel neurophysiological biomarker of prognosis after cardiac arrest–a pilot study," *Resuscitation*, vol. 93, pp. 128–135, 2015.

[128] S. K. Kessler, A. A. Topjian, A. M. Gutierrez-Colina, R. N. Ichord, M. Donnelly, V. M. Nadkarni, R. A. Berg, D. J. Dlugos, R. R. Clancy, and N. S. Abend, "Short-term outcome prediction by electroencephalographic features in children treated with therapeutic hypothermia after cardiac arrest," *Neurocritical care*, vol. 14, no. 1, pp. 37–43, 2011.

[129] C. Storm, J. Nee, A. Jörres, C. Leithner, D. Hasper, and C. Ploner, "Serial measurement of neuron specific enolase improves prognostication in cardiac arrest patients treated with hypothermia: A prospective study," *Scandanavian Journal of Trauma Resuscitation and Emergency Medicine*, vol. 20, no. 6, 2012.

[130] A. O. Rossetti, M. Oddo, G. Logroscino, and P. W. Kaplan, "Prognostication after cardiac arrest and hypothermia: a prospective study," *Annals of neurology*, vol. 67, no. 3, pp. 301–307, 2010.

[131] C. Adrie, A. Cariou, B. Mourvillier, I. Laurent, H. Dabbane, F. Hantala, A. Rhaoui, M. Thuong, and M. Monchi, "Predicting survival with good neurological recovery at hospital admission after successful resuscitation of out-of-hospital cardiac arrest: The ohca score," *European Heart Journal*, vol. 27, no. 23, pp. 2840–2845, 2006.

[132] I. G. Steffen, D. Hasper, C. J. Ploner, J. C. Schefold, E. Dietz, F. Martens, J. Nee, A. Krueger, A. Jörres, and C. Storm, "Mild therapeutic hypothermia alters neuron specific enolase as an outcome predictor after resuscitation: 97 prospective hypothermia patients compared to 133 historical non-hypothermia patients," *Critical care*, vol. 14, no. 2, p. R69, 2010.

[133] M. Oddo and A. Rossetti, "Early multimodal outcome prediction after cardiac arrest in patients treated with hypothermia.," *Critical Care Medicine*, 2014.

[134] J. Hofmeijer, T. Beernink, F. Bosch, A. Beishuizen, M. Tjepkema-Cloostermans, and M. van Putten, "Early eeg contributes to multimodal outcome prediction of postanoxic coma," *Neurology*, vol. 85, no. 2, pp. 137–143, 2015.

[135] D. Seder, G. Fraser, T. Robbins, L. Libby, and R. Riker, "The bispectral index and suppression ratio are very early predictors of neurological outcome during therapeutic hypothermia after cardiac arrest," *Intensive Care Medicine*, vol. 36, no. 2, pp. 281–288, 2010.

[136] C. Bassetti, F. Bomio, J. Mathis, and C. W. Hess, "Early prognosis in coma after cardiac arrest: a prospective clinical, electrophysiological, and biochemical study of 60 patients.," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 61, no. 6, pp. 610–615, 1996.

[137] C. Roger, L. Palmier, B. Louart, N. Molinari, P. Claret, J. de la Coussaye, J. Lefrant, and L. Muller, "Neuron specific enolase and glasgow motor score remain useful tools for assessing neurological prognosis after out-of-hospital cardiac arrest treated with therapeutic hypothermia," *Anaesthesia Critical Care & Pain Medicine*, vol. 34, no. 4, pp. 231–237, 2015.

[138] J. Reisinger, K. Höllinger, W. Lang, C. Steiner, T. Winter, E. Zeindlhofer, M. Mori, A. Schiller, A. Lindorfer, K. Wiesinger, *et al.*, "Prediction of neurological outcome after cardiopulmonary resuscitation by serial determination of serum neuron-specific enolase," *European heart journal*, vol. 28, no. 1, pp. 52–58, 2007.

[139] C. M. Booth, R. H. Boone, G. Tomlinson, and A. S. Detsky, "Is this patient dead, vegetative, or severely neurologically impaired?: assessing outcome for comatose survivors of cardiac arrest," *Jama*, vol. 291, no. 7, pp. 870–879, 2004.

[140] C. Youn, C. Callaway, and J. Rittenberger, "Combination of initial neurologic examination and continuous eeg to predict survival after cardiac arrest," *Resuscitation*, vol. 94, pp. 73–79, 2015.

[141] N. Bigdely-Shamlo, T. Mullen, C. Kothe, K.-M. Su, and K. A. Robbins, "The prep pipeline: standardized preprocessing for large-scale eeg analysis," *Frontiers in neuroinformatics*, vol. 9, p. 16, 2015.

[142] A. Delorme, S. Makeig, T. Jung, and T. Sejnowski, "Automatic rejection of event-related potential trials and components using independent component analysis," in *Society for Neuroscience Abstracts*, vol. 27, 2001.

[143] M. S. Grewal, "Kalman filtering," in *International Encyclopedia of Statistical Science*, pp. 705–708, Springer, 2011.

[144] Z. Li, J. E. O'Doherty, T. L. Hanson, M. A. Lebedev, C. S. Henriquez, and M. A. Nicolelis, "Unscented kalman filter for brain-machine interfaces," *PloS one*, vol. 4, no. 7, p. e6243, 2009.

[145] M. Rohál'ová, P. Sykacek, M. Koskaand, and G. Dorffner, "Detection of the eeg artifacts by the means of the (extended) kalman filter," *Meas. Sci. Rev*, vol. 1, no. 1, pp. 59–62, 2001.

[146] M. J. Barton, P. A. Robinson, S. Kumar, A. Galka, H. F. Durrant-Whyte, J. Guivant, and T. Ozaki, "Evaluating the performance of kalman-filter-based eeg source localization," *IEEE transactions on biomedical engineering*, vol. 56, no. 1, pp. 122–136, 2009.

[147] N. R. Cook and P. M. Ridker, "The use and magnitude of reclassification measures for individual predictors of global cardiovascular risk," *Annals of internal medicine*, vol. 150, no. 11, p. 795, 2009.

[148] L. Beretta and A. Santaniello, "Nearest neighbor imputation algorithms: a critical evaluation," *BMC medical informatics and decision making*, vol. 16, no. 3, p. 74, 2016.

[149] M. M. Ghassemi, "Features for eeg analysis." `https://github.com/deskool/ComaPrognosticanUsingEEG`, 2017.

[150] M. M. Ghassemi and T. Alhanai, "How to cook your results (without even knowing it)." `https://github.com/deskool/Cooking_Data`, 2017.

[151] O. Gosseries, C. Schnakers, D. Ledoux, A. Vanhaudenhuyse, M.-A. Bruno, A. Demertzi, Q. Noirhomme, R. Lehembre, P. Damas, S. Goldman, *et al.*, "Automated eeg entropy measurements in coma, vegetative state/unresponsive wakefulness syndrome and minimally conscious state," *Functional neurology*, vol. 26, no. 1, p. 25, 2011.

[152] J. E. Wennervirta, M. J. Ermes, S. M. Tiainen, T. K. Salmi, M. S. Hynninen, M. O. Särkelä, M. J. Hynynen, U.-H. Stenman, H. E. Viertiö-Oja, K.-P. Saastamoinen, *et al.*, "Hypothermia-treated cardiac arrest patients with good neurological outcome differ early in quantitative variables of eeg suppression and epileptiform activity," *Critical care medicine*, vol. 37, no. 8, pp. 2427–2435, 2009.

[153] J. Elmer, J. J. Gianakas, J. C. Rittenberger, M. E. Baldwin, J. Faro, C. Plummer, L. A. Shutter, C. L. Wassel, C. W. Callaway, A. Fabio, *et al.*, "Group-based trajectory modeling of suppression ratio after cardiac arrest," *Neurocritical care*, vol. 25, no. 3, pp. 415–423, 2016.

[154] D. B. Seder, K. Sunde, S. Rubertsson, M. Mooney, P. Stammet, R. R. Riker, K. B. Kern, B. Unger, T. Cronberg, J. Dziodzio, *et al.*, "Neurologic outcomes and postresuscitation care of patients with myoclonus following cardiac arrest," *Critical care medicine*, vol. 43, no. 5, pp. 965–972, 2015.

[155] E. Amorim, J. C. Rittenberger, J. J. Zheng, M. B. Westover, M. E. Baldwin, C. W. Callaway, A. Popescu, *et al.*, "Continuous eeg monitoring enhances multimodal outcome prediction in hypoxic–ischemic brain injury," *Resuscitation*, vol. 109, pp. 121–126, 2016.

[156] E. Amorim, J. C. Rittenberger, M. E. Baldwin, C. W. Callaway, A. Popescu, and P. C. A. Service, "Malignant eeg patterns in cardiac arrest patients treated with targeted temperature management who survive to hospital discharge," *Resuscitation*, vol. 90, pp. 127–132, 2015.

[157] A. O. Rossetti, L. A. Urbano, F. Delodder, P. W. Kaplan, and M. Oddo, "Prognostic value of continuous eeg monitoring during therapeutic hypothermia after cardiac arrest," *Critical care*, vol. 14, no. 5, p. R173, 2010.

[158] J. E. Fugate, E. F. Wijdicks, J. Mandrekar, D. O. Claassen, E. M. Manno, R. D. White, M. R. Bell, and A. A. Rabinstein, "Predictors of neurologic outcome in hypothermia after cardiac arrest," *Annals of neurology*, vol. 68, no. 6, pp. 907–914, 2010.

[159] B. J. Ruijter, J. Hofmeijer, H. G. E. Meijer, and M. J. A. M. van Putten, "Synaptic damage underlies eeg abnormalities in postanoxic encephalopathy: A computational study," *Clinical neurophysiology*, vol. 128, no. 9, pp. 1682–1695, 2017.

[160] A. Anglemyer, H. T. Horvath, and L. Bero, "Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials," *The Cochrane Library*, 2014.

[161] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[162] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[163] J. Donahue, L. A. Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell, "Long-term recurrent convolutional networks for visual recognition and description," in *CVPR*, 2015.

[164] P. Bashivan, I. Rish, M. Yeasin, and N. Codella, "Learning representations from eeg with deep recurrent-convolutional neural networks," *arXiv preprint arXiv:1511.06448*, 2015.

[165] L. Deng, J. Li, J.-T. Huang, K. Yao, D. Yu, F. Seide, M. Seltzer, G. Zweig, X. He, J. Williams, *et al.*, "Recent advances in deep learning for speech research at microsoft," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pp. 8604–8608, IEEE, 2013.

[166] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," *arXiv preprint arXiv:1706.04599*, 2017.

[167] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," in *International Conference on Artificial Neural Networks*, pp. 52–59, Springer, 2011.

[168] M. B. Christopher, *PATTERN RECOGNITION AND MACHINE LEARNING.* Springer-Verlag New York, 2016.

[169] M. Coakley, M. Leerkes, J. Barnett, A. Gabrielian, K. Noble, M. Weber, and Y. Huyen, "Unlocking the power of big data at the national institutes of health," *Big Data*, 2013.

[170] N. Halpern and S. Pastores, "Critical care medicine in the united states 2000–2005: An analysis of bed numbers, occupancy rates, payer mix, and costs*," *Critical Care Medicine*, vol. 38, no. 1, pp. 65–71, 2010.

[171] H. Goulet, V. Guerand, B. Bloom, P. Martel, P. Aegerter, E. Casalino, B. Riou, and Y. Freund, "Unexpected death within 72 hours of emergency department visit: Were those deaths preventable?," *Critical Care*, vol. 19, no. 1, p. 154, 2015.

[172] A. Delorme and S. Makeig, "Eeglab: An open source toolbox for analysis of single-trial eeg dynamics including independent component analysis," *Journal of Neuroscience Methods*, vol. 134, no. 1, pp. 9–21, 2004.

[173] N. Abend, A. Topjian, and S. Williams, "Could eeg monitoring in critically ill children be a cost-effective neuroprotective strategy?," *Journal of Clinical Neurophysiology*, vol. 32, no. 6, pp. 486–494, 2015.