Characterizing Uncertainty in Sensor Fusion to Improve Predictive Models

---

A Dissertation

Presented to

the faculty of the School of Engineering and Applied Science

University of Virginia

---

in partial fulfillment
of the requirements for the degree

Doctor of Philosophy

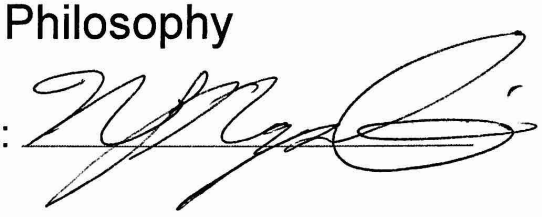
by

Nicholas J. Napoli

May 2018

APPROVAL SHEET

This Dissertation
is submitted in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

Author Signature: 

This Dissertation has been read and approved by the examining committee:

Advisor: Dr. Laura E. Barnes

Committee Member: Dr. William T. Scherer

Committee Member: Dr. Scott Acton

Committee Member: Dr. Gerard Learmonth

Committee Member: Dr. Laura E. Barnes

Committee Member: Dr. Angela R. Harrivel

Accepted for the School of Engineering and Applied Science:

Craig H. Benson, School of Engineering and Applied Science

May 2018

Abstract

Uncertainty is inevitably introduced into machine learning paradigms and threatens the validity and robustness of class prediction. This uncertainty can be introduced at the model level or from the utilized features that are input into the model. This dissertation presents multiple methodologies for evaluating and reducing uncertainty to improve model performance. The work is divided into two sections, which address two types of machine learning problems: handling uncertainty for signature detection (i.e., template matching paradigms) and handling uncertainty for predictive problems for systems with diverse information sources. This work utilizes model fusion methodologies to build upon previously developed machine learning methods to improve the predictive power. The contribution in Part I is a theoretical framework that quantifies and handles model uncertainty assignments allowing multiple sources of information to be fused together in order to improve detection and scalability of temporal and spatial signatures. The contribution in Part II is novel feature engineering techniques and a model fusion approach for improving the downstream effects of uncertainty. Overall, the focus of this work is on the various avenues of how uncertainty enters systems. Through different methods of handling uncertainty, we demonstrate that we can improve the performance of predictive models.

Dedications

This is dedicated to people who have entered into my life that have shaped me into who I am today! My mother who has always been there for me and has provided me the tools to accomplish my goals and dreams. My grandfather, Joseph Fiero, who has made into the man I am today. My brother who has guided me and has watched over me. I love you. Additionally, Alex Vallejo, Alex Castro, and Avtar Singh three influential people when I was an undergraduate that taught me how to believe in myself.

Acknowledgments

My graduate doctoral experience has been a long enduring continual test of being able to push myself and the research boundary. There are several people to whom I owe tremendous gratitude for achieving my doctorate.

I would like to especially thank my immediate adviser, Dr. Laura E. Barnes for creating and providing the numerous fellowship, collaborations, and grant opportunities for me conduct my research. Moreover, her flexibility in providing me the time and ability to explore my research interests has allowed the development of these novel designs to push these research boundaries. Thank you!

To the NASA Langley Research Center's Crew Systems Operations Team, more specifically Chad L. Stephens, Kellie D. Kennedy, Angela Harrivel, Brenda L. Crook, Randall E. Bailey, and Alan T. Pope, thank you for the last three years of working with you that have created an unforgettable experience. I hope our late night phone calls and dinner research conversations will continue as my research career continues to development. Your expertise in experimental design, humans factors, psychophysiology and biomedical engineering has been invaluable for my dissertation and personal development.

To my two pivotal graduate student colleagues and officemates, Matt Demas (a.k.a David) and Kevin Leach, thank you being wonderful co-authors, friends, and collaborators.

To the three people from my past, Ranga Dabarera, Anthony Mixco, and Kamal Premaratne, that continue to always be there for me to discuss technical problems with when I am in a jam.

To my girlfriend, Madeline Kotoriy, for her unwavering support for me to be able to push through my dissertation in the darkest and more stressful hours. Furthermore, thank you for your brilliant edits, health care research collaborations, and simply just being a great listener. I will always love you, no matter what!

Contents

Abstract i

Dedications ii

Acknowledgments iii

Contents vii

1 Introduction 1

I Handling Uncertainty for Signature Detection 4

2 Template Matching: Accounting for Uncertainty in Selecting the Winner 7
- 2.1 Introduction . . . 8
- 2.2 Basic Notions of DS Theory . . . 10
- 2.3 Proposed DST Model . . . 12
- 2.4 Evaluation and Discussion . . . 20
  - 2.4.1 RQ1: Generalized Template Match Model Validation . . . 20
  - 2.4.2 RQ2: Correlation Coefficients as a a form of Evidence . . . 26
- 2.5 Conclusion . . . 29

3 DS MapReduce Paradigm 32
- 3.1 Background . . . 35
  - 3.1.1 MapReduce . . . 35
  - 3.1.2 Dempster-Shafer Theory Background and Notation . . . 36
- 3.2 MapReduce Framework for Template Fusion . . . 38
  - 3.2.1 Source Dataset . . . 38
  - 3.2.2 Correlation to DS Framework: Map . . . 39
  - 3.2.3 Set of Correlations Coefficients to DS Masses . . . 41
  - 3.2.4 DCR Fusion: TreeReduce . . . 42
- 3.3 Architecture and Implementation . . . 42
  - 3.3.1 Testing Methodology . . . 45
  - 3.3.2 MapReduce Implementation . . . 45
  - 3.3.3 Pthreads Implementation . . . 46
- 3.4 Evaluation and Discussion . . . 47

3.4.1 RQ1: Scalability . . . 48
3.4.2 RQ2: Robustness against noise . . . 49
3.4.3 RQ3: Reducing uncertainty with more sources . . . 50
3.5 Conclusion . . . 52

4 Handling Time Series Data: ECG Corruption 54
4.1 Introduction . . . 54
4.2 Methods . . . 58
4.2.1 Template Framework. . . . 58
4.2.2 Templates to DS Masses . . . 60
4.3 Evaluation and Discussion . . . 63
4.3.1 RQ1: Capturing belief and uncertainty . . . 65
4.3.2 RQ2: Robustness against noise . . . 66
4.3.3 RQ3: Little Prior Knowledge . . . 68
4.4 Conclusion . . . 68

II Handling Uncertainty for Predictive Problems for Systems with Diverse Information Sources 70

5 Evaluating Sensor Uncertainty due to Noise 73
5.1 Introduction . . . 74
5.2 Methods . . . 80
5.2.1 Monte Carlo Simulation: Addressing RQ1 to RQ3 . . . 80
5.2.2 Sympathetic Response Dataset: Addressing RQ4 to RQ6 . . . 84
5.3 Evaluation and Discussion . . . 86
5.3.1 RQ1: Fiducial R-peak Marker, Type of Noise, and SNR . . . 87
5.3.2 RQ2: Type of Entropy and Precision . . . 87
5.3.3 RQ3: Noisy Time Segments and HRC Variance . . . 89
5.3.4 RQ4: Insight HRC Sensitivity . . . 92
5.3.5 RQ5: What is the Directional Change in Entropy . . . 93
5.3.6 RQ6: Indistinguishable Independent Distributions . . . 93
5.4 Conclusion . . . 96

6 Featuring Engineering to Provide Additional Model Dimensionality 98
6.1 Introduction . . . 99
6.2 Methods . . . 108
6.2.1 Hypoxia Data Set . . . 109
6.2.2 Filter Bank Intensity Method . . . 110
6.2.3 Activation Complexity: . . . 121
6.2.4 Rank Order Complexity: . . . 122
6.3 Evaluation and Discussion . . . 124
6.3.1 RQ1: Simulated Continuous EEG Intensity Measurement . . . 125
6.3.2 RQ2: Activation Complexity . . . 128
6.3.3 RQ3: Rank Order Complexity . . . 129

6.3.4 RQ4: Classical Methods 133
6.4 Conclusions 138

7 Naive Adaptive Probabilistic Sensor Fusion **140**
7.1 Introduction 141
7.2 Methods 146
7.2.1 Experimental Design and Data Collection 146
7.2.2 Naive Adaptive Probabilistic Sensor (NAPS) Fusion 149
7.3 Evaluations and Discussion 157
7.3.1 RQ1: Dimensionality Reduction for Multi-Class Problems 158
7.3.2 RQ2: Dimensionality Reduction Combined with Augmented Response Variables 161
7.3.3 RQ3: NAPS Fusion Approach 162
7.4 Conclusion 166

8 Conclusion **167**

Life isn't always black and white. You're not always sure you're right. At least I'm not.

Brad Paisley

1

Introduction

In practice, when a decision is to be rendered through machine learning (ML) methods, there are two main types of uncertainties one may encounter: aleatory uncertainty or epistemic uncertainty. Aleatory uncertainty refers to a system behaving in a "random" way and represents unknowns that differ each time the system is observed. This type of uncertainty is not reducible, but it can be accounted for and modeled using historic data via probability distribution models. Epistemic uncertainty is caused by a lack of knowledge which can be reduced through increased understanding. One way of increasing understanding around specific hypotheses is through obtaining more information in the form of new features or sensor

modalities. Thus, ML algorithms can be improved by reducing epistemic uncertainty and appropriately handling aleatory uncertainty, which, in turn, helps ensure a robust system design.

The work in this dissertation discusses uncertainty in two different ML architectures. Part I of the dissertation focuses on detecting signatures using basic matching algorithms. This work is applied to images and time series data that are highly corrupted, and then addresses the scalability of the paradigm in a MapReduce framework. Part II of the dissertation examines more complicated systems that handle a multitude of features where we apply typical ML paradigms to cross-sectional data sets. This work discusses the development of bio-signal processing methods for featuring engineering, a paradigm to quantify uncertainty in the designed ML model over numerous sensors and approaches to handle low samples of data that contain class imbalances. Each of the two partitions in this body of work consists of three chapters, where each chapter is independent of each other and addresses overall arching themes for handling uncertainty in either signature detection (Part I) or complicated, predictive system design (Part II). Each chapter is its own body of work and is structured to be published by itself.

Each partition has developed an augmented uncertainty theoretical framework that is used in conjunction with a prior ML methodology. The application of the uncertainty framework is stacked upon the ML framework. ML Stacking is a generalized terminology that refers to use of combining different predictive models for improved accuracy. The stacked uncertainty framework is developed around Dempster-Shafer (DS) Theory which is a method that accounts for epistemic uncertainty and combines evidence to reduce uncertainty while achieving a more accurate decision. However, in the each body of evidence (BoE) within the DS framework uncertainty is distributed to the set of propositions. Thus, properly characterizing uncertainty within the BoE is pivotal to improve the overall predictive framework. Likewise neglecting or improperly characterizing the uncertainty can lead to poor performing

algorithms. Thus, we must appropriately reduce and address uncertainty to ensure successful behavior of the system. This enables the paradigm to handle conflicting information and essentially weigh the various sources of information differently in order to support the set propositions (i.e., hypotheses). In addition, these bodies of work extend beyond the developed theory, which address uncertainty in other facets, practical applications of the framework, and the scalability of the paradigms.

Part I

Handling Uncertainty for Signature Detection

Part I of the this dissertation strictly discusses the detection of temporal and spacial signatures, typically referred to as template matching or motifs. Part I is partitioned into three conference proceeding papers, each making up a chapter (i.e., Chapters 2, 3, and 4). Each individual chapter demonstrates specific contributions, key-points, and conclusions. Chapter 2, “Template Matching: Accounting for Uncertainty in Selecting the Winner," is a theoretical paper that discusses template matching and how an uncertainty framework is assigned to sets of templates. Chapter 3, “D.S. MapReduce Paradigm," demonstrates how this framework can be implemented for distributed computing to improve computational speed and reduce uncertainty on large fused data sets. The last chapter in Part I, “Handling Time Series Data: ECG Corruption," demonstrates the framework for a practical time series application - electrocardiograms - which highlights how utilizing multiple fused sources of information reduces uncertainty and increases predictive power.

Although this work was designed and implemented using correlation coefficients (a feature depicting the degree of a match between templates), it does not require the use of correlation values. The theoretical framework only requires an input that is bounded and normalized, thus methods such as Normalized Match Filtering and Normalized Dynamic Time Warping would be applicable for this framework as well. Although one of the easiest methods to implement, correlation is a rudimentary method and more sophisticated methods may be better. Thus the correlation coefficients one produces are not always the best to use for detection problems and provides, at times, a high amount of variance and conflict between matches. However, the main point of this work is not to use correlation coefficients for detection but rather to exemplify how the quantified uncertainty paradigm can capture this uncertainty (i.e., variance) and conflict among disputed winners. The overall points to note within this compilation are 1) how the theoretic framework for quantifying uncertainty provides additional contextual meaning (i.e. belief, uncertainty and plausibly) for detecting the winning template; 2) how fused information from additional sources provides lift in detection performance; and 3) how the various use-cases can be augmented for future work.

When one admits that nothing is certain one must, I think, also add that some things are more nearly certain than others.

Bertrand Russell

2

Template Matching: Accounting for Uncertainty in Selecting the Winner

The problem of selecting a template that matches a given candidate signal is applicable across a wide variety of domains (e.g. character recognition, radar and sonar). Using the correlation coefficient as the avenue for selecting the winning (or most highly-matching) template is perhaps the most common technique. The challenge lies in selecting the winning template when there is no clear separation between the correlation coefficient values of the winning template and the others. In this paper, we present a simple Dempster-Shafer (DS)

theoretic model that enables one to capture the uncertainty regarding the winner selection in correlation-coefficient-based template matching. The DS theoretic framework provides an avenue to develop the model with few resources and little to no prior knowledge. We validate the model using several numerical examples and a numerical character recognition application where the evidence provided by several sets of templates are combined using a DS theoretic fusion strategy to arrive at a better decision.

2.1 Introduction

The correlation coefficient is widely used as a measure that captures the strength of the linear relationship between two variables. When applied to random variables, the correlation coefficient between the random variables $\mathbf{X}$ and $\mathbf{Y}$ usually takes the form [95]

$$r_{\mathbf{X},\mathbf{Y}} = \frac{\mathrm{cov}(\mathbf{X},\mathbf{Y})}{\sigma_{\mathbf{X}}\sigma_{\mathbf{Y}}} = \frac{E[(\mathbf{X}-\overline{\mathbf{X}})(\mathbf{Y}-\overline{\mathbf{Y}})}{\sigma_{\mathbf{X}}\sigma_{\mathbf{Y}}}, \tag{2.1}$$

where $E[.]$ denotes the expectation, $\overline{\mathbf{X}} = E[\mathbf{X}]$ and $\overline{\mathbf{Y}} = E[\mathbf{Y}]$ denote the means, and $\sigma_{\mathbf{X}} = E[\mathbf{X}^2] - E^2[\mathbf{X}]$ and $\sigma_{\mathbf{Y}} = E[\mathbf{X}^2] - E^2[\mathbf{X}]$ denote the standard deviations. The correlation coefficient takes values in $[-1,+1]$: values closer to $+1$ indicate a strong positive relationship; values closer to -1 indicate a strong negative relationship; and values closer to 0 indicate weak a or non-existent relationship.

The correlation coefficient is also widely employed to gauge the 'closeness' or similarity between two data vectors, in which case it usually takes the form [107]

$$r_{\mathbf{x},\mathbf{y}} = \frac{\sum_{i=1}^{n}(x_i-\overline{\mathbf{x}})(y_i-\overline{\mathbf{y}})}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{\mathbf{x}})^2\sum_{j=1}^{n}(y_j-\overline{\mathbf{y}})^2}}, \tag{2.2}$$

where $\overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\overline{\mathbf{y}} = \frac{1}{n}\sum_{j=1}^{n} y_j$ denote the 'sample' means of the real-valued time series data vectors $\mathbf{x} = [x_1,\ldots,x_N]^T$ and $\mathbf{y} = [y_1,\ldots,y_N]^T$, respectively.

Prior Work. In practice, one is often interested in determining how well a *template* signal vector belonging to a set of templates matches a given *candidate* signal. The template which yields the highest correlation coefficient with the candidate can be considered the 'winner' thus implying that the candidate indicates the presence of the winning template [4, 8, 24, 47, 81]. Over the years, this simple and popular method based on correlation coefficients has also been employed when there is a *set* of candidate signals to be considered [34, 82, 86, 102].

Challenges. However, this technique of selecting the winning template may not be satisfactory when multiple templates yield high correlation coefficients with no significant separation of values between the correlation coefficient of one template and the others.

For instance, two templates which yield correlation coefficient values that are close to each other creates an uncertainty regarding which template should be declared the winner. Approaches that attempt to deal with such situations tend to employ various weighting and voting strategies to select the winning candidate [55, 82, 102]. Another source of difficulty in selecting the winner is the presence of noise. For example, within the context of reconstructing physiological signals, the work in [55] employs additional leads to mitigate problems caused by sources corrupted from noise. Fuzzy reasoning has been suggested as a way to improve the detection process in such situations [34].

These methods however do not provide a satisfactory solution for capturing the uncertainty associated with assigning a template to the candidate signal. To account for noisy signals, multiple templates, and multiple candidates, researchers have also embraced the use of more elaborate Kalman filtering and other statistical approaches [96].

Contributions. How should we handle such uncertainties in selecting the winning template? While probability theory is perhaps the most widely used approach for representing and handling imperfect information, probabilistic methods usually call for a priori assumptions regarding the underlying distributions and priors for handling data uncertainties. On the other hand, alternate imprecise probability formalisms, such as the Demspter-Shafer (DS)

belief theoretic approach, provide ways to represent and deal with data uncertainties while requiring only little more information than voting and set intersection techniques [20,63,135, 149]. DS theoretic (DST) methods can represent a wider variety of data imperfections in a more intuitive manner [20, 63]; they are more robust to modeling errors [15, 140], and when compared to what alternate frameworks provide, the DST belief and plausibility measures enable a decision to be made with a better understanding of the associated uncertainties [160]. When there is no uncertainty regarding the underlying distribution, these DST measures equal probability thus allowing seamless integration of DST methods with probabilisitic methods [20, 63, 135, 160, 162]. This is a unique feature of the DST framework.

In this paper, we present a simple DST model which can be used to capture the uncertainty associated with assigning a template to the candidate signal. This model, which is only slightly more complicated than a probabilistic model, is based on the correlation coefficients between the candidate and the templates, and it allows us to view the candidate as an evidence source which indicates the 'presence' or 'absence' of each template. One may then utilize a DST evidence fusion strategy to combine multiple evidence sources, where each source provides evidence towards the 'presence' or 'absence' of each template, to make a decision regarding the winning template. Moreover, the availability of DST measures of belief and plausibility provide valuable information regarding the confidence one can place on this decision.

2.2 Basic Notions of DS Theory

The *frame of discernment (FoD)* Θ refers to the set of mutually exclusive and exhaustive propositions of interest. We take the FoD to be $\Theta = \{\theta_1, \ldots, \theta_n\}$, i.e., Θ is finite and composed of n *singleton* propositions. The power set of Θ is denoted by 2^{Θ}. A *basic probability assignment (BPA),* otherwise referred to as a *mass function,* is a function m :

$2^{\Theta} \rightarrow [0,1]$ such that

$$m(\emptyset) = 0; \quad \sum_{A \subseteq \Theta} m(A) = 1. \tag{2.3}$$

Contextual considerations (e.g., accuracy, source reliability, source conflicts, etc.) all play a role in determining the mass to be allocated to a given proposition [133]. A proposition which has been allocated a non-zero mass is referred to as a *focal element.* The *core* $\mathfrak{F}$ refers to the set of focal elements and the *body of evidence (BoE)* $\mathcal{E}$ refers to the triplet $\{\Theta, \mathfrak{F}, m(.)\}$.

By allowing the allocation of masses directly to non-singleton or *composite* propositions, DS theory provides an avenue to capture *ignorance* and *uncertainty.* While $m(A)$ measures the support that is directly assigned to proposition $A \subseteq \Theta$ *only,* the *belief* $Bl(A)$ represents the total support that can move into A without any ambiguity; *plausibility* $Pl(A)$ represents the extent to which one finds A plausible. So,

$$Bl(A) = \sum_{B \subseteq A} m(B); \quad Pl(A) = \sum_{B \cap A \neq \emptyset} m(B). \tag{2.4}$$

These DST belief and plausibility measures are closely related to the inner and outer measures of a non-measurable event $A \subseteq \Theta$ with respect to probability mass functions (p.m.f.s) defined on Θ. Furthermore, when focal elements are constituted of singletons only, the mass, belief and plausibility all reduce to a p.m.f. The *uncertainty interval* $Un(A) = [Bl(A), Pl(A)]$ provides information regarding the support for $A \subseteq \Theta$.

Dempster's combination rule (DCR) allows one to combine or fuse evidence represented as DST models [135]:

$$m(A) = \frac{\sum_{B \cap C = A} m_1(B)\, m_2(C)}{1 - \sum_{B \cap C = \emptyset} m_1(B)\, m_2(C)}, \tag{2.5}$$

where $\mathcal{E}_1 = \{\Theta, \mathfrak{F}_1, m_1(.)\}$ and $\mathcal{E}_2 = \{\Theta, \mathfrak{F}_2, m_2(.)\}$ are the BoE being fused to generate the fused BoE $\mathcal{E} = \{\Theta, \mathfrak{F}, m(.)\}$. The fused mass and BoE generated by the DCR are usually

denoted by $m = m_1 \oplus m_2$ and $\mathcal{E} = \mathcal{E}_1 \oplus \mathcal{E}_2$, respectively. The DCR is commutative and associative, thus allowing one to fuse multiple sources of evidence with ease.

2.3 Proposed DST Model

Consider the correlation vector corresponding to a single candidate vector and a set of N templates $\{T_1, \ldots, T_N\}$:

$$\mathbf{v} = \begin{bmatrix} v_1 & v_2 & \cdots & v_N \end{bmatrix}^T, \tag{2.6}$$

where $v_i \geq 0$, $i = 1, \ldots, N$, denotes the positive correlation coefficient between the candidate vector and the template T_i. In what appears below, we will deal with only non-negative values of correlation coefficients. In most applications, all the information regarding the identity of the template is captured by the absolute value of the correlation coefficient. In some application contexts, a negative correlation coefficient usually implies the absence of the corresponding template and treating such coefficients as having a value of zero would force the corresponding DST mass to be zero as well. On the other hand, if negative correlation coefficients provide additional information regarding the identity of the template, one can easily modify the algorithm we present below to account for negative coefficients values.

In constructing a DST model to represent the correlation coefficient vector $\mathbf{v}$ in (3.4), we must ensure that the model captures the potential conflicts among templates that are 'competing' for a match with the candidate signal reasonably well. For example, a correlation vector with multiple entries having a value of $+1$ would indicate that multiple templates are perfect matches for the candidate.

Weighting Matrix. First, consider the following $(N \times N)$-sized matrix $\Delta\mathbf{V} = \{\Delta V_{ij}\}$:

$$\Delta\mathbf{V} = \{\Delta V_{ij}\}, \text{ where } \Delta V_{ij} = v_i - v_j. \tag{2.7}$$

Note that $\Delta V_{ij} \in [-1, +1]$, $\forall i, j = 1, \ldots, N$, can be considered the 'distance' between the correlation coefficients for templates T_i and T_j; the sum of the entries in the j-th column of $\Delta\mathbf{V}$ is the 'sum of distances' $\sum_{i=1}^{N} \Delta V_{ij} = \sum_{i=1}^{N}(v_i - v_j)$ from the correlation coefficient of template T_j.

However, the entries in the $\Delta\mathbf{V}$ matrix still lack information about the 'strength' of the correlations. For example, take $\mathbf{v} = [0.4, 0.3, 0.7, 0.6]^T$. This yields $(v_1 - v_2) = (v_3 - v_4) = 0.1$, irrespective of the fact that the values v_3 and v_4 are significantly higher (thus more correlated with the corresponding two templates) than the values v_1 and v_2 (which are less correlated). To account for this, we utilize the following weighting strategy:

$$\Delta\mathbf{W} = \{\Delta W_{ij}\} = \Delta\mathbf{v} \cdot \mathbf{D}_{\mathbf{V}}, \tag{2.8}$$

where $\mathbf{D}_{\mathbf{v}} = \text{diag}\{v_1, \ldots, v_N\}$ denotes the diagonal matrix with the diagonal entries $\{v_1, \ldots, v_N\}$. Column j of $\Delta\mathbf{W}$ is simply column j of $\Delta\mathbf{V}$ weighted by the correlation coefficient v_j corresponding to template T_j. Taking the same example $\mathbf{v} = [0.4, 0.3, 0.7, 0.6]^T$, we get $v_1\Delta V_{12} = (0.4)(0.4 - 0.3) = 0.04$ and $v_3\Delta V_{34} = (0.7)(0.7 - 0.6) = 0.07$, which now accounts for the strength of the correlation coefficients.

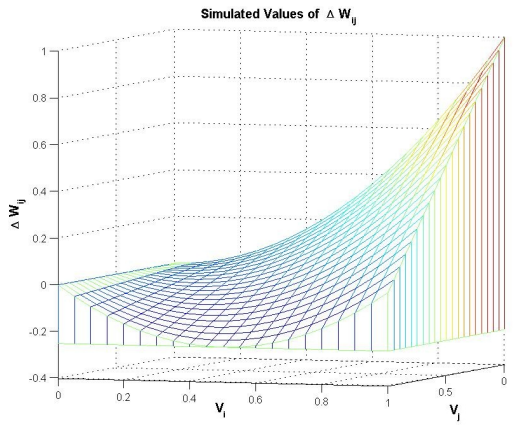


Figure 2.1: Plot of $\Delta W_{ij} = v_i\,\Delta V_{ij}$ versus (v_i, v_j), for $v_i, v_j \in [0, 1]$.

Note that each column of $\Delta\mathbf{W}$ evaluates the correlation coefficient corresponding to a specific

template against the correlation coefficients of the entire set of templates. For example, column j of $\Delta\mathbf{W}$ captures the strength of the correlation coefficient v_j and the distance between v_j and all other correlation coefficients $(v_i - v_j)$. In essence, the matrix $\Delta\mathbf{W} = \{\Delta W_{ij}\}$ accounts for both the *strength* of the correlation to each template and also the *distance* between the correlations of pairs of templates. When one or both aspects are low, the corresponding entries in $\Delta\mathbf{W}$ will take on lower values. This feature of $\Delta\mathbf{W}$ is important when attempting to fit our DST model. Fig. 2.1 demonstrates the mapping $\Delta W_{ij} = v_i\,\Delta V_{ij}$ versus (v_i, v_j). Note that the maximum and minimum achievable values of ΔW_{ij} are $+1$ and $-1/4$, respectively; and these values are achieved when $(v_i, v_j) = (1, 0)$ and $(v_i, v_j) = (1/2, 1)$, respectively. Note how ΔW_{ij} rewards higher strength of correlation coefficients and larger distances of pairs of templates.

Column Weights Vector. The sum of column j of $\Delta\mathbf{W}$ informs us about how well the template T_j matches the candidate and how different the correlation coefficient associated with T_j is when compared to the other correlation coefficients. We refer to the vector created by the sum of the entries in each column of the weighting matrix $\Delta\mathbf{W}$ as its *column weight vector* $\mathbf{c}$*:*

$$\mathbf{c} = \begin{bmatrix} c_1 & \cdots & c_N \end{bmatrix} = \mathbf{1}_{1\times N}\Delta\mathbf{W}, \tag{2.9}$$

where $\mathbf{1}_{1\times N} = [1,\ 1,\ \ldots,\ 1]$, i.e., the $(1\times N)$-sized row vector with all entries taking the value 1. Note that the diagonal entries of $\Delta\mathbf{W}$ are always 0. Hence, $c_i \in [-(N-1)/4, (N-1)]$, $\forall\ i \in \overline{1, N}$. These column weights allow us to identify the rival templates that are competing to be a match for the candidate. Let us take some examples to demonstrate this.

Mass Measure Vector. Non-positive entries of $\mathbf{c}$ indicate templates that cannot compete for being a match to the candidate vector. So, such entries are replaced with values of 0 to get the *mass measure vector* $\mathbf{h}$*:*

$$\mathbf{h} = [h_1,\ h_2\ \cdots,\ h_N],\ \text{where } h_i = \frac{c_i + |c_i|}{2}. \tag{2.10}$$

Clearly, $h_i \in [0, (N-1)]$, $\forall i \in \overline{1, N}$. Note that, $\mathbf{h}$ is identical to $\mathbf{c}$, except that it substitutes 0 for all the non-positive elements of $\mathbf{c}$. We will later use the entries of $\mathbf{h}$ to generate the DST masses associated with the templates T_j. For convenience, we use the notation

$$S_{\mathbf{v}} = \sum_{i=1}^{N} v_i; \quad S_{\mathbf{h}} = \sum_{i=1}^{N} h_i. \tag{2.11}$$

As it turns out, $S_{\mathbf{h}}$ denotes the total of DST masses that will be assigned to the singletons $\{T_i\}$.

Table 2.1: Some Examples

v	**c**	**h**
(1) $[1, 0, 0, 0]^T$	[3, 0, 0, 0]	[3, 0, 0, 0]
(2) $[1, 1, 0, 0]^T$	[2, 2, 0, 0]	[2, 2, 0, 0]
(3) $[1, 1, 0.9, 0]^T$	[1.10, 1.10, 0.63, 0]	[1.10, 1.10, 0.63, 0]
(4) $[1, 1, 0.9, 0.1]^T$	[1, 1, 0.54, -0.26]	[1, 1, 0.54, 0]

Examples. Table 2.1 show some examples of how $\mathbf{h}$ can be used to capture those templates that can be a match for the candidate. Note the following:

Example (1): Template T_1 is the only match, and $\mathbf{h}$ puts the maximum weight on T_1.

Example (2): Templates T_1 and T_2 are both competing for being a match, and $\mathbf{h}$ puts equal weights to T_1 and T_2. This weight is however less than what T_1 gets in Example (1) because of the presence of potentially two matching templates.

Example (3): Templates T_1-T_3 are competing for being a match, with template T_3 being the slightly weaker, and $\mathbf{h}$ distributes its weights among T_1-T_3.

Example (4): All templates have positive correlation coefficients, but only T_1-T_3 are legitimate potential matches. Note how $\mathbf{h}$ discards T_4, but puts less values for T_1-T_3 because $c_4 < 0$. ■

'Extreme' Case. Examples 1-2 above are instances of an 'extreme' situation when the correlation coefficients of all the templates take on values 1 or 0 only, i.e., $v_i \in \{0, 1\}$, $\forall i \in \overline{1, N}$.

We refer to this case as the *extreme case.* For this case, without loss of generality, we may assume that the templates yielding a perfect match appear as the top elements of $\mathbf{v}$, i.e.,

$$\mathbf{v} = [\overbrace{1, \ldots, 1}^{P \text{ of } 1\text{s}}, \overbrace{0, \ldots, 0}^{(N-P) \text{ of } 0\text{s}}]^T. \tag{2.12}$$

This yields

$$\begin{aligned} \Delta\mathbf{W} &= \begin{bmatrix} \emptyset_{P\times P} & \emptyset_{P\times(N-P)} \\ \mathbf{1}_{(N-P)\times P} & \emptyset_{(N-P)\times(N-P)} \end{bmatrix} \\ \mathbf{c} = \mathbf{h} &= [\overbrace{(N-P), \ldots, (N-P)}^{P \text{ of } (N-P)\text{s}}, 0, \ldots, 0]. \end{aligned} \tag{2.13}$$

Here $\emptyset_{K\times L}$ and $\mathbf{1}_{K\times L}$ denotes the $(K \times L)$-sized matrices with entries 0 and 1, respectively. Note that, $S_{\mathbf{h}} = P(N-P)$.

Focal Elements. The FoD of the proposed DST model is the set of N templates, i.e., $\Theta = \{T_1, \ldots, T_N\}$. We use the column weights vector $\mathbf{c}$ to identify the focal elements of our DST model. First, we make the following observations:

(a) When $v_i \to 1$ and $v_j \to 0$, $\forall j \neq i$, the column weights $c_i \to (N-1)$ and $c_j \to 0$, $\forall j \neq i$, i.e., the uncertainty regarding the template T_i being a match decreases (see Example 1 above). Conversely, the higher the difference between $(N-1)$ and a column weight, the less likely the corresponding template is the correct match. When the column weight is zero or negative, we become more certain that the corresponding template is not a match.

(b) When a template is unable to compete against the other templates, the corresponding column weight becomes zero or negative (see Example 4 above). So, in our model, we neglect the templates corresponding to non-positive values of c_i, thus preventing them from becoming focal elements in our DST model. This strategy restricts the domain of candidate templates that are potential focal elements thus avoiding having to assign DST masses to

'weak' candidates.

The number of non-zero entries of $\mathbf{h}$ (or equivalently, the number of positive entries of $\mathbf{c}$) determine the number P of singleton focal elements of our DST model:

$$P = \sum_{i=1}^{N} t_i, \text{ where } t_i = \begin{cases} 1, & \textit{if } h_i > 0; \\ 0, & \textit{if } h_i = 0. \end{cases} \tag{2.14}$$

DST Mass Allocation. We will develop the model first for the extreme case, and then extend it to the more general case. The focal elements of the proposed DST model are

$$\mathfrak{F} = \{T_1, T_2, \ldots, T_P, \Theta\}. \tag{2.15}$$

We will assume that $P \in \overline{1, N}$; $P = 0$ case implies complete lack of information, and we trivially assign the DST model $m(\Theta) = 1$.

Extreme Case. a) Masses Allocated to Singletons $m(T_i)$: From the examples in Table 2.1, we notice how the values h_i appear to weigh templates according to how well they match the candidate. With this observation in mind, we propose

$$m(T_i) \propto h_i \implies m(T_i) = \frac{h_i}{K} = \frac{N - P}{K}, \tag{2.16}$$

where $K > 0$ is the proportionality constant.

b) Mass Allocated to Complete Ignorance $m(\Theta)$: Since the masses allocated to all the focal elements must add to 1, we must have

$$m(\Theta) = 1 - \sum_{i=1}^{P} m(T_i) = 1 - \frac{S_{\mathbf{h}}}{K} = 1 - \frac{P(N - P)}{K}. \tag{2.17}$$

On the other hand, in the extreme case, consider (2.12) and (2.13) where P templates are

yielding perfect matches. Clearly, when $P = 1$, template T_1 provides a perfect match while the others offer 0 correlation. We should then allocate no mass for Θ, i.e., $m(\Theta) = 0$. On the other hand, when $P = N$, all the templates exhibit equally 'perfect' matches, and no decision can be made in favor of any template. We should then allocate $m(\Theta) = 1$. Using a linear relationship, for the extreme case, we then use

$$m(\Theta) = \frac{P-1}{N-1}. \tag{2.18}$$

Compare (2.17) and (2.18) to get

$$K = (N-1)P. \tag{2.19}$$

Thus we arrive at the following DST model:

$$m(A) = \begin{cases} \dfrac{N-P}{P(N-1)}, & \text{for } A = T_i,\ i \in \overline{1,P}; \\ \dfrac{P-1}{N-1}, & \text{for } A = \Theta; \\ 0, & \text{otherwise.} \end{cases} \tag{2.20}$$

General Case. c) Masses Allocated to Singletons $m(T_i)$: As we did for the extreme case, we again propose

$$m(T_i) \propto h_i \implies m(T_i) = \frac{h_i}{K}, \tag{2.21}$$

where $K > 0$ is the proportionality constant.

Mass Allocated to Complete Ignorance $m(\Theta)$ Again, as before, we must have

$$m(\Theta) = 1 - \sum_{i=1}^{P} m(T_i) = 1 - \frac{S_{\mathbf{h}}}{K}. \tag{2.22}$$

In the extreme case, we know that $S_{\mathbf{h}} = P(N-P)$. This corresponds to the case when P templates have perfect matches, viz., $v_i = 1$ and $c_i = h_i = N - P$, for $i \in \overline{1, P}$ (see (2.12) and (2.13)). In the general case however, when one or more of the P templates have correlation coefficients that are less than 1, as we show in Appendix 2.5, $S_{\mathbf{h}} < P(N-P)$. In other words, for a given value of P, the extreme case (where $v_i = 1$, $i \in \overline{1, P}$) yields the maximum value for $S_{\mathbf{h}}$. So,

$$m(\Theta) = 1 - \frac{S_{\mathbf{h}}}{K} > 1 - \frac{P(N-P)}{K}. \tag{2.23}$$

The amount of the increase in $m(\Theta)$ from the extreme case should be $[P(N-P) - S_{\mathbf{h}}]/K$. Thus we would get

$$m(\Theta) = \frac{P-1}{N-1} + \frac{P(N-P) - S_{\mathbf{h}}}{K}. \tag{2.24}$$

Since the masses should add to 1, considering (2.21) and (2.24), we get

$$\frac{S_{\mathbf{h}}}{K} + \frac{P-1}{N-1} + \frac{P(N-P) - S_{\mathbf{h}}}{K} = 1. \tag{2.25}$$

This yields

$$K = P(N-1). \tag{2.26}$$

Thus, we arrive at the following DST model:

$$m(A) = \begin{cases} \dfrac{h_i}{P(N-1)}, & \text{for } A = T_i,\ i \in \overline{1,P}; \\ 1 - \dfrac{S_{\mathbf{h}}}{P(N-1)}, & \text{for } A = \Theta; \\ 0, & \text{otherwise.} \end{cases} \tag{2.27}$$

Noting that $h_i = N - P$, $t \in \overline{1,P}$, for the extreme case, we note that the DST model in (2.27) is valid for the extreme case as well. Finally, we note that, for $i \in \overline{1,P}$,

$$Bl(T_i) = \frac{h_i}{P(N-1)};\ \ Pl(T_i) = 1 - \frac{\sum_{j \neq i} h_j}{P(N-1)}. \tag{2.28}$$

2.4 Evaluation and Discussion

In this section, we address the following research questions regarding the a bounded matching algorithm (i.e. correlations coefficients) being implemented for the use of DS Theory.

RQ1 Does the proposed DS mass assignments hold for theoretical generalizable cases?

RQ2 How robust is the proposed DS algorithm when applied within the context of numerical character recognition in images?

2.4.1 RQ1: Generalized Template Match Model Validation

In this section, we provide some examples to demonstrate the validity of the proposed DST model. All examples use $N = 4$, and we will use $\mathbf{m}$ to denote the vector of mass assignments,

i.e.,

$$\mathbf{m} = [m(T_1),\, m(T_2),\, m(T_3),\, m(T_4),\, m(\Theta)]^T. \tag{2.29}$$

Some General Observations. *Example 1:* $\mathbf{v} = [y,\, y,\, y,\, y]^T$. This example demonstrates the case when all the correlation coefficients are identical. In this case, we get

$$\Delta\mathbf{W} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

and

$$\mathbf{c} = \mathbf{h} = [0,\, 0,\, 0,\, 0].$$

Thus, $P = 0$, i.e., we have no singleton focal elements. The corresponding DST model is

$$\mathbf{m} = [0,\, 0,\, 0,\, 0,\, 1]^T.$$

Note that, $Un(T_i) = [0, 1]$, $i \in \overline{1, 4}$, which shows the complete lack of evidence to make any decision. ■

Example 2: $\mathbf{v} = [y,\, y,\, y,\, z]^T$ *with* $z > y$. Here, all the correlation coefficients are identical, except one which has a higher coefficient. In this case, we get

$$\Delta\mathbf{W} = \begin{bmatrix} 0 & 0 & 0 & z(z-y) \\ 0 & 0 & 0 & z(z-y) \\ 0 & 0 & 0 & z(z-y) \\ y(y-z) & y(y-z) & y(y-z) & 0 \end{bmatrix},$$

and

$$\mathbf{c} = [y(y-z),\ y(y-z),\ y(y-z),\ 3z(z-y)];$$
$$\mathbf{h} = [0,\ 0,\ 0,\ 3z(z-y)].$$

Thus, $P = 1$. The corresponding DST model is

$$\mathbf{m} = [0,\ 0,\ 0,\ 3z(z-y)/3,\ 1 - 3z(z-y)/3]^T.$$

Note that the mass assignments for T_4 and Θ are dependent on the distance $(z-y)$. As $(z-y)$ increases, the $m(\Theta)$ decreases and $m(T_4)$ increases. The maximum distance achievable is when $z = 1$ and $y = 0$, yielding $m(T_4) = 1$. However, as $(z-y)$ decreases, the $m(\Theta)$ increases and $m(T_4)$ decreases. The minimum achievable distance occurs when $z = y$. As we approach this minimum distance, we converge to the scenario of maximum conflict where $m(\Theta) = 1$. Also note that $Un(T_4) = [3z(z-y)/3, 1]$. ■

Example 3: $\mathbf{v} = [y,\ y,\ z,\ z]^T$ *with* $z > y$. Here, half the correlation coefficients are identical; the other half also has identical coefficients but with a higher value. In this case, we get

$$\Delta\mathbf{W} = \begin{bmatrix} 0 & 0 & z(z-y) & z(z-y) \\ 0 & 0 & z(z-y) & z(z-y) \\ y(y-z) & y(y-z) & 0 & 0 \\ y(y-z) & y(y-z) & 0 & 0 \end{bmatrix},$$

and

$$\mathbf{c} = [2y(y-z),\ 2y(y-z),\ 2z(z-y),\ 2z(z-y)];$$
$$\mathbf{h} = [0,\ 0,\ 2z(z-y),\ 2z(z-y)].$$

Thus, $P = 2$. The corresponding DST model is

$$\mathbf{m} = [0,\ 0,\ 2z(z-y)/6,\ 2z(z-y)/6,\ 1-4z(z-y)/6]^T.$$

Compare this DST model, Example 3, with that of Example 2 above. While the masses are still dependent on the distance $(z-y)$, the values of the masses for T_3 and T_4 have been reduced. This lowering of the value can be attributed to having additional competing templates. Furthermore, notice how $m(\Theta)$ has been increased; it will never be reduced to 0, even if a maximum distance between z and y is reached. The range of of values for $m(\Theta)$ is $[0.33, 1]$ because of the amount of focal elements present in the vector. Also note that $Un(T_3) = [Un(T_4) = 2z(z-y)/6, 1-2z(z-y)/6]$. ■

Numerical Examples. *Example 1:* $\mathbf{v} = [0.01,\ 0.03,\ 0.12,\ 0.98]^T$. Here, the template T_4 is the only strong match. Note that we have

$$\Delta\mathbf{W} = \begin{bmatrix} 0 & -0.125 & -0.232 & -0.204 \\ 0.147 & 0 & -0.172 & -0.159 \\ 0.568 & 0.357 & 0 & -0.030 \\ 0.666 & 0.440 & 0.040 & 0 \end{bmatrix},$$

and

$$\mathbf{c} = [-0.0110,\ -0.0306,\ -0.0792,\ 2.7244];$$
$$\mathbf{h} = [0,\ 0,\ 0,\ 2.7244].$$

Thus, $P = 1$. The corresponding DST model is

$$\mathbf{m} = [0,\ 0,\ 0,\ 0.9081,\ 0.0919]^T.$$

Also note that $Un(T_4) = [0.9081, 1]$. ■

Example 2: $\mathbf{v} = [0.98, 0.83, 0.40, 0.30]^T$. Here, the templates T_1 and T_2 act as strong matches. Note that we have

$$\Delta\mathbf{W} = \begin{bmatrix} 0 & -0.125 & -0.232 & -0.204 \\ 0.147 & 0 & -0.172 & -0.159 \\ 0.568 & 0.357 & 0 & -0.030 \\ 0.666 & 0.440 & 0.040 & 0 \end{bmatrix},$$

and

$$\mathbf{c} = [1.3818, 0.6723, -0.3640, -0.3930];$$

$$\mathbf{h} = [1.3818, 0.6723, 0, 0].$$

Thus, $P = 2$. The corresponding DST model is

$$\mathbf{m} = [0.2303, 0.1120, 0, 0, 0.6577]^T.$$

Also note that

$$Un(T_1) = [0.2303, 0.8880];\ Un(T_2) = [0.1120, 0.7697].$$ ■

Example 3: $\mathbf{v} = [0.31, 0.32, 0.43, 0.44]^T$. Here, all the templates are weak. Note that

$$\Delta\mathbf{W} = \begin{bmatrix} 0 & 0.0032 & 0.0516 & 0.0572 \\ -0.0031 & 0.0473 & 0.0528 & \\ -0.0372 & -0.0352 & 0 & 0.0044 \\ -0.0403 & -0.0384 & -0.0043 & 0 \end{bmatrix},$$

and

$$\mathbf{c} = [-0.0806,\ -0.0704,\ 0.0946,\ 0.1144];$$

$$\mathbf{h} = [0,\ 0,\ 0.0946,\ 0.1144].$$

Thus, $P = 2$. The corresponding DST model is

$$\mathbf{m} = [0,\ 0,\ 0.0158,\ 0.0191,\ 0.9652]^T.$$

This illustrates that, with our proposed model, even with a weak set, masses are still assigned to the stronger templates, but with an increased value for $m(\Theta)$. Also note that

$$Un(T_1) = Un(T_2) = [0, 0.8652];$$

$$Un(T_3) = [0.0158, 0.9810];\ Un(T_4) = [0.0191, 0.9843].$$ ■

Example 4: $\mathbf{v} = [0.31,\ 0.32,\ 0.73,\ 0.74]^T$. We use this example to compare with Example 3 to demonstrate how uncertainty decreases if T_3 and T_4 become stronger candidates. We have

$$\Delta\mathbf{W} = \begin{bmatrix} 0 & 0.0032 & 0.3066 & 0.3182 \\ -0.0031 & 0 & 0.2993 & 0.3108 \\ -0.1302 & -0.1312 & 0 & 0.0074 \\ -0.1333 & -0.1344 & -0.0073 & 0 \end{bmatrix},$$

and

$$\mathbf{c} = [-0.2666,\ -0.2624,\ 0.5986,\ 0.6364];$$

$$\mathbf{h} = [0,\ 0,\ 0.5986,\ 0.6364].$$

Thus, $P = 2$. The corresponding DST model is

$$\mathbf{m} = [0,\ 0,\ 0.0998,\ 0.1061,\ 0.7942]^T.$$

Notice how T_3 and T_4 are being supported more, while the support for $m(\Theta)$ is being reduced. Also note that

$$Un(T_1) = Un(T_2) = [0, 0.7942];$$
$$Un(T_3) = [0.0998, 0.8940];\ Un(T_4) = [0.1061, 0.9003].$$ ■

2.4.2 RQ2: Correlation Coefficients as a a form of Evidence

In this subsection, we demonstrate the applicability and robustness of the proposed model within the context of numerical character recognition within images.

Data Set. We used the data set in [147] for detecting numerical characters from computer fonts with 4 variations (combinations of italic, bold, and normal) and in 85 font types. Each font type provides a set of 10 images thus creating a total of 850 template images within the data base.

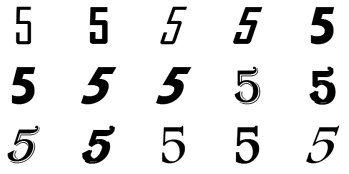


Figure 2.2: Template T_5 in 15 different font types.

Formulation of Correlation Coefficients as Evidence. For our experiment, we used only 15 font types. With the templates $\{T_0, \ldots, T_9\}$ denoting the numerical characters $\{0, \ldots, 9\}$, respectively, we therefore used a set of 150 templates. Fig. 2.2 shows the 15 font types corresponding to the template T_5 (which corresponds to the numerical character "5"). The data base was then partitioned into two groups. The first partition was designated as

the training set (i.e., evidence sources) and the other as the testing set. It is important to emphasize that the purpose of this experiment is to simply evaluate the DST model. We have undertaken no preprocessing of the images.

With each font giving a set of 10 templates, when correlated with an unknown image, we obtain the vector $\mathbf{v}$ in (3.4). We treat this vector corresponding to each font as one source of evidence. With the 15 font types, we therefore obtain 15 evidence sources and their corresponding 15 DST models. These sources are then fused using the DCR in (7.8). For the testing data set, unknown candidate images are randomly selected from the remaining 700 templates and then we introduce 'salt and pepper' noise to corrupt the image prior to generating the correlation coefficients. Several different noise density, d, levels of 'salt and pepper' noise were employed. Figs 2.3, 2.4, and 2.5 show three levels of d. The noise level applied to a candidate image was quantified via the peak signal-to-noise ratio (PSNR) [87].

Results and Discussion. No noise was added initially to determine the baseline accuracy

Figure 2.3: $d = 0.55$

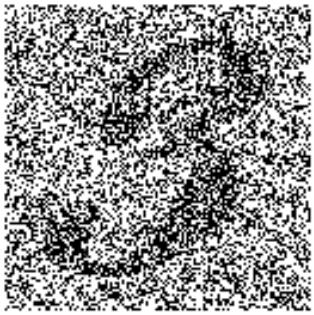


Figure 2.4: $d = 0.75$

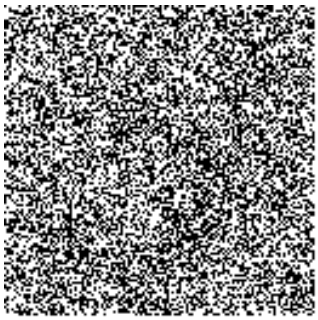


Figure 2.5: $d = 0.90$

which was 93.17% for the detection of 600 randomly selected samples using all 15 evidence sources. It typically required a minimum of 3 fusion combinations to converge to the correct match, thus making it obvious that simply taking the highest correlation coefficient would be an inadequate approach. The confusion matrix in Table 2.2 demonstrates the difficultly in distinguishing a specific template from one another (in particular, for T_4 and T_1).

When high distortion is introduced into the samples ($d = 0.9$, PSNR = 35.02), the overall accuracy is 76.7%. The corresponding confusion matrix appears in Table 2.3, where an

Table 2.2: Confusion Matrix: $d = 0$, accuracy = 93.17%

	T_0	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9
T_0	55	0	0	0	0	0	0	0	2	0
T_1	0	49	0	0	15	0	0	0	0	0
T_2	0	0	71	0	0	0	0	0	0	0
T_3	0	0	0	54	0	0	0	0	0	0
T_4	0	5	0	0	53	0	0	0	0	0
T_5	0	0	0	0	0	58	0	0	0	0
T_6	0	0	0	0	2	9	40	0	4	0
T_7	0	1	0	0	0	0	0	69	0	0
T_8	0	0	0	0	0	0	0	0	52	0
T_9	3	0	0	0	0	0	0	0	0	58

increase in false positives and false negatives across all templates is observed.

Table 2.3: Confusion Matrix: $d = 0.9$, accuracy = 76.7%

	T_0	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9
T_0	42	0	0	0	2	0	2	3	3	0
T_1	0	43	1	0	7	0	0	0	0	1
T_2	1	0	55	4	0	0	0	11	0	0
T_3	0	1	0	45	0	8	0	0	11	1
T_4	0	5	0	0	67	2	0	1	0	1
T_5	5	1	0	1	0	52	5	0	1	0
T_6	8	0	0	1	4	12	33	0	6	2
T_7	0	3	2	0	1	0	0	46	0	0
T_8	3	0	0	1	0	1	3	0	42	2
T_9	11	5	0	0	2	0	1	4	0	34

Table 2.2 demonstrates that detection errors were mostly attributable to a sample that may not fit well with our 150 evidence sources. This suggests that we could possibly pick different sources of evidence that could better represent the database for improved detection. On the other hand, Table 2.3 demonstrates that a failure in detection is attributable to noise distortion.

The importance of our proposed DST model is demonstrated by how it captures the uncertainty as noise is introduced into the sample. As Table 2.4 shows, when noise increases, we are unable to reduce the uncertainty even when multiple sources of evidence are fused. We are still able to detect the winner (by simply picking the template with the highest mass),

but now, the DST model provides us invaluable information regarding our confidence (or lack thereof) in the match. This is an important and useful feature of our model.

Table 2.4: Noise Levels Vs Uncertainty

$\mathbf{N_D}$	**.05**	**.25**	**.45**	**0.65**	**0.85**	**0.95**
$\mathbf{P_{SNR}}$	47.46	40.52	37.91	36.37	34.70	35.16
$\mathbf{\Theta}$	0.57	0.74	0.86	0.94	0.98	0.99

2.5 Conclusion

We have proposed a simple DST model for capturing the uncertainty associated with allocating a template for a given candidate signal. This model can be especially useful in situations where there is no clear winning template in terms of the correlation coefficient values. Such a situation creates uncertainty as to the template to be declared the winner. The DST framework allows the model to be developed with little prior knowledge, and the DST uncertainty interval (constructed from the belief and plausibility values) provide valuable information regarding the confidence one can place on this decision. Moreover, DST evidence fusion strategies can be utilized to fuse evidence generated from different sources thus allowing the decision to be refined and improved. This is exactly the strategy that we followed in the experiment carried out in Section 2.4.2.

We emphasize that the results in Section 2.4.2 are extracted simply by using the correlation coefficient as applied *directly* to the templates and the candidate. In practice, when one employs correlation coefficients for template matching, template and candidate signals are pre-filtered (e.g., light compensation, rotation, etc., for images). The results we give do not employ any pre-filtering, and the decisions can be significantly improved by employing such schemes.

The model proposed in this work is simple in the sense that it captures uncertainty via the assignment of a mass to the complete ambiguity (i.e., Θ). A better model would be to

allow other non-singleton propositions to be focal elements. For example, if the correlation coefficients corresponding to the templates T_1 and T_2 are high and the coefficients for the others are low, a better model would generate a focal element from (T_1, T_2). The model in this paper employs Θ as the only non-singleton focal element.

Chapter Appendix: Maximum of $S_{\mathbf{h}} = \sum_{i=1}^{N} h_i$

Consider the correlation coefficient vector $\mathbf{v}$ in (3.4), where $v_i \in [0, 1]$, $\forall i \in \overline{1, N}$. Generate the following matrix associated with $\mathbf{V}$:

$$\Delta\mathbf{W} = \begin{bmatrix} v_1\Delta v_{11} & v_2\Delta v_{21} & \dots & v_N\Delta v_{N1} \\ v_1\Delta v_{12} & v_2\Delta v_{22} & \dots & v_N\Delta v_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ v_1\Delta v_{1N} & v_2\Delta v_{2N} & \dots & v_N\Delta v_{NN} \end{bmatrix}, \tag{2.30}$$

where $\Delta v_{ij} = (v_i - v_j)$. Next, generate a column weights vector, $\mathbf{c}$, by summing the elements of each column of $\Delta\mathbf{W}$:

$$\mathbf{c} \quad = \quad \left[N\, v_1^2 - v_1 S_{\mathbf{v}},\ N\, v_2^2 - v_2 S_{\mathbf{v}},\ \dots, N\, v_j^2 - v_j S_{\mathbf{v}},\ \dots,\ N\, v_N^2 - v_N S_{\mathbf{v}}\right]. \tag{2.31}$$

Without loss of generality, let us assume that only the first P elements of $\mathbf{c}$ are positive which implies that

$$N\, v_i^2 - v_i S_{\mathbf{v}} \begin{cases} > 0, & \text{for } i \in \overline{1, P}; \\ \leq 0, & \text{for } i \in \overline{P+1, N}, \end{cases}$$

$$\iff v_i \begin{cases} > S_{\mathbf{v}}/N, & \text{for } i \in \overline{1, P}; \\ \leq S_{\mathbf{v}}/N, & \text{for } i \in \overline{P+1, N}. \end{cases} \tag{2.32}$$

The mass measure vector thus generated is

$$\mathbf{h} \quad = \quad \left[N\,v_1^2 - v_1 S_{\mathbf{v}},\ N\,v_2^2 - v_2 S_{\mathbf{v}}, N\,v_P^2 - v_P S_{\mathbf{v}},\ 0,\ \ldots,\ 0\right]. \quad (2.33)$$

Then,

$$S_{\mathbf{h}} = \sum_{j=1}^{P} \left(N\,v_j^2 - v_j S_{\mathbf{v}}\right) = N \sum_{j=1}^{P} v_j^2 - S_{\mathbf{v}} \left(\sum_{j=1}^{P} v_j\right). \quad (2.34)$$

We now consider the following problem: how much can we increase $S_{\mathbf{h}}$ by changing v_j, $j \in \overline{1, P}$, while making sure that only the first P elements of $\mathbf{c}$ are positive? We note the following:

- For v_i, $i \in \overline{1, P}$, use (2.32) to get $\partial S_{\mathbf{h}}/\partial v_i = 2Nv_i - S_{\mathbf{v}} - \sum_{j=1}^{P} v_j > 0$.
- For v_i, $i \in \overline{P+1, N}$, $\partial S_{\mathbf{h}}/\partial v_i = -\sum_{j=1}^{P} v_j < 0$.

So, it is clear that the maximum of $S_{\mathbf{h}}$ is achieved when v_i, $i \in \overline{1, P}$ are increased to their maximum value (viz., 1) and when v_i, $i \in \overline{P+1, N}$, are decreased to their minimum value (viz., 0). In doing so, we do not violate the conditions in (2.32) and therefore we ensure that only the first P elements of $\mathbf{c}$ are positive. So, putting $v_i = 1$, $i \in \overline{1, P}$, and $v_i = 0$, ß $\in$ $\overline{P+1, N}$, in (2.34), we get

$$\max S_{\mathbf{h}} = NP - P^2 = P(N - P). \quad (2.35)$$

The best teamwork comes from men who are working independently toward one goal in unison.

James Cash Penney

3

DS MapReduce Paradigm

Normalized cross-correlation template matching is used as a detection method in many scientific domains. To be practical, template matching must scale to large datasets while handling ambiguity, uncertainty, and noisy data. We propose a novel approach based on Dempster-Shafer (DS) Theory and MapReduce parallelism. DS Theory addresses conflicts between data sources, noisy data, and uncertainty, but is traditionally serial. However, we use the commutative and associative nature of Dempster's Combination Rule to perform a parallel computation of DS masses and a logarithmic hierarchical fusion of these DS masses. This parallelism is particularly important because additional data sources allow DS-based

template matching to maintain accuracy and refine uncertainty in the face of noisy data. We validate the parallelism, accuracy, and uncertainty of our implementation as a function of the size and noise of the input dataset, finding that it scales linearly and can retain accuracy and improve uncertainty in the face of noise for large datasets.

Template Matching methods are ubiquitous in the scientific community. Matching approaches have been implemented across many scientific disciplines ranging from electrocardiogram analysis in medicine [82, 86], to gene expression in biology [129], to signature detection in communications [9], to analysis of seismic signals in geo-science [7, 22], and to image tracking and recognition in computer vision [24, 83, 163]. Traditional template matching methods are popular because they are effective, easily interpreted, and can be evaluated or prototyped rapidly. The fundamental concept behind template matching is to quantify similarity between two *objects*, which are single- or multi-dimensional signals such as time series information or images. There are several ways to quantify similarity, including sum of absolute difference, histogram matching, and selective correlation coefficients [83, 101, 163]. This paper focuses on the popular method of correlation to quantify object similarity. This statistical approach measures the linear relationship between two objects by providing a correlation coefficient as a metric indicating evidence of similarity.

Prior Work. Traditional single-template cross-correlation matching techniques establish the statistical linear relationship between two objects' unique features [107]. A template is serially translated over a signal or image to search for the strongest correlation coefficient [47, 65, 82], and if a threshold criterion is exceeded a match is concluded. Contemporary approaches also consider sets of templates. A template set acts as a source of evidence for detecting various features within a single signal of interest [86]. Each template in the set represents a specific hypothesis; the template with the highest coefficient is reported as the winner. This work raises three interesting questions: 1) what can be done if this serial algorithm does not scale to larger datasets, 2) what should be returned if multiple templates

report similar correlation coefficients, and 3) what is the certainty that the reported template is the true match?

Challenges. This issue of ambiguous winners creates an uncertainty regarding which template is the true match [110]. The complexity of the decision is increased when additional sources of information are provided as sets of multiple templates. Current approaches employ voting and weighting strategies [55,82,102]. However, these approaches do not appropriately capture the uncertainty associated with deciding a winner [110]. This situation is further compounded when working with high-volume datasets that require significant computation time. Additionally, the veracity of real-world data is frequently degraded by noise and measurement error. We desire a scalable template matching approach that handles multiple templates and a high volume of noisy source data while still capturing ambiguity.

Insights. As relevant datasets become larger and more readily available, there is evidence that increasing the size of the dataset can increase detection specificity and sensitivity [167]. Such benefits are contingent on proper regularization and treatment of noisy samples [167]. Dempster-Shafer (DS) theory offers *combination rules* that can properly account for the regularization of noise [29]. These combination paradigms allow fusion of evidence sources into a single set of hypotheses. Fusing sources allows contextual considerations to be captured, such as conflicts between sources, corrupt information, uncertainty, source reliability, and accuracy [134]. We propose to use a DS theory-based approach for template matching to address the issues of multiple templates and ambiguity; however, the benefits of DS theory cannot be fully leveraged unless the approach applies efficiently to many templates and sources.

The MapReduce distributed programming paradigm has become increasingly well-supported by companies and computing clusters [45]. MapReduce takes advantage of insights from functional and parallel programming to gain high performance, but requires that computations be structured and data be staged with Map and Reduce tasks operating over $\langle key, value \rangle$

pairs. While MapReduce has been used successfully in research [90, 106] and industrial practice [45, 136], to our knowledge it has not been used to support DS theory. This may result from the relative nascency of MapReduce and the variety and novelty of DS theory frameworks to form evidence. We exploit the associativity and commutativity of Dempster's Combination Rule to produce a parallel Map and a hierarchical logarithmic Reduce over $\langle key, value \rangle$ pairs representing DS correlation information, to scale template matching to large datasets.

Contributions. We design and implement a parallelized and distributed framework based on the MapReduce paradigm that carries out Dempster-Shafer theory calculations to perform multiple-set template matching in a manner that handles large datasets, ambiguity, and noisy data. The contributions of this work are:

1. We develop a MapReduce framework for an evidence fusion methodology that can leverage large volumes of templates and sources to improve uncertainty.
2. We demonstrate that our framework scales to high volumes of data.
3. We demonstrate that our framework is robust against noisy data.

3.1 Background

In this section, we introduce the MapReduce framework for distributed computation and the Dempster-Shafer Theory approach evidence fusion.

3.1.1 MapReduce

MapReduce [45] is a programming paradigm intended for distributing computations over large datasets on a cluster. In principle, MapReduce consists of two phases: *Map* and *Reduce*, named after the map and reduce (or fold) functions in functional programming. MapReduce

is powerful in situations where a large amount of input can be processed independently (e.g., embarrassingly parallel applications). The Map phase transforms, filters or sorts data in parallel. Map operates on each element of the input (represented as a $\langle key, value \rangle$ pair) and produces zero or more $\langle key, value \rangle$ pairs as output. Map should be stateless, operating only on its input. Many instances of the Map function can execute simultaneously (e.g., on different nodes) because there are no dependencies between pairs. After the Map phase, the output may be staged or exchanged between nodes, and $\langle key, value \rangle$ pairs are sorted and assigned (or partitioned) to nodes for reduction. The Reduce function is applied once for each *key*, accessing all of the *value*s associated with that *key* and producing zero or more outputs. MapReduce frameworks or implementations handle staging, marshaling, and data transfer aspects while the user provides a few specified functions (e.g., reading the input files, Map, Reduce, writing the output, etc.). In this paper, we take advantage of the commutativity and associativity of Dempster's Combination Rule to realize a highly parallelized framework for DS fusion. Traditionally, MapReduce parallelism benefits come from the Map stage; hierarchical reductions such as the one proposed in Section 3.2.4 are not part of most MapReduce frameworks.

3.1.2 Dempster-Shafer Theory Background and Notation

Dempster-Shafer Theory is an evidence-based approach which develops support for hypotheses. Evidence is constructed around information regarding occurrences of events [75]. A single piece of evidence may support multiple hypotheses. To create the necessary evidence, information is gather from *sources*, which can be sensors, organizations, databases, people, or other entities [75].

DS Theory is a generalization of classical probability theory [145], where the support of hypotheses can be considered propositions. These propositions, referred to as the *frame of discernment* (FOD), are mutually exclusive and exhaustive. We take the FOD Ω to be a

finite set (i.e., $\Omega = \{\theta_1, \ldots, \theta_N\}$). Ω is finite and composed of N *singleton* propositions. The *basic probability assignment* (BPA), otherwise referred to as a *basic belief assignment* or *mass function*, is a function $m : 2^{\Omega} \rightarrow [0, 1]$, where 2^{Ω} is the power set of Ω, such that all probabilities sum to 1:

$$m(\emptyset) = 0; \quad \sum_{X_i \subseteq 2^{\Omega}} m(X_i) = 1. \tag{3.1}$$

While $m(X)$ measures *only* the support that is directly assigned to proposition $X \subseteq \Omega$, the *belief* $Bl(X)$ represents the total support that can move into X from all other propositions that contain X. Thus, $Bl(X) = \sum_{Y \subseteq X} m(Y)$. Belief is the minimum amount of support that is given for a specific proposition. For the singleton case, the DS mass of the proposition is equal to the belief.

Capturing Uncertainty via DS Theory. Uncertainty has many ways of entering a model. There are two classifications of uncertainty: *aleatory uncertainty* and *epistemic uncertainty*. Aleatory uncertainty represents unknowns that differ each time the system is observed, and can be accounted for using historic data. In this work, we focus on epistemic uncertainty, caused by a lack of knowledge, which is reduced through increased understanding [134, 145]. Hence, when more evidence is provided a refinement of the uncertainty for the decision is possible through increased knowledge.

DS Theory is a major tool for deciphering uncertainty. It uses data fusion techniques to reduce uncertainty from imperfect data (e.g., source information that is conflicting or sources reporting similar information) [91, 134, 155]. Combining additional sources of evidence, referred to as *evidence fusion* [109], reduces uncertainty and redistributes masses to the propositions. Therefore, when one source of evidence indicates a specific template and another source conflicts with that prediction, the uncertainty in the model increases. Likewise, if there is no conflict and both sources report similar findings, the uncertainty decreases.

The evidence fusion approach implemented in this paper uses *Dempster's Combination Rule* (DCR):

$$m(X_i) = \frac{\sum_{X_p \cap X_q = X_i} m_1(X_p)\, m_2(X_q)}{1 - \sum_{X_p \cap X_q = \emptyset} m_1(X_p)\, m_2(X_q)}, \tag{3.2}$$

where the evidence provided by the mass functions m_1 and m_2 are combined to obtain the fused mass function m. We write $m = m_1 \oplus m_2$ to denote that m is the fused mass function produced by combining m_1 and m_2.

Finally, we note that the DCR $\oplus$ has two important mathematical properties: it is commutative (i.e., $m_1 \oplus m_2 = m_2 \oplus m_1$) and associative (i.e., $(m_1 \oplus m_2) \oplus m_3 = m_1 \oplus (m_2 \oplus m_3)$). While these properties of DCR are well-known in the literature [75, 134], to our knowledge they have not been previously used to support a distributed implementation of DS Theory.

3.2 MapReduce Framework for Template Fusion

To position our framework within a MapReduce environment, we describe two algorithms, `Map` and `TreeReduce`, that compute a winning match. Typical implementations compute a match serially. However, by noting the associativity and commutativity of Dempster's Combination Rule, we can execute our Map and TreeReduce functions in parallel, gaining significant speedups when a computing cluster is available. In addition, the structure of our inputs is pivotal to the correct execution of these algorithms.

3.2.1 Source Dataset

The input dataset is structured to properly combine MapReduce with DS Theory. This dataset contains D different sources, with an individual source denoted by $\mathbf{S}_d$. Each source is considered to provide evidence for its own FOD, and all sources contain the same number

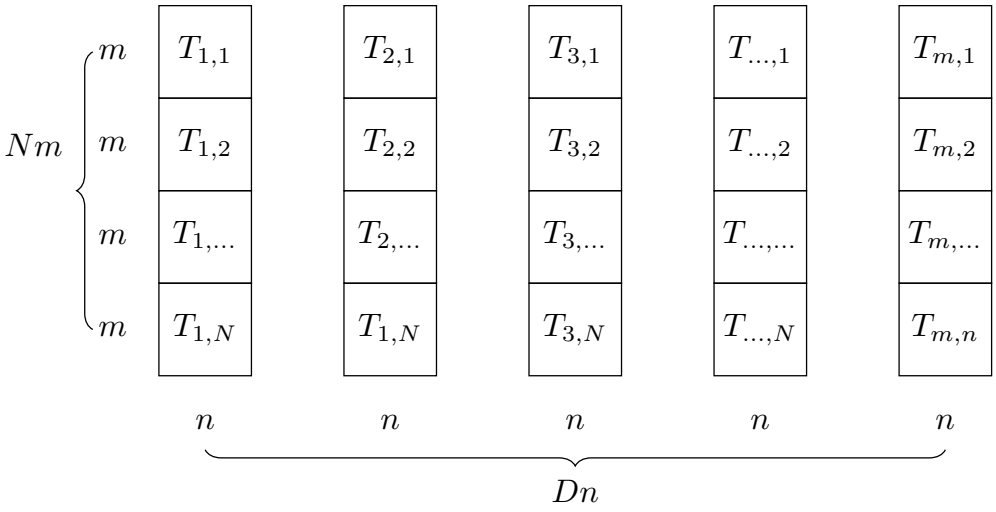


Figure 3.1: Visual representation of the structure of input sources. Each template is an $m \times n$ signal, each source contains N different templates. There are D sources in the dataset.

of templates, N, in the same order. A specific template i associated with source d is denoted as $T_{d,i}$. Each template i supports a specific hypothesis. As each source contains a series of templates in the same order, each source can thus be thought of a series of hypotheses. For example, template $T_{1,i}$ in source $\mathbf{S}_1$ reflects the same proposition/hypothesis generated by template $T_{2,i}$. Furthermore, every template in every source is identical in size.[1] This is a necessary precondition for the computation of correlation coefficients. In practice, templates can be uniformly resized in a preprocessing step. Each set of templates is examined to create our support for each proposition in the set. This structure allows each source to be reconstructed into its own FOD (Ω), which is crucial for MapReduce because a single FOD can be represented by a single $\langle key, value \rangle$ pair.

3.2.2 Correlation to DS Framework: Map

The Map function's task is to analyze a source's information and develop evidence using correlation. This set of correlation coefficients is then placed into a DS Framework [110].

[1] In our evaluation, we investigate an example image processing task in which we consider every image to have the same dimensions.

There are two inputs to the Map function. The first input is an unknown 2-dimensional signal, A, with dimensions m by n. The second input is a Source, $\mathbf{S}_d$, composed of a set of N templates, where each template's dimension are m by n.

Forming Evidence Using Correlation

To develop evidence, correlation between the unknown template and a known template, $T_{d,i}$, produces a correlation coefficient. This correlation coefficient is assigned to the template, providing evidence of the match. A cross-correlation value of $\rho = 1$ denotes a perfect match between the unknown and known templates. When $\rho = 0$, there is no match, and when $\rho = -1$, the two signals are out of phase or negatively correlated. Since we are examining a set of correlation coefficients from a source, this framework can be applied to multidimensional template matching. In this paper, our work considers the two dimensional case (e.g., images), where the correlation coefficient can be expressed as

$$\rho_{\mathbf{A},\mathbf{B}} = \frac{\sum_{j=1}^{m}\sum_{i=1}^{n}(A_{ij} - \overline{\mathbf{A}})(B_{ij} - \overline{\mathbf{B}})}{\sqrt{\left(\sum_{j=1}^{m}\sum_{i=1}^{n}(A_{ij} - \overline{\mathbf{A}})^2\right)\left(\sum_{j=1}^{m}\sum_{i=1}^{n}(B_{ij} - \overline{\mathbf{B}})^2\right)}}, \tag{3.3}$$

where we denote the 'sample' means of images A and B as

$$\overline{\mathbf{A}} = \frac{1}{mn}\sum_{i=1}^{n}\sum_{j=1}^{m}A_{ij}, \quad \text{and} \quad \overline{\mathbf{B}} = \frac{1}{mn}\sum_{i=1}^{n}\sum_{j=1}^{m}B_{ij}.$$

This process normalizes the correlation coefficient, ρ, where $\rho \in [-1, 1]$. A set of normalized correlation coefficients is generated from $\mathbf{S_d}$'s set of templates by using equation 4.2 and the unknown template thus take the form

$$\mathbf{V} = \begin{bmatrix} V_1 & V_2 & \cdots & V_N \end{bmatrix}^T, \tag{3.4}$$

where $V_i \in [0, 1]$ to only denote the positive normalized correlation coefficient. The negative correlation coefficients are set to zero. Further discussion can be found in Napoli *et al.* [110]. Each V_i provides evidence on how similar $T_{d,i}$ is to the unknown template. This correlation calculation is carried out for all sources in a parallelized manner.

3.2.3 Set of Correlations Coefficients to DS Masses

We follow the DS framework of Napoli *et al.* [109] for correlation coefficients to capture the overall certainty of each element in $\mathbf{V}$ by utilizing a weighting strategy as

$$\begin{aligned}\Delta\mathbf{W} &= \Delta\mathbf{V} \circ \mathbf{J}_{NN}\mathbf{D}_N \\ &= \begin{pmatrix} V_1\Delta V_{11} & V_2\Delta V_{21} & \cdots & V_N\Delta V_{N1} \\ V_1\Delta V_{12} & V_2\Delta V_{22} & \cdots & V_N\Delta V_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ V_1\Delta V_{1N} & V_2\Delta V_{2N} & \ldots & V_N\Delta V_{NN} \end{pmatrix},\end{aligned} \tag{3.5}$$

where $\circ$ denotes the matrix Hadamard product, $\Delta V_{ij} = (V_i - V_j) \in [-1, +1]$. $\mathbf{J}_{NM}$ denotes the $N \times M$ matrix with each entry being 1 and $\mathbf{D}_N = diag\,[V_1, V_2 \cdots, V_N]$ denotes the diagonal matrix with the diagonal entries being $\{V_1, V_2, \cdots, V_N\}$. The columns of $\Delta\mathbf{W}$ are references for each V_i proposition being analyzed. The summation of these column vectors informs us the strength of a proposition being a winner relative to other propositions in $\mathbf{V}$. The sum of all the elements in a column vector is referred to as the *column weight*. The *column weights*, C_i, are calculated as

$$\mathbf{C} = \begin{bmatrix} C_1 & C_2 & \cdots & C_N \end{bmatrix}^T = (\mathbf{J}_{1N}\Delta\mathbf{W})^T, \tag{3.6}$$

where $C_i \in [-(N-1)/4, (N-1)]$. We determine P different *focal elements* from the positive values in the column weights vector, $\mathbf{C}$. These focal elements are assigned DS masses. We

define a *mass measure vector* $\mathbf{H} = [H_1, H_2, \cdots, H_N]$ as

$$H_i = \frac{C_i + |C_i|}{2}. \tag{3.7}$$

We first observe that each $H_i \in [0, (N-1)]$. More generally, $\mathbf{H}_d$ is associated with source $\mathbf{S}_d$. Note that, $\mathbf{H}_d$ is associated with source $\mathbf{S}_d$. The mass measure is adjusted to DS masses via

$$M(A) = \begin{cases} 1 - \left(\frac{Y}{(N-1)P}\right), & \text{for } A = \Theta; \\ \\ H_i \left(\frac{1-m(\Theta)}{Y}\right), & \text{for } A = H_i, \end{cases} \tag{3.8}$$

where $\Theta = \{V_1, V_2, \cdots, V_N\}$ and $Y = \sum_{i=1}^{N} H_i$ is the FOD consisting of the propositions H_i, $i \in 1, \ldots, N$. Once the correlation coefficients for a single source have been converted to masses, those mass vectors can be fused.

3.2.4 DCR Fusion: TreeReduce

Our TreeReduce function takes any two FOD vectors V_1 and V_2 as input. The TreeReduce function fuses the two vectors into a single DS mass vector of the same length. The TreeReduce function can then be executed hierarchically to fuse all D FOD vectors in $\log(D)$ steps until one fused vector remains. This final output vector contains the fused DS masses from all the processed sources. From this output, we account for the belief and uncertainty of the winning template based upon the maximal mass in the vector.

3.3 Architecture and Implementation

In this section, we describe the architecture and implementation of two prototypes for our parallel framework. The first prototype for running our framework uses Hadoop [1], which

Algorithm 1 — **Map**: calculate DS mass from one source.

Input: A, and unknown input observation to match
Input: B, a single source containing N templates
Input: corr, a subroutine implementing Equation 4.2
Output: correlation vector OUT of length $N + 1$

$count = 0$
for $i = 0$ to $N - 1$ **do**
  $C[i] = \text{corr}(A, B[i])$
for $i = 0$ to $N - 1$ **do**
  for $k = 0$ to $N - 1$ **do**
    $OUT[i] = OUT[i] + (C[i] - C[k])^*C[i]$
  if $OUT[i] < 0$ **then**
    $OUT[i] = 0$
  else
    $count = count + 1$
$sum = \text{sum}(OUT)$
if $sum > 0$ **then**
  $uncertainty = 1 - sum/count$
  for $i = 0$ to $N - 1$ **do**
    $OUT[i] = OUT[i] * ((1 - uncertainty)/sum)$
  $OUT[N] = uncertainty$
else
  $OUT[N] = 1$

Algorithm 2 — **TreeReduce**: Dempster's Combination Rule

Input: V_1 and V_2, mass vectors of length $N + 1$
Output: OUT, a fused vector of length $N + 1$

$tmp = (N + 1) \times (N + 1)$ array
$norm = 0$
for $i = 0$ to $N + 1$ **do**
  for $k = 0$ to $N + 1$ **do**
    $tmp[i][k] = V_1[i] * V_2[k]$
for $i = 0$ to $N + 1$ **do**
  $norm = norm + tmp[i][i] + tmp[i][N] + tmp[N][i]$
for $i = 0$ to N **do**
  $OUT[i] = (tmp[i][i] + tmp[i][N] + tmp[N][i])/norm$
$OUT[N] = tmp[N][N]/norm$

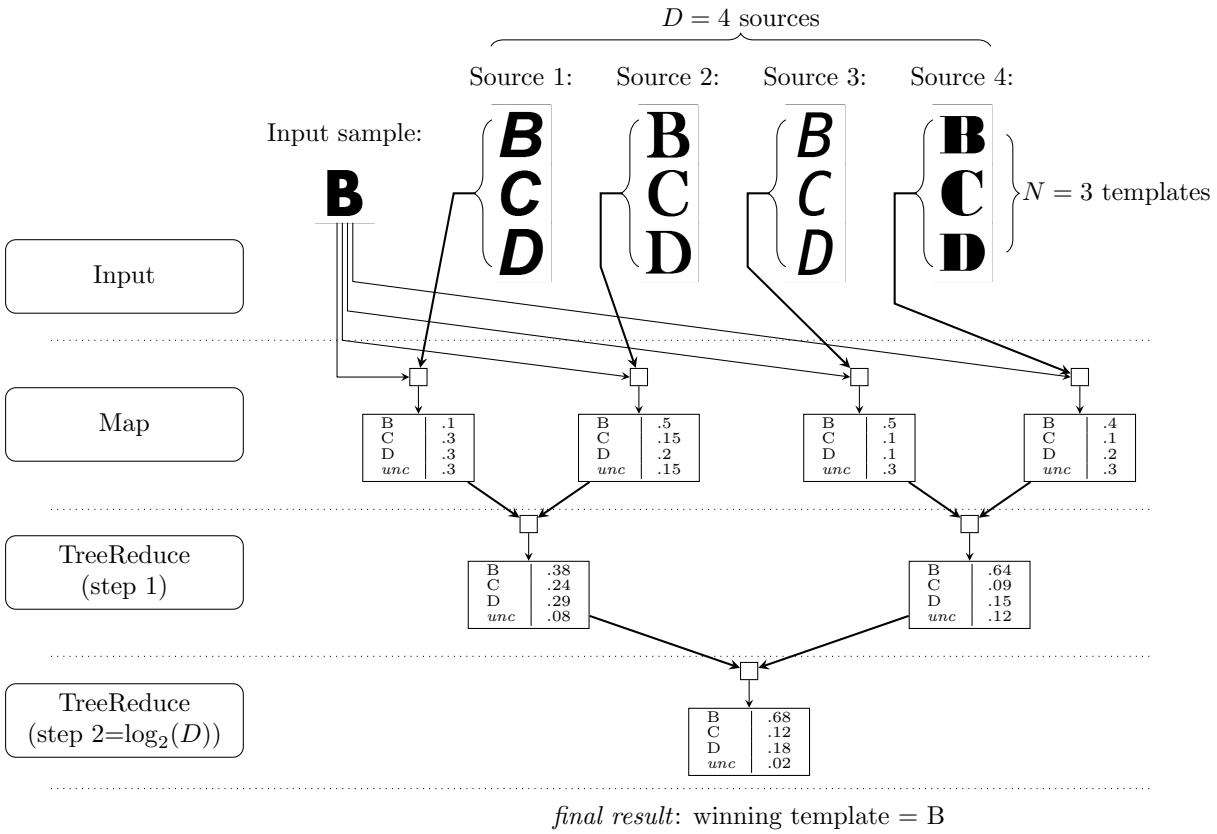


Figure 3.2: Architecture of the proposed framework on a handwriting identification task using template matching. D refers to the number of sources present in the system, N refers to the total number of templates each source provides. The Map operation correlates the input image with the ith template image of the dth source ($T_{d,i}$), producing a vector of masses. Pairs of masses are provided to each TreeReduce operation, which fuses the two vectors together. In particular, the fused vector has an increased mass for templates that already had higher masses in the input vector. *unc* represents the uncertainty.

allows arbitrary scalability depending on the number of nodes and cores available in a distributed compute cluster. The second prototype is a flexible pthreads implementation in Linux that uses the MapReduce paradigm akin to Phoenix [124]. This enables rapid debugging of our framework on a single node in instances where a Hadoop cluster is not practical.

3.3.1 Testing Methodology

We examine changes in mass vectors as a function of source count and compute core count. The system's workflow is illustrated in Figure 3.2 using a simple handwriting letter classification task. Each source contains an array of template images corresponding to alphanumeric characters. As illustrated in Figure 3.2, all templates across all sources maintain the same order (described in Section 3.2.1). The parallel Map operation produces an mass vector using the unknown input image and a particular source. The TreeReduce operation then applies DCR in parallel, fusing pairs of mass vectors together and outputting a single updated mass vector. The TreeReduce operation is repeated $\log(D)$ times until a single fused mass vector remains.

3.3.2 MapReduce Implementation

Our framework takes advantage of popular open source tools for massively parallel computation. We use Apache Hadoop [1], Hadoop FS [138], and Apache YARN [153] for cluster and computational resource management. These tools apply our framework to any general Hadoop cluster with minimal human effort. In addition, we take advantage of MapReduce4C (MR4C [60]), a C wrapper to run jobs on a Hadoop cluster. MR4C allows executing an arbitrary C function in a Hadoop task managed by YARN. Our overall framework enables the use of all the cores in a cluster to efficiently execute our implementation while reaping the benefits of temporal parallelism.

The first part of the implementation described in Algorithm 1 serves as the Map function. The $\langle key, value \rangle$ pair identifies the source and the array of mass values, respectively. In MR4C, the *key* would be the name of a source file and the *value* would be its contents. The output of Map is then a single $\langle key, value \rangle$ pair: the *key* is unchanged, but the output *value* is a vector of DS masses produced by the data values from that source and the unknown input sample. Because the inputs are independent, many instances of the Map function can execute simultaneously.

The second part of the implementation described in Algorithm 2 serves as the Reduce function. This function takes two $\langle key, value \rangle$ pairs and applies Dempster's Combination Rule ($\oplus$), producing a single $\langle key, value \rangle$ pair. To further leverage the associative and commutative properties of DCR, a logarithmic TreeReduce is implemented rather than a traditional serial reduction. Ultimately, this improves the total Reduce latency from D to $\log(D)$.

3.3.3 Pthreads Implementation

To demonstrate the generality of our framework, we also implement it using pthreads as the underlying parallelism framework. While pthreads cannot directly apply to a distributed environment, it can be used in rapid prototyping situations to take advantage of all of the cores on a machine.

In a parent process, we load the input sample and the sources into global memory (e.g., via `mmap()`). We then create a new thread for each source (and thus each instance of the Map function). Each thread j can independently read from global memory to compute its own mass vector (relating source j to the input sample). We synchronize all threads with a barrier before, we hierarchically execute our TreeReduce function in multiple threads to fuse the mass vectors produced in the Map step. Each level of the TreeReduce hierarchy is similarly synchronized with a barrier before continuing. While threading is limited to shared memory machines, our approach can take advantage of thread pools (e.g., with BoostThreads [44])

in situations where the source count is extremely large.[2]

3.4 Evaluation and Discussion

In this section, we validate our claims that our distributed MapReduce implementation of template matching using DS theory. This Framework combines the handling of uncertainty using DS Theory with the efficiency and scalability of MapReduce. We seek to answer the following research questions:

RQ1 Does our framework scale as the number of sources increases and as the number of computational resources increases?

RQ2 Is our framework robust against noise that corrupts the data sources?

RQ3 Does increasing the number of sources reduce the amount of ambiguity and uncertainty in our final winning template?

For evaluation purposes, we consider the task of two-dimensional letter classification. In practice, there are many other domain-specific approaches for handwriting recognition (e.g., [36, 46,88]). We consider this application because it simplifies the generation of additional sources as well as the introduction and interpretation of noise and uncertainty in a way we can control. However, our approach can be applied to other domains as well. So long as the input dataset meets the structure criteria described in Section 3.2.1, DS mass values can be computed for propositions related to that dataset.

We consider two datasets: the Chars74K [147] dataset of computer fonts, and the MNIST [89] handwriting dataset. Chars74K contains images of numbers and characters using about 1000 different fonts—we consider each font to be a different source of information for our purposes. The MNIST handwriting dataset contains handwritten numbers from unique writers—here,

[2]Linux often soft-caps thread count to about 32,000; having more sources would be inappropriate for pthreads alone.

each writer is a different source. Varying amounts of salt-and-pepper noise [30] were added to the data source images. The noise is characterized by a noise density (ϵ). An ϵ of 0.5 equates to 50% of pixels randomly assigned either black or white. Figure 3.5 illustrates how this noise affects the template images. The prediction accuracy is measured as the fraction of trials that yielded a correct prediction over the number of total trials. A correct prediction is a True Positive (TP); an incorrect one is a False Positive (FP).

We discuss and present experimental results that address each of the research questions below. In each experiment, we also verified the results of our parallel or distributed algorithm against a previous independent serial implementation: in each trial our algorithm and the reference provided the same answer.

3.4.1 RQ1: Scalability

We address the scalability question by measuring total latency as a function of both the number of sources and the number of compute nodes available. We vary the number of sources D used when executing our framework, noting the starting and completion times to determine the total amount of time elapsed (with $n = 50$ repeated trials per measurement). We also repeat this experiment by varying the number of compute cores available to the framework. We normalize the results according to the longest time taken during serial operation.

Figure 3.3 shows the results of this timing experiment for our pthreads implementation. The Hadoop results are similar. Each line corresponds to a different number of cores available to the framework during execution. The lower the slope of the line, the faster it completes the matching task. We note a linear relationship in the number of sources—the total latency is directly related to the number of sources, indicating scalability. Additionally, by taking ratios of the slopes of the lines, we can compute the amount of speedup obtained by doubling the number of available cores. We find that doubling the number of available cores

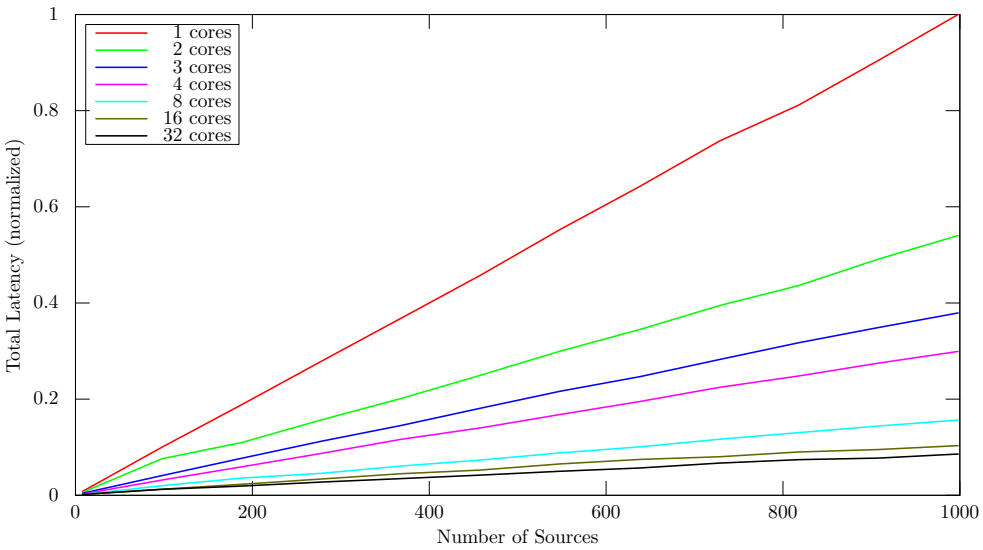
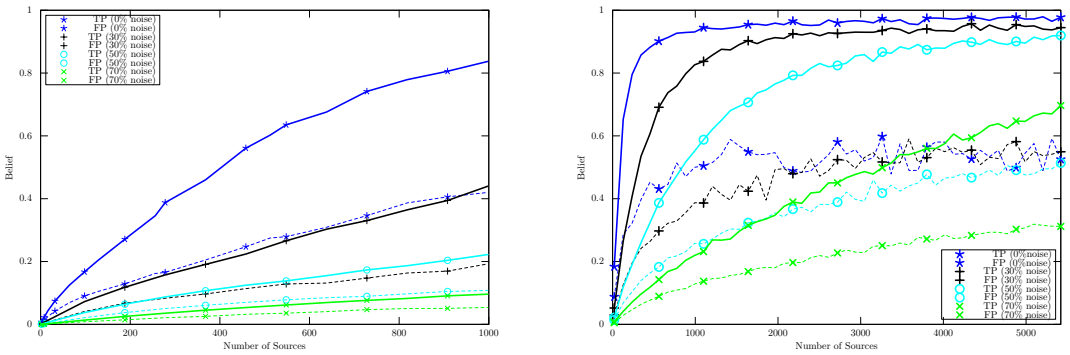


Figure 3.3: Total execution latency as a function of source count. The line slopes indicate linear scalability with additional computational resources.



(a) (b)

Figure 3.4: Belief vs. number of sources at different noise densities. 3.4a shows results for the Chars74K dataset, while 3.4b shows the MNIST results.

approximately yields a 1.9x speedup. We suspect that switching, staging, and networking overhead contribute to performance degradation at a high number of sources.

3.4.2 RQ2: Robustness against noise

We hypothesize that having many sources, even in the presence of noise, benefits our DS framework by allowing refinement of the uncertainty and DS masses. We executed our framework, over $n = 1000$ trials, with varying numbers of sources and measured the accuracy

of our predictions. Once the trail is complete, we select highest mass to predict in the input image. Figure 3.6 shows that when we consider a high number of sources (the top two lines), increasing noise does not degrade the accuracy of our framework. These results indicate that our highly-parallel MapReduce framework allows us to leverage a high number of sources to make robust decisions in the presence of large, corrupted information sources.

3.4.3 RQ3: Reducing uncertainty with more sources

One key advantage of DS Theory for matching is that each classification answer comes equipped with an uncertainty value for your decision. Scientists can make use of quantified uncertainty to help interpret results: high belief and lower uncertainty in a prediction is valued. We demonstrate that increasing the number of sources results in higher belief and lower uncertainty for predictions from our framework, even in the presence of noise.

Figure 3.4a shows belief as a function of sources and noise in the Chars74K dataset, broken down by FP vs. TP predictions. Consider the lines corresponding to 0% noise: at 200 sources, correct answers come with a belief of about 0.25 while incorrect answers come with a belief of about 0.15: a small margin. By contrast, at 1000 sources, the margin has spread to about 0.8 vs. 0.4. The trends for data with reasonable amounts of noise (e.g., 30%) are similar. Figure 3.4b shows the same results using the MNIST handwriting dataset. Again, at 1000 sources, the margin for noiseless predictions 0.9 vs. 0.45.

Additionally, the data suggest that adding many sources in the presence of noise helps the system gain more confidence in its predictions. In other words, our MapReduce framework

Figure 3.5: A template image shown with varying noise densities ϵ which represents the fraction of pixels that are impacted by noise.

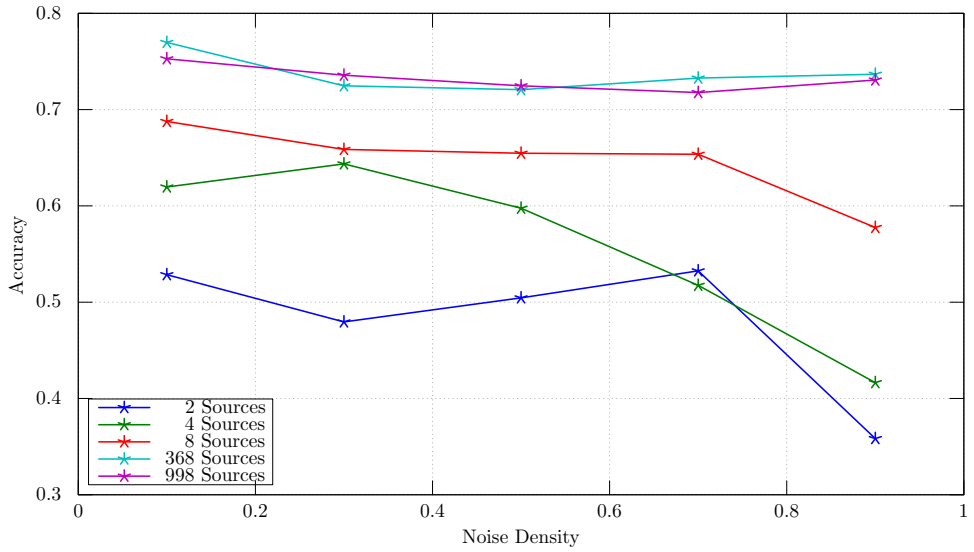


Figure 3.6: Prediction accuracy as a function of noise density and source count. The horizontal lines at the top demonstrate that many sources allow for high accuracy even in the presence of noisy data.

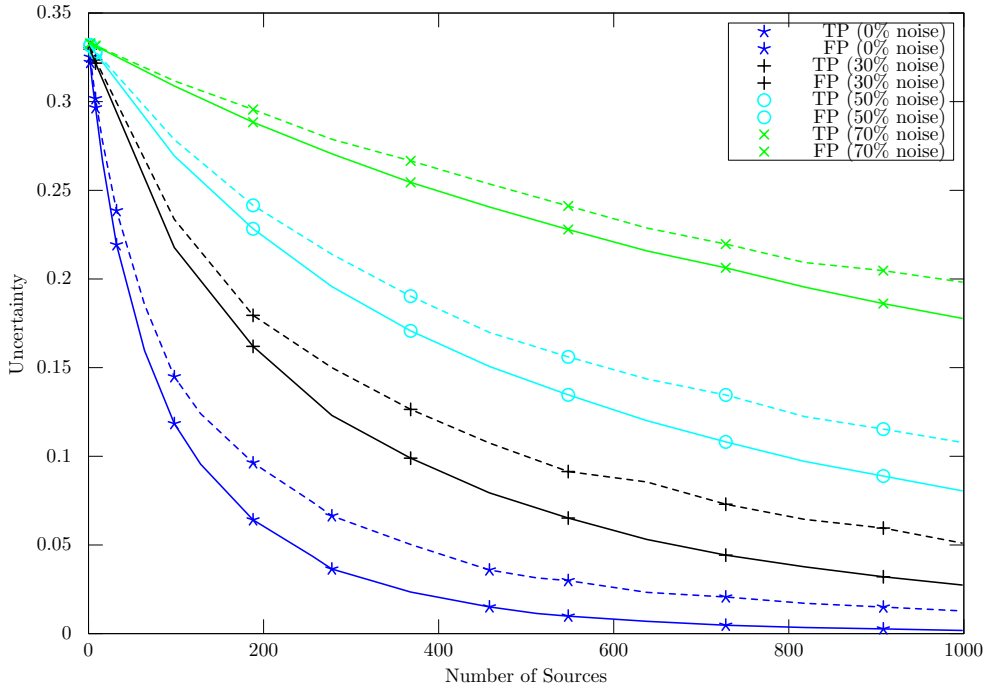


Figure 3.7: Uncertainty vs. number of sources at different noise densities in the Chars74K dataset.

is capable of lowering uncertainty in the presence of noisy data by leveraging high amounts of noisy sources. In particular, Figure 3.7 shows the uncertainty for correct and incorrect predictions at varying levels of noise. These results show how increasing the number of sources leads to a more robust prediction.

We note that our current framework considers only singleton cases—that is, single hypotheses. Restricting attention to singletons admits a lower complexity class for the TreeReduce function ($\mathcal{O}(n^2)$a for Algorithm 2). Addressing non-singletons increases the complexity of the TreeReduce function as it requires computing matrix cross-products ($\mathcal{O}(n^3)$). We leave the non-singleton case for future work, but note that it creates further potential for exploiting the massive parallelism we achieved to reduce uncertainty and resolve ambiguities in a practical runtime with a large number of sources.

3.5 Conclusion

In this paper, we present a novel MapReduce framework for improving uncertainty in DS template matching. By taking advantage of the associativity and commutativity of Dempster's Combination Rule, we construct a highly scalable framework for fusing evidence in the form of DS mass vectors from many sources of information.

We implement two prototypes of our framework using open source software—one using Hadoop, and the other using pthreads. The pthreads implementation allows rapid development and debugging cycle where testing is required before full-scale deployment. The Hadoop implementation allows rapid deployment since there is very little effort required to deploy a MapReduce algorithm on a Hadoop cluster once written.

We demonstrate the scalability of our system by comparing total execution latency as a function of both source count and the computational elements available. Our system yields about a 1.9x speedup every time the number of cores is doubled. This linear scalability

allows practitioners to consider more sources,yielding benefits for accuracy, belief, and uncertainty in the face of noise. We show that providing many sources to our framework allows it to maintain a high level of accuracy even when source data is affected by noise, while maintaining practicality. Finally, we show that our system improves belief and uncertainty when considering many sources.

It is a very sad thing that nowadays there is so little useless information.

Oscar Wilde

4

Handling Time Series Data: ECG Corruption

4.1 Introduction

Intensive Care Units (ICUs) have started looking toward the future with real-time data predictive models to aid physicians at the bedside. These algorithms are intended to detect latent physiological indicators that provide information to the physician about the patient's trajectory and imminent risk [40]. The most commonly used information for determining a patient's trajectory is the analysis of electrocardiogram (ECG) signals. Methods such as Heart Rate Variability (HRV) and Heart Rate Complexity (HRC) have been demonstrated

to be predictive of numerous types of physiological ailments such as myocardial infarction, mortality, autonomic responses, and hypoglycemia, [72, 76]. The measurement of HRV and HRC rely on calculating the time delay between heart beats; this term is clinically known as the R-R intervals. In order to calculate R-R intervals, it is critical to detect the QRS complexes within the ECG signal, which corresponds to the depolarization of the ventricles in the heart. Thus, predicting potential complications for a patient using HRV and HRC is dependent on accurate and precise detection of QRS complexes in the ECG signal.

Prior Work. Standard benchmarking methodologies utilized for QRS detection are cross-correlation and non-syntactic algorithms [86,115]. These approaches are well developed, easy to implement, and have been around for decades. Cross-Correlation examines how well the shape of the waveform coincides with a particular template [150]. The basic idea is that these approaches acquire a template and pass the template through the signal searching for a set threshold in the correlation coefficient in order to assume a match [82]. Non-syntactic algorithms manipulate the signal to form peaks at the locations of the QRS complexes and attenuate the other waveforms in the signal. A threshold is set for the height of the peaks to determine which peaks indicate a QRS complex; however, when noise is introduced in the signal, this threshold fails to provide information on the quality or context of the prediction. These methods raise three interesting questions: 1) How can we determine the quality of the QRS complex prediction, 2) what can be done if the ECG lead or multiple leads are corrupted or inconsistent in their predictions, 3) can an algorithm address these questions using little or no prior knowledge?

Challenges. The current discussed QRS detection paradigms pose a challenging problem because the real-time data from the ICU faces corruption and a multitude of noise artifacts. These problematic conditions are attributed to missing data, physical activities, muscle artifacts, electromagnetic interference and baseline wandering [55]. This noise is typically combated by attempting to manually or dynamically choose which lead (source) has the

best signal-to-noise ratio (SNR), but this practice still detects QRS complexes from only a single source of information [81, 82, 116]. In order for this work to be practical, real-time implementation of these models is necessary. Thus, the approach must account for ingesting data that is continually streaming regardless of the quality and reliability of the information.

Insights. There are a number of factors that can contribute to the corruption of an ECG lead. However, a majority of these QRS detection algorithms require a supervised approach to account for their parameters (thresholds, windowing lengths, filtering). Furthermore, these algorithms continue to analyze information from only a single lead for QRS detection. Dempster-Shafer (DS) Theory is an approach that allows us to quantify uncertainty and fuse data to then refine uncertainty from imperfect data from multiple sources (such as conflicting source information or sources reporting similar information) [134]. DS Theory offers combination rules that fuse multiple evidence sources into a single set of hypotheses. Fusing sources captures contextual considerations, such as conflicts between sources, corrupt information, uncertainty, source reliability, and accuracy [134]. Recent work has developed a method for quantifying uncertainty in a set of correlation coefficients using a DS framework [111]. We propose building upon this cross-correlation method to address the three questions posed in the Prior Work section.

The proposed DS theory-based approach for template matching addresses the issues of multiple templates and ambiguity. Using this methodology, we are able to capture different levels of ambiguity that occur when sources report either the same or conflicting information. If one ECG lead indicates QRS detection while another lead conflicts with that prediction, the uncertainty in the model should increase. Likewise, if there is no conflict (both sources report similar findings), then the uncertainty should decrease. Therefore, we sought a DS approach with the ability to use all available ECG leads, appropriately deal with corruption, conflict, and uncertainty, and use little to no prior knowledge.

Contributions. This DS framework is then designed to be applied to quasi-periodic signals (such as ECG data [55]) using a set of cross-correlated ECG templates as a form of evidence. The contributions of this work are:

1. We develop an uncertainty and 'probability' value at each point in time of the ECG signal to provide information on the quality of each QRS prediction.
2. We develop a template cross-correlation approach using DS theory to fuse ECG leads to overcome corruption.
3. We demonstrate that the only prior information used was a single set of templates for each ECG lead.

Cross-Correlation.

The central idea of cross-correlation is to examine how well the shape of the waveform coincides with a particular template [150]. This correlation coefficient is assigned to the template for a particular point in time within the ECG signal, providing evidence of a match. A correlation coefficient of $\rho = 1$ demonstrates a perfect match between the template and the signal, a coefficient of $\rho = 0$ indicates no match, and when $\rho = -1$ the topology is the same but the two signals are out of phase or negatively correlated. A recently explored method reduces the original template into multiple components which searches for specific components within the ECG waveform [86].

Non-Syntactic.

The Non-Syntactic approaches tend to have two stages, a pre-processing stage and a decision stage [86,116]. There are many designs for a non-syntactic approach; however, [116] has one of the most accepted approaches for QRS detection. The pre-processing stage filters the signal to remove baseline wandering and increase the SNR. The signal is then differentiated

to provide slope information of the QRS complex and rectified using a squaring function [116]. A moving window integration is applied to obtain features of the QRS and essentially defines how many peaks are produced from a QRS complex. Fiducial Marks are then placed on the rising edges to indicate the locations of the assumed QRS complex.

4.2 Methods

The overview of the methodology is to construct a template for each lead, capturing the inherent characteristics of the topology for an ECG lead. This template is partitioned to compartmentalize critical waveforms of the ECG signal (P-wave, T-Wave, QRS Complex). These partitioned components of the template are then cross-correlated with their associated lead producing correlation coefficients for each component over the time span of the signal. Using these correlation coefficients for each lead, we apply the DS Framework to the correlation coefficients to produce DS Masses. We then apply Dempster Combination Rule (DCR) to analyze conflict between leads and refine our detection by evaluating the belief and uncertainty of the propositions. The proposed template fusion algorithm and the non-syntactic model, otherwise known as the Pan Algorithm, described by [116] are compared to evaluate performance under corrupted signal conditions.

4.2.1 Template Framework.

Creating a proper template framework is one of the cornerstones of this algorithm. First, the template must accurately develop and compartmentalize the waveforms of interest that comprise the ECG signal. Moreover, for unsupervised detection regarding the partitioning, it is crucial to know where the compartmentalized critical waveforms are located within the master template.

Master Template.

We first construct a *master template* for each lead using a sample of the first 20 seconds of data. The template is developed by segmenting the raw ECG waveform over multiple iterations. In the time-domain, the ECG waveforms are optimally aligned using cross-correlation coefficients and the lag before averaging all of the time series together, a process similar to [59, 77]. This provides a stronger model by accentuating prevalent waveform features of the signal and filtering out the noise.

Compartmentalizing Template Components.

A multi-component template method is developed around the master template for each lead. This method aids us in avoiding previous thresholding techniques in the detection process [116], assessing conflict between competing components, and incorporating a DS framework using this conflict between competing coefficients [109]. To incorporate this DS Theory into a multi-component template method, the compartmentalization within the master template needs to be formalized with meaningful evidence to satisfy the FOD framework. Therefore, we propose a windowing scheme that will capture a set of *critical events* to be $\mathbf{K} = \{\kappa_1, \dots \kappa_D\}$, of the ECG signal, where we consider a single critical event in the set as κ_i. For our purposes, we defined these critical events as the P-Wave, T-Wave, and QRS Complex with a window length n. In addition, a fifty percent window overlap scheme is implemented to capture transitions from one criteria event, κ_i, to the next criteria event, κ_{i+1}. Thus, we form additional events called *transitional events*. The number of events, N, using a fifty percent windowing scheme is defined as

$$N = |\mathbf{ev}| = 2|K|. \tag{4.1}$$

Each event has fixed windowed length, n, segmenting the master template into $N/2$ critical events. Therefore we produce N events, ev_i, which are considered as a doubletons, due to the overlapping windows having intersections within the finite set, shown in Figure 4.1. Formulating this windowing framework will later enable us to apply DCR as a fusion method by treating the additional leads as evidence sources by accounting for these intersections, which will aid in refining our decision.

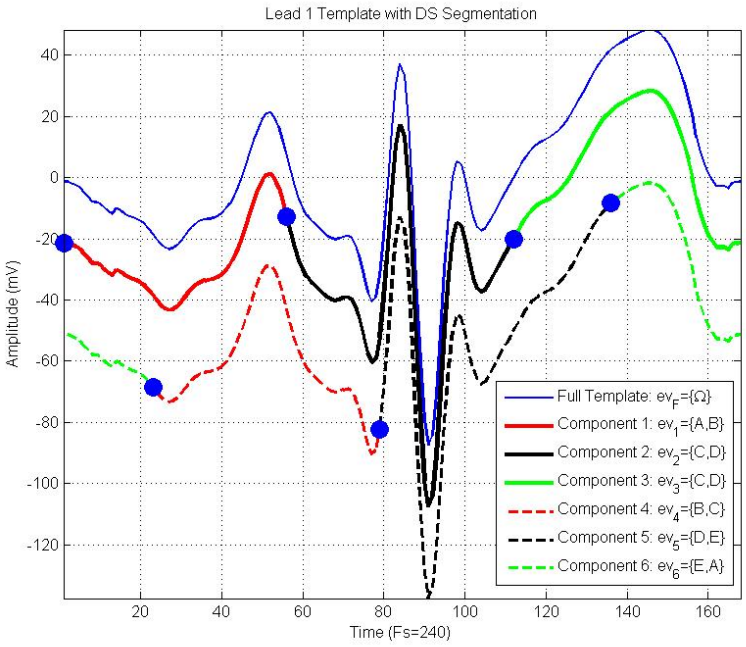


Figure 4.1: Compartmentalized Evidence Templates

4.2.2 Templates to DS Masses

Cross Correlation of Components.

In the implementation of any template matching scheme, the evidence that is obtained informs us of how well a segmented component or event matches a specific time point within the signal. We quantify this evidence for each ECG lead by calculating a cross-correlation coefficient using Equation 4.2 for every time instance, t, in the signal. We quantify the

strength of the match between the two signals with a correlation coefficient, which gauges the 'closeness' or similarity between two signals. This quantification can be expressed as:

$$\rho_{\mathbf{x},\mathbf{y}}(\tau) = \frac{\sum_{i=1}^{n}(x_i - \overline{\mathbf{x}})(y_{i+\tau} - \overline{\mathbf{y}})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{\mathbf{x}})^2 \sum_{j=1}^{n}(y_{j+\tau} - \overline{\mathbf{y}})^2}}, \tag{4.2}$$

where the correlation coefficient $\rho \in [-1, 1]$, $\overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\overline{\mathbf{y}} = \frac{1}{n}\sum_{j=1}^{n} y_j$ denote the 'sample' means of the real-valued time series data vectors $\mathbf{x} = [x_1, \ldots, x_n]^T$ and $\mathbf{y} = [y_1, \ldots, y_n]^T$, respectively.

Correlations to DS Framework.

Evidence is formed around each indexed time point, where we produce a vector of normalized correlation coefficients, $\mathbf{V}$, for each ECG Lead. Within $\mathbf{V}$, we have N elements, where each element is a coefficient that represents the correlation of a specific event, ev_i, to the signal at time t. Therefore, each ECG lead can be considered a separate, independent source of evidence, producing a set of N correlation coefficients. The correlation coefficients are developed into a DS Frameworks by analyzing the conflict and penalizing the propositions that are weak within the set described in [111]. This framework is appropriately suited for the singleton cases. However, due to the size of set (n) and uniqueness of the QRS in the set, this framework should roughly capture the conflict appropriately. Applying the windowing scheme discussed above for a single lead, a vector of normalized correlation coefficients are generated, thus taking the form $\mathbf{V} = \begin{bmatrix} V_1 & V_2 & \cdots & V_N \end{bmatrix}^T$, where $V_i \in [-1, 1]$ and V_i, $i = 1, \ldots, N$, denotes only the positive normalized correlation coefficient between a corresponding windowed event within the ECG signal and the i^{th} event in the template data set. We consider negative correlations events that have already passed in time because of the phase change in the signal. We follow [109]'s DS framework for correlation coefficients to capture

the overall magnitude of each element in $\mathbf{V}$ by utilizing a weighting strategy as

$$\begin{aligned}\Delta\mathbf{W} &= \Delta\mathbf{V} \circ \mathbf{J}_{NN}\mathbf{D}_N \\ &= \begin{pmatrix} V_1\Delta V_{11} & V_2\Delta V_{21} & \cdots & V_N\Delta V_{N1} \\ V_1\Delta V_{12} & Ev_2\Delta V_{22} & \cdots & V_N\Delta V_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ V_1\Delta V_{1N} & V_2\Delta V_{2N} & \dots & V_N\Delta V_{NN} \end{pmatrix},\end{aligned} \tag{4.3}$$

where $\circ$ denotes the matrix Hadamard product, $\Delta V_{ij} = (V_i - V_j) \in [-1, +1]$, $\forall i, j \in \overline{1, N}$. $\mathbf{J}_{NM}$ denotes the $N \times M$ matrix with each entry being 1 and $\mathbf{D}_N = diag\,[V_1, V_2 \cdots, V_N]$ denotes the diagonal matrix with the diagonal entries being $\{V_1, V_2, \cdots, V_N\}$. The columns of $\Delta\mathbf{W}$ compares the distance of an element V_i with all the elements in $\mathbf{V}$ and weights V_{ij} with V_i. This weighting of V_i could be thought of as an indication of the "strength" of the corresponding prototype being a match. Thus, the *column weight*, the summation of the column vectors, informs us how each element in $\mathbf{V}$ is different from the other elements and how strongly it matches a specific event. The *column weights*, C_i, are calculated as

$$\mathbf{C} = \begin{bmatrix} C_1 & C_2 & \cdots & C_N \end{bmatrix}^T = (\mathbf{J}_{1N}\Delta\mathbf{W})^T, \tag{4.4}$$

where $C_i \in [-(N-1)/4, (N-1)]$, $\forall\ i \in \overline{1, N}$.

A reduction of propositions is done by applying a constraint to the column weights vector, $\mathbf{C}$. This determines the propositions that are assigned DS masses, known as a focal element. The criterion to determine the number of focal elements, P, is the number of elements whose C_i is positive. This determines the propositions that are assigned DS masses, known as a focal element. The criterion to determine the number of focal elements, P, is the number of elements whose C_i is positive. The mass measure vector $\mathbf{H} = [H_1, H_2 \cdots, H_N]$ is defined as $H_i = \frac{C_i + |C_i|}{2}$. Thus, $H_i \in [0, (N-1)]$, $\forall\ i \in \overline{1, N}$. The mass measure vector's elements are

calculated to DS masses by

$$m(A) = \begin{cases} 1 - \left(\frac{S}{(N-1)P}\right), & \text{for } A = \Theta; \\ \\ H_i \left(\frac{1-m(\Theta)}{S}\right), & \text{for } A = H_i, \end{cases} \tag{4.5}$$

where $\Theta = \{V_1, V_2, \cdots, V_N\}$ and $S = \sum_{i=1}^{N} \mathbf{H}_i$ is the the FoD consisting of the propositions H_i, $i \in \overline{1, N}$.

Template Fusion.

The fusion of evidence sources (ECG Leads) allows us to refine our supports and our uncertainty. Thus, as additional information (ECG sources) is provided, we can refine our epistemic uncertainty by quantifying conflict within a single source and the conflict between sources. This ability to handle conflict allow us to work with imperfect data in an effective and more intuitive manner [92]. The fusion method utilizes the *Dempster's Combination Rule (DCR)* [19]:

$$m(A_i) = \frac{\sum\limits_{A_p \cap A_q = A_i \neq \emptyset} m_1(A_p)\, m_2(A_q)}{1 - \sum\limits_{A_p \cap A_q = \emptyset} m_1(A_p)\, m_2(A_q)}, \tag{4.6}$$

where the evidence provided by the mass functions m_1 (ECG lead 1) and m_2 (ECG lead 2) are combined to get the fused mass function m, denoted as $m = m_1 \oplus m_2$.

4.3 Evaluation and Discussion

This section evaluates and validates the performance of the proposed fusion algorithm using the following three research questions:

RQ1 Does our template fusion framework provide and capture contextual meaning about the quality of the information and the QRS prediction?

RQ2 Does our framework appropriately handle corrupted and inconsistent leads?

RQ3 Does this template approach require a large amount of historic information for its decision or training processes?

The non-syntactic model described by [116] and the proposed fusion model were evaluated using the annotations and data from the MIT-BIH Normal Sinus Rhythm Database [57]. The traditional template methodology was not used in the comparison since the proposed methodology is built upon this traditional approach and therefore would be expected to produce similar, if not better, results.

These signals are a two lead extended ECG recording from subjects with no significant arrhythmias that were taken at the Beth Israel Deaconess Medical Center, with annotated QRS complexes. Each original signal, X_s, from the Beth Israel Deaconess Medical Center was not initially filtered to enhance its SNR. Furthermore, Gaussian additive noise, N_g, was introduced by $X_c = X_s + N_g$, producing a further corrupted signal, X_c. The original signal, X_s, was bandpass filtered to produce a further degraded signal, X_E, and assess the SNR by $(E[X_E^2]/E[X_c^2])$ for quantifying the signal quality.

The performance was calculated using the number of false positives, F_P and false negatives, F_N. False positives occur when the method indicated that a QRS complex occurred but it actually did not. The number of false negatives is when a QRS complex was not detected when one actually did occur. The QRS failure rate is defined by,

$$\epsilon = \frac{F_P + F_N}{T_{QRS}} \tag{4.7}$$

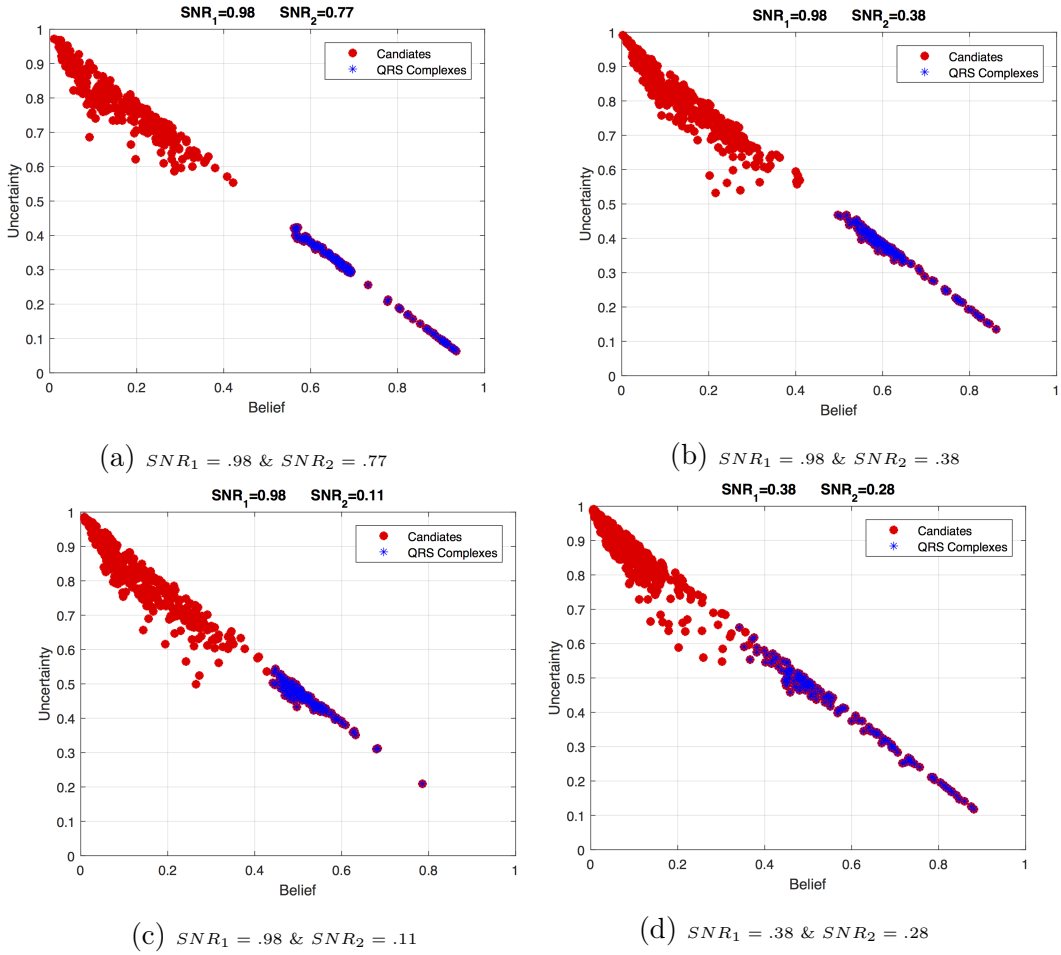
where T_{QRS} is the total number of QRS complexes in the signal [116]. Although, the precision and recall metric are typically used in analytics, this domain has used this metric for decades

[116]. Since an hour of data would produce over a million samples, but only 4000 QRS complex. The goal is to strictly highlight the misses and detections of the QRS complex, rather than being over shadowed by total number of samples causing a small change in the performance metric. We discuss and present experimental results that address each of the research questions below. In each experiment, we address a research question demonstrating our findings associated with that specific question.

4.3.1 RQ1: Capturing belief and uncertainty

We address the question regarding capturing contextual meaning by running various experiments with different SNRs providing a graphical representation of belief vs uncertainty for QRS candidates. Figure **??** depicts the calculated belief and uncertainty for instances in the signal that indicated a possible 'candidate' of a QRS complex. A 'candidate', in red, is indicated by having the highest belief within the set after applying DCR for a particular time instance. The actual QRS complexes, in blue, are plotted over the candidates to demonstrate their belief and uncertainty relative to the other candidates. Note, if we took all the proposed candidates our error rate would be very high. However, we can note the linear separability of the two classes even when the SNR is below common standards (SNR less than 1), seen in Figure 4.2c. As the SNR degrades further, the class separation becomes more complex.

We can note the capturing of contextual meaning of the quality of the information by how the uncertainty graphically increases and belief decreases as SNR further degrades. Similar quantitative results can also be observed in Tables 4.1 and 4.2 discussed in research question two, where the average belief (Bel_μ) and uncertainty (Ω_μ) were listed.



(a) $SNR_1 = .98$ & $SNR_2 = .77$

(b) $SNR_1 = .98$ & $SNR_2 = .38$

(c) $SNR_1 = .98$ & $SNR_2 = .11$

(d) $SNR_1 = .38$ & $SNR_2 = .28$

Figure 4.2: Uncertainty vs Belief

4.3.2 RQ2: Robustness against noise

Here we address whether our framework appropriately handles corrupted and inconsistent leads. Regardless of what pre-processing methods are utilized, imminent sporadic events of noise will always occur, which drastically affect the calculations for HRC and HRV and have downstream effects on predictive algorithms.The simulated noise that was added to these unfiltered signals was done to demonstrate the effectiveness of detection when these sporadic events occur and when pre-processing the signal fails to obtain adequate SNR. We also hypothesize that the proposed DS template algorithm is more effective than the benchmark non-syntactic methodology signal in corrupted situations. For simplicity, we took an arbitrary classification stance where, in order to classify a candidate as a QRS

complex, it must have a $Bel > .4$ and $\Omega < .55$. However, it is recognized that optimizing these constraints on these given conditions would provide superior results for DS template method.

Table 4.1: Corruption: MIT Data Set 16265 ($T_{QRS} = 253$)

Corruption		Pan Algo.		D.S. Fusion Algo.		
SNR_1	SNR_2	ϵ_{S_1}	ϵ_{S_2}	Bel_μ	Ω_μ	$\epsilon_{F_{12}}$
.96	.74	0	1.18	.66	.32	1.18
.47	.66	129.6	1.18	.64	.34	0
.29	.29	146.2	62.8	.58	.40	24.5
.13	.11	203.2	221.3	.52	.46	80.2

The simulation in Table 4.1 and the simulation in Table 4.2 were done for a small segment of the time series, associated with Data Sets 16125 and 16272, respectively. Equation 4.7 was applied providing the performance for the fusion algorithm ($\epsilon_{F_{12}}$) and the non-syntactic algorithm for each ECG lead (ϵ_{S_1} and ϵ_{S_2}). The two QRS algorithms have comparable performance at the higher SNR cases; however, the non-syntactic algorithms' performance deteriorates as corruption of the signal occurs. More interestingly, the proposed DS template algorithm is able to handle inconsistent leads. This can be seen when the performance of the non-syntactic algorithm deteriorates on a single lead or on both leads, and the marginal error for the DS template method is substantially smaller than the marginal errors for the non-syntactic algorithms.

Table 4.2: Corruption: MIT Data Set 16272 ($T_{QRS} = 164$)

Corruption		Pan Algo.		D.S. Fusion Algo.		
SNR_1	SNR_2	ϵ_{S_1}	ϵ_{S_2}	Bel_μ	Ω_μ	$\epsilon_{F_{12}}$
.70	.78	1.83	31.7	.76	.23	1.83
.64	.44	1.83	39.0	.74	.25	1.23
.52	.39	3.65	47.0	.70	.29	1.22
.40	.32	58.5	145.7	.66	.32	5.49

The experiment was then extended for approximately one hour's worth of data, providing a more accurate measure of performance, Table 4.3. We can note that for this specific example of the fusion algorithm performance, $\epsilon_{F_{12}}$, it outperformed the non-syntactic algorithm in

both leads, S_1 and S_2. Obviously, S_2 had a higher QRS failure detection percentage of 54.5%, since the SNR for that particular signal was much lower. However, it is worth noting that we were able to obtain an improved detection using the degraded signal from S_2. This ultimately cuts our error by half.

Table 4.3: Method Comparison: MIT Data Set 16265

M	SNR	T_{QRS}	F_P	F_N	$\#\epsilon$	ϵ
F_{12}	.82/.31	4647	25	51	76	1.6%
S_1	.82	4647	128	19	147	3.2%
S_2	.31	4647	1680	857	2537	54.5%

4.3.3 RQ3: Little Prior Knowledge

The proposed approach requires very little to no information, where no training is required. The amount of information required is equivalent for the traditional template algorithm for a single ECG lead. which is important for the use of algorithms in clinical setting. Algorithms that are reliant on large data sets for clinical research can contain hidden biases that might restrict the algorithm's accuracy and application on heterogeneous data sources. Hence, this design was built around the idea of template matching. However, more data with respects to the number of ECG leads would further refine the support of the DS framework for further enhanced performance.

4.4 Conclusion

Future work will focus on optimization methods for defining the proper constraints for the belief and uncertainty values to enhance classification, and on producing a better framework to handle the doubleton case. We have demonstrated strong potential for further improvement, which could include utilizing more ECG leads as evidence sources, parallelizing the process [112], and detection of other cardiac waveforms using this template paradigm.

The proposed DS fusion multiple template approach over multiple leads has demonstrated strong evidence for being a superior detection methodology when signals are corrupted with noise. This work demonstrates that using little to no prior knowledge, contextualized QRS detection provides powerful evidence to support a decision. Unlike the non-syntactic and conventional cross-correlation methods, where an arbitrary threshold is set and requires adjusting, this contextual meaning provides information on how corrupt the signals are by quantifying the uncertainty that is associated with the present conflict between competing templates. More importantly, unreliable information from corrupted signals still provides value to the detection process and are leveraged by addressing the conflicting information sources across additional leads.

Part II

Handling Uncertainty for Predictive Problems for Systems with Diverse Information Sources

Part II of this dissertation analyzes key components of where uncertainty enters a predictive system for detecting human performance error in pilots. Part II is also partitioned into three individualized papers that make up an additional three chapters (Chapters 5, 6, and 7). The first two chapters (Chapters 5 and 6) focus on feature engineering and Chapter 7 focuses on modeling paradigms. The two feature engineering chapters do not directly discuss uncertainty and its downstream effects on model performance, but rather core featuring engineering methods to reduce uncertainty that is introduced into our predictive models. This is demonstrated in Chapter 5, “Evaluating Sensor Metric Uncertainty Due to Noise," by examining how noise alters reported metrics or features through a large scale Monte Carlo Simulation approach for heart variable metrics. The simulation approach presented in this chapter can be utilized to evaluate other metrics that are vulnerable to real-time noisy conditions. Chapter 6, "Featuring Engineering to Provide Additional Model Dimensionality," focuses on engineering new features that capture additional information (that previous features do not) to reduce epistemic uncertainty. The second takeaway from this work is that the fundamental design for features is not dependent on changes from technical variability. This variability could be caused by a number of reasons such as subject variability, signal drifting, sensor problems, conductivity of electrodes, etc. From a data science perspective, standardization is usually viewed as a means for trying to control input from technical variability that goes into the model, but maintains biologically driven global variation. However this cannot be fully controlled, and feature engineering design should incorporate these considerations into the design process. We demonstrate this by examining electroencephalogram intensity, which varies from subject to subject (due to the technical variability in the conductance of an electrode placed on the scalp of a subject), through features that are linked to timing rather than conductance. Chapter 7, "Naive Adaptive Probabilistic Sensor Fusion," then utilizes these features and others to develop ML paradigms to detect human performance. This work explores novel designs that address and quantify model uncertainty.

Errors using inadequate data are much less than those using no data at all.

Charles Babbage

5

Evaluating Sensor Uncertainty due to Noise

Heart rate complexity (HRC) is a proven metric for gaining insight into human stress and physiological deterioration. To calculate HRC, the detection of the exact instance of when the heart beats, the R-peak, is necessary. Electrocardiogram (ECG) signals can often be corrupted by environmental noise (e.g., from electromagnetic interference, movement artifacts), which can potentially alter the HRC measurement, producing erroneous inputs which feed into complex decision models. Current literature has only investigated how HRC is affected by noise when R-peak detection errors occur (false positives and false negatives). However, the numerical methods used to calculate HRC are also sensitive to the specific location of the

fiducial point of the R-peak. This raises many questions regarding how this fiducial point is altered by noise, the resulting impact on the measured HRC, and how we can account for noisy HRC measures as inputs into our decision models. This work uses Monte Carlo simulations to systematically add white and pink noise at different permutations of signal-to-noise ratios (SNRs), time segments and HRC measurements to characterize the influence of noise on the HRC measure by altering the fiducial point of the R-peak. Using the generated information from these simulations provides improved decision processes for system design which address key concerns such as permutation entropy being a more precise, reliable, less biased, and more sensitive measurement for HRC than sample and approximate entropy.

5.1 Introduction

The growing field of physiological telemetry systems is a continual source of research in numerous domains such as health care, aerospace, and nuclear power (e.g., [66, 100, 105]). These telemetry systems are designed to provide informative decision support and predictive analytics to gain physiological insight on health monitoring [100] and cognitive states identification (e.g., cognitive workload) [28, 31, 159]. One major physiological component that is leveraged in order to gain these insights is the sympathetic nervous system, which is the part of the autonomic nervous system responsible for monitoring and adapting to stress imposed on the body [71, 121, 130]. An extended increase in arousal of the sympathetic nervous system has been associated with a decrease in performance [148], increased cognitive workload [11, 35], and deteriorating health for patients in a clinical setting [146, 166].

One of the most relevant and informative measures associated with sympathetic response is heart rate complexity (HRC) [146]. HRC algorithms have been developed as an integral latent physiological indicator being utilized in numerous studies assessing cognitive workload [17, 84, 132], a patient's physiological health and imminent risk [76, 100, 108], and

analysis of depression [93]. HRC is often operationalized in the form of entropy measures which are calculated using a person's heart rate variability (HRV) [100]. These measures require accurate, precise detection of individual heart beats. While there are many methods to measure heart rate, such as electrocardiograms (ECGs), photoplethysmograms, and optical heart rate monitors, ECGs are the only method that provide information about the QRS waveform - the so-called combination of the three deflections typically seen in the ECG - which corresponds to the depolarization of the heart ventricles (a single heart beat). These QRS waveforms are critical to precisely calculating R-R intervals and thus HRV. HRV is then used to quantify the fluctuations in HRC. Thus, ECG telemetry stands as the only proven method for calculating HRC accurately. For these predictive technologies to provide insights, computational algorithms are required to examine variations from normal physiology [165]. Since many of these technologies require consistent monitoring, intermittent noise is inevitably introduced into the system. Thus noisy occurrences and detection of physiological anomalies generate ambiguities, leading to false detection, inaccurate decision support, and alarm fatigue [37, 99, 152]. However, much of the previous, original work surrounding HRV detection was developed on cleaned, retrospective datasets [58, 100, 108, 125]; thus, we cannot assume that complexity algorithms perform appropriately on noisy and corrupt data.

Understanding the robustness of HRC measurements under noisy conditions allows for corrective computational approaches and system design [99]. These corrective measurements are becoming a paramount objective as systems are expected to work in real time which leads to high risk of signal corruption. In this investigation, we aim to fill this research gap by evaluating complexity algorithms under noisy and corrupted conditions.

Prior Work. The majority of work on HRC was conducted by Costa et al. [39], based on Pincus's work with approximate entropy [119]. These ideas, concepts, and applications used to evaluate heart rate entropy spawned a multitude of measures [125,126]. The two most uti-

lized complexity measures since are approximate and sample entropy. In line with Pincus's and Costa's prior work, these complexity measures have been used on retrospective data to demonstrate predictive model feasibility and practicality in classifying patients in clinical settings [39]. While extensive studies have been conducted on various HRC methods and applications in the past few decades, few studies exist that evaluate the influence of noise and erroneous behavior on these measures. These works have demonstrated that missed detection of QRS complexes (heart beats) could drastically alter downstream entropy measurements [78, 99]. This was demonstrated by examining the effects of corruption through the false positive (FP) and false negative (FN) rates of the detection algorithms rather than looking at how the SNR itself alters entropy [78, 99]. This was done to control for the false positive and false negatives in the detection algorithms. These approaches manipulate the HRV signal through the concatenation of sequences or random down and up sampling, which changes the entropy measurement associated with the HRC [78, 99]. Although robust and accurate QRS complex detection algorithms are highly desirable and have been achieved through various types of machine learning approaches [108], QRS detection is not the only avenue through which erroneous entropy measurement can occur. ECG sampling frequency and R-peak interpolation have demonstrated importance as methodological considerations in obtaining consistent HRV signals [48]. These concerns underline the importance of the fiducial marker for the R-peak and its relationship to producing reliable HRV measurements, thus impacting the HRC. This can be depicted in Figure 5.1, an illustration where the ECG signal R-peak has an arbitrary confidence interval which therefore induces uncertainty in the HRV signal below. Noise plays an integral role in where the fiducial marker for the R-peak is placed, which affects downstream applications. However, neither the extent to which noise alters HRV and HRC measurement through changes in the fiducial marker nor the downstream implications of these fiducial markers on statistical models have been characterized. Ultimately, current literature's focus has been on the advancement of complexity algorithms and their utility for classification, rather than understanding how and when these algorithms

fail. Developing this understanding is critical for appropriate system design in preventing alarm fatigue, erroneous prediction, and decision support systems [99, 152].

This work raises six relevant research questions (RQs): 1) How does the SNR and the color of the noise alter the fiducial marker of the R-peak? 2) Is any particular entropy measurement more precise under corruption than others? 3) Does signal length contribute to a more robust measurement in the presence of corruption? 4) Can a precise entropy measurement still be statistically sensitive in differing HRC dynamics expected in study populations? 5) If the fiducial point is altered by noise, how does this effect the direction of the measured entropy calculation? 6) How much impact does noise and the fiducial marker downstream have on the statistical implications of a study?

Challenges. Due to the multiple stages involved in the process of calculating HRC, it is difficult to pinpoint exactly where noise alters an entropy measure's reliability. Noise can alter HRC reliability through numerous avenues such as False Positives for the QRS detection, false negatives of the QRS detection, or the QRS waveform fiducial point shifting. However, the robustness of current methods has only been evaluated by resampling HRV signals, thus only examining the false positive and false negative rate [78–80]. Therefore, the fundamental issue of how corruption in an ECG signal alters the HRV signal through slight shifts in fiducial points and QRS waveform timing has been neglected and is difficult to capture. Additionally, the numerous types of complexity measures produce different outputs with different scales, further complicating the validation process and comparative analysis.

Insight. A systematic process of validation in which noise can be exactly replicated through the seeding of a random number generator and introduced to a signal provides control and reproducibility when applied to testing datasets. Monte Carlo simulations allow for an increased number of controlled iterations enabling the evaluation of the aleatory uncertainty from a set of selected parameters (e.g SNR, type of entropy measurement) to form distri-

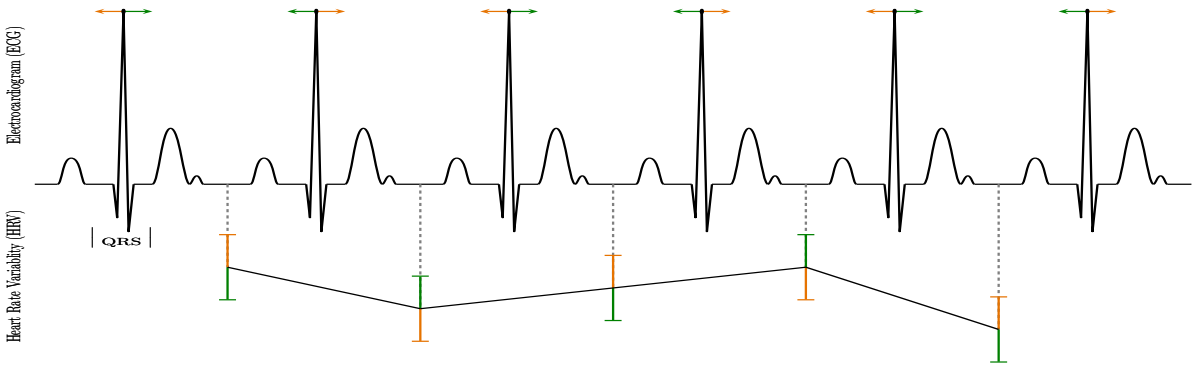


Figure 5.1: This illustration demonstrates the variation in the identified location of R-wave peaks result in uncertainty in the heart rate variability (HRV). Variation in the identified location of the R-peak from the QRS complex results in uncertainty in the heart rate variability (HRV) signal.

butions that characterize errors caused by the fiducial R-peak. This controlled simulation enables rigorous analysis on how the fiducial points impose errors in the system. Here we strictly analyze shifts in the mean and variance associated with the fiducial point at which noise enters the system. Using these approaches, we can overcome scaling differences from different entropy measures by evaluating the percent error (PE) in the various reported entropy distributions with respect to that found without the introduced noise.

Contributions. We designed and implemented a Monte Carlo simulation to evaluate the effects of time window size and SNR level of ECG signals on the percent error of three entropy calculation methods applied to those ECG signals. This method provided a controlled framework for evaluating the effects of ECG corruption on the reliability of three entropy measures.

The contributions of this work are as follows:

1. We characterize how SNR alters the locations of the fiducial point of the R-peak via a probability mass function (PMF), and demonstrate that the fiducial point for R-peak is most affected by white noise at varying SNR levels, as compared to pink noise.

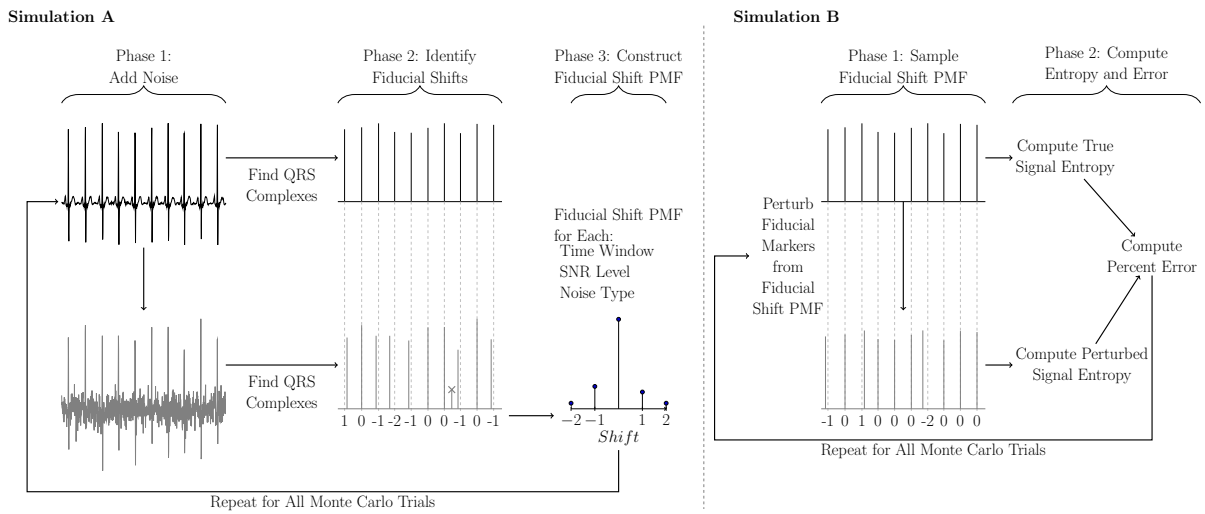


Figure 5.2: Flow Chart Process of Simulation

2. We demonstrate that permutation entropy is more precise than both approximate and sample entropy (i.e., it has the lowest standard deviation).

3. We demonstrate that, regardless of the entropy measurement or type of noise, as the time window segment increases, the precision of the measured entropy improves.

4. We demonstrate that increased precision does not imply a lack of sensitivity for demonstrating significant differences in HRC.

5. We demonstrate that as the SNR increases (altering the fiducial marker), there is a directional change in the measured entropy.

6. We show the level of SNR that affects the likelihood of making a distinguishable distribution statistically indistinguishable.

All of these findings ultimately aid us in future design methodologies to reduce predictive modeling error downstream from upstream feature extraction methodologies.

5.2 Methods

The methods section is split into two sections. The first method section addresses research questions RQ1 to RQ3 and the second section addresses RQ4 to RQ6.

5.2.1 Monte Carlo Simulation: Addressing RQ1 to RQ3

This section outlines the Monte Carlo simulation implementation method shown in Figure 5.2. These simulations were designed to study the effects of time window segmentation length and signal corruption resulting from multiple "colors" of noise on the prevalence of fiducial marker shifts on approximate, permutation, and sample entropy calculations.

1) Normal Sinus Rhythm Data: Single lead ECG data from 10 subjects (130 min, 128 Hz) selected from the MIT-BIH Normal Sinus Rhythm Database on PhysioNet [58] were utilized. The Normal Sinus Rhythm database was selected to reduce measurement variability that could be attributed to physiological dysfunction such as atrial fibrillation, ectopy, and other disorders [39]. Each subject's ECG signal was divided into 65, 26 and 13 intervals for 2, 5, and 10 minute time windows, respectively. Following signal preprocessing, noise was added, and the prevalence of shifts in the fiducial marker (i.e., the location of the R-peak) was collected to form a PMF.

2) Signal Preprocessing: After segmenting the ECG signals into intervals, a high pass and low pass finite impulse response Butterworth zero phase filter was applied to each time window. This produced our cleaned ECG signal, X_S. The zero phase filter avoids distortion in the phase of signal [62]. A key component for also avoiding fiducial marker shifting.

3) Simulation A: Determine Effect of Noise on Prevalence of Fiducial Shifts: The following section describes the procedure for creating a PMF of the prevalence of fiducial marker shifts caused by the addition of either white or pink noise to ECG signals.

a) Step 1: Add Noise: In Step 1, a white or pink noise signal, X_N, was added to the cleaned signal X_S comprising signal-to-noise ratios (SNRs) from 2 to 20 in increments of 2. The SNR level X_N added to X_S is defined as

$$SNR = 10\log_{10}\sqrt{\frac{\sum\left(X_S \cdot \bar{X}_S\right)^2}{\sum\left(X_N \cdot \bar{X}_N\right)^2}}, \tag{5.1}$$

where $\bar{X}_N$ and $\bar{X}_S$ are the complex conjugates of X_N and X_S. A randomly seeded X_N of each SNR was added to each subject's cleaned ECG signal interval 100, 250, or 500 times for each 2, 5, and 10 minute interval, respectively.

White noise is designed into the implementation of the Monte Carlo Simulation to evaluate the effects of electromagnetic interference, problematic sensors, or issues with wireless devices [33,97]. A uniform distribution was used to model white noise in the frequency domain by sampling from a random Gaussian distribution in the time domain sequence. The SNR is altered by adjusting the variance of the Gaussian distribution.

Pink noise was selected to evaluate the effect of correlated noise typically associated with observation noise on the ECG [38]. Pink noise was modeled as a decreasing function $(1/f)$ in the frequency domain and has close similarity to brownian motion-like noise which is modeled as a decreasing function $(1/f^2)$ and is related to electrode movement noise [38]. Pink noise was implemented using a noise generator package on MATLAB's file exchange service based on the theory for discrete simulations of colored noise by Kasdin [74].

b) Steps 2 and 3: Identify Fiducial Shifts and Construct Fiducial Shift PMFs: In Step 2, a QRS complex detection algorithm was applied to both the cleaned and noisy data to identify individual heart beats, R-R intervals, and heart rate variability (HRV) [117]. Using the Pan-Thompkins algorithm, the fiduical mark's location for the R-peak is determined from the rising edge of the waveform [117]. The locations of each R-peak extracted from the noisy signal were compared to the locations of the R-peaks extracted from the filtered signal. In

Step 3, the prevalence of noisy R-peaks within ±39.1 ms of their filtered counterparts were collected to form a PMF. This particular time increment was due to our sampling frequency of 128 Hz, thus $F_s^{-1} = 128^{-1} = 39.1\ ms$. False positives (i.e., noisy R-peaks detected outside of this window) and false negatives were not included in the construction of these PMFs. This process of creating PMFs was repeated for all subjects' time intervals (2, 5, 10 mins), resulting in PMFs for each noise type (white or pink), and for each SNR level (2,4,. . . ,20) equaling a total of 60 PMFs (i.e. 3 Time Intervals x 10 SNR Levels x 2 Types of Noise).

4) RQ1: Effects of Noise Level and Type on Fiducial Shift PMFs: We used visual inspection of PMFs and the corresponding tabulated shift frequency values to evaluate the effects of different SNRs and noise types on the prevalence of fiducial marker shifts.

5) Simulation B: Determine Effect of Fiducial Shift Prevalence on Heart Rate Complexity: The following section discusses the process of determining the effects on entropy measures resulting from perturbing the location of R-peaks extracted from filtered (cleaned) signals by an amount sampled from the PMFs constructed during Simulation A. *a) Step 1: Sample Fiducial Shift PMF:* In Step 1, the location of R-peaks of subjects' filtered ECG signal intervals were perturbed with a probability determined from the fiducial shift PMFs constructed in Simulation A. A HRV signal was then computed for both the filtered and perturbed signals.

b) Step 2: Compute Entropy and Percent Error: In Step 2, three entropy measures were applied to each of the HRV signals–approximate entropy [119], permutation entropy [13,126], and sample entropy [125].

Approximate entropy is defined as

$$ApEn(m, r, N) = \Phi^m(r) - \Phi^{m+1}(r), \tag{5.2}$$

where $\Phi^m(r) = (N - m + 1)^{-1} \sum_{i=1}^{N-m+1} \log C_i^m(r)$. $C_i^m(r)$ is the number of matches to

template i of length m within a tolerance of r (*including* self matches) [119].

Sample Entropy is defined as

$$SampEn(m, r, N) = -\ln\left[\frac{A^m(r)}{B^m(r)}\right] \tag{5.3}$$

where $A^m(r)$ and $B^m(r)$ are the probabilities of two sequences matching for $m+1$ and m data points respectively within a tolerance of r (*excluding* self matches) [125].

Permutation Entropy employs rank order (i.e., not exact distance) to quantify time series similarity. This is defined as

$$H_n = -\sum_{j=1}^{n!} p'_j \log_2 p'_j \tag{5.4}$$

where p'_j is the proportion of the occurrence of th j^{th} template (of length m) in the signal [13, 126]. Approximate [119] and sample entropy [125] both use distance as the measurement to examine similarity in time series to quantify entropy. The main difference between the two is that unlike approximate entropy, sample entropy excludes self matches. On the other hand, permutation entropy employs rank order (i.e., not exact distance) to quantify time series similarity [13, 126]. In Step 2, approximate, permutation, and sample entropy values were computed for both the filtered and perturbed HRV signals and compared to calculate percent error 100, 250, or 500 times for each 2, 5, or 10 minute interval, respectively; each noise type (white or pink); and each SNR level (2,4,. . . ,20).

6) RQ2: Comparison of Entropy Method Reliability: We computed the standard deviation of each time interval and then applied the natural logarithm to the data to obtain relatively equal variances and reasonably symmetric distributions. We then tested for the differences in the mean log standard deviation of the percent error using a one-way ANOVA.

7) RQ3: Effects of Window Size on Reliability: Using the data from Simulation B, conditional distributions were constructed of the percent error in entropy for time window

sizes of 2, 5, and 10 minutes and for both pink and white noise by collapsing over SNR, subject, and time interval number. Using visual comparison and descriptive statistics, the effects of window size on entropy reliability were addressed for both white and pink noise.

5.2.2 Sympathetic Response Dataset: Addressing RQ4 to RQ6

The experience of hypoxia is known to result in autonomic-nervous-system-driven changes (sympathetic arousal) in the cardiac and respiratory systems [61]. We utilize a dataset which characterizes hypoxic responses because of these known physiological changes in sympathetic arousal to address research questions RQ4 to RQ6.

1) NASA Hypoxia Data: The dataset was collected by a research team at NASA Langley Research Center (LaRC) who subjected 49 volunteers (all with current hypoxia training certificates) to normobaric hypoxia to study the impact on aircraft pilot performance [141, 142].

The goal of the study was to understand cognitive impairment due to exposure to mild hypoxia in order to improve the safety of psychophysiologically-based automation interfaces. Subjects in the study experienced simulated altitudes of Sea Level (21% O2) and 15,000 feet (11.2% O2) induced by an Environics, Inc. Reduced Oxygen Breathing Device (ROBD-2). During non-hypoxic and hypoxic exposures each subject performed a battery of written, computer-based, and flight simulation tasks each lasting 10 minute. In each exposure the research team collected task performance measures, a subjective self-report of workload (NASA Task Load Index), and multiple physiological responses (including ECG). This article discusses only the ECG data and task performance/conditions collected during hypoxic and non-hypoxic exposures.

2) RQ4: Evaluating Entropy Method Sensitivity: To evaluate the sensitivity of entropy measures in detecting mild hypoxia, approximate, permutation, and sample entropy

were calculated for ECG data collected during the final 2 min and final 5 min of non-hypoxic and hypoxic exposures. Ten minute time segments were excluded from this analysis because hypoxic exposures were only ten minutes long and thus not all subjects demonstrated indicators of mild hypoxia (e.g., $Sp0^2 < 80\%$) until serveral minutes into the $15,000$ ft exposures. Due to small sample size and non-normality, the non-parametric Wilcoxon rank-sum test was employed to test for differences between the hypoxic and non-hypoxic cohorts with smaller p-values indicating a greater ability for an entropy calculation method to discriminate between hypoxic and non-hypoxic states. The initialized parameters used in calculating the entropy measurements were the same as in Simulation B ($m = 2$ & $r = 0.15\sigma$ for sample and approximate entropy and $m = 3$ for permutation entropy order).

3) Simulation C: Effect of Fiducial Shifts on Entropy Data from Hypoxic Subjects: Using the NASA hypoxia ECG dataset and the PMFs of the fiducial marker shifting found in Simulation A, we corrupt the HRV signals by altering the locations of the R-peaks for 2 and 5 minute time segments at various SNR levels. Each subject and their respective cohort's (hypoxia and non-hypoxia) fiducial markers are corrupted 250 times for each subject to examine the raw change in the entropy distribution.

4) RQ5: Effects of Fiducial Shifts on Entropy Skew and Bias in Hypoxic Subjects: We pooled the approximate, permutation, and sample entropy values for all subjects' Monte Carlo trials for both hypoxic and non-hypoxic exposures at each SNR level and time window size (2 and 5 min). Kernel density plots of each pooled distribution were obtained and visually inspected for skewness. For each conditional pooled distribution, the mean entropy value was computed and qualitatively assessed for general trends in the data.

5) RQ6: Effects of Fiducial Shifting on Cohort Discriminability: We demonstrate this through an approach similar to that of RQ3. Using the NASA hypoxia dataset and the probability distributions associated with the fiducial shifts, we alter the fiducial markers of the R-peaks in the ECG signals to simulate corruption for each subject. This process is

simulated 250 times for each time window segment length, cohort, and entropy measurement at each SNR level.

During each simulation trial the non-parametric Wilcoxon rank-sum test evaluated differences between the hypoxic and non-hypoxic distributions with the null-hypothesis significance level set at 0.05. After all simulation trials were complete, the percentage of tests rejecting the null-hypothesis (i.e., having $p < 0.05$) were collected. A greater percentage of null-hypothesis rejections indicates a greater ability to resolve hypoxic from non-hypoxic states.

5.3 Evaluation and Discussion

In this section, we address the following research questions regarding the effects of ECG corruption on entropy dynamics which were introduced in Section I:

RQ1 How does SNR and various colored noise effect the fiducial point of the R-peak?

RQ2 If the fiducial point of the R-peak shifts, does a single entropy method demonstrate superior precision from the proposed three types of entropy calculations that measure HRC?

RQ3 Does increasing the HRV signal time segment length enhance HRC variance (entropy dynamics)?

RQ4 Is increased precision associated with lack of sensitivity for demonstrating significant differences in HRC?

RQ5 Is there a directional change in the measured entropy upon corruption in which the entropy increases, decreases, or does only the variance symmetrically increase providing no actual change in complexity?

RQ6 At what simulated SNR level do fiducial shifts render two previously distinguishable

distributions indistinguishable?

Table 5.1: Pink Noise Monte Carlo Simulation Summary of Probability Distribution of the Fiducial Shift (Trials=$\sim 65,000$).

Target	**Distribution of Fiducial Shift by Milliseconds**				
SNR	**-15.62ms**	**-7.81ms**	**0ms**	**7.81ms**	**15.62ms**
2	0.0	4.63	90.39	4.95	0.0
4	0.0	3.68	92.41	3.89	0.0
6	0.0	2.94	93.97	3.07	0.0
8	0.0	2.33	95.22	2.42	0.0
10	0.0	1.85	96.21	1.92	0.0
12	0.0	1.47	96.99	1.53	0.0
14	0.0	1.18	97.60	1.21	0.0
16	0.0	0.94	98.10	0.96	0.0
18	0.0	0.75	98.48	0.76	0.0
20	0.0	0.60	98.79	0.60	0.0

5.3.1 RQ1: Fiducial R-peak Marker, Type of Noise, and SNR

We hypothesized that the fiducial marker will change as a function of the SNR and type of noise. Each row in Tables 5.1 and 5.2 present a PMF of the fiducial marker shifting for white and pink noise, respectively, as a function of the targeted SNR. The PMF was generated by part A of the simulation discussed in Figure 5.2 and describes the probability of the fiducial marker shifting. Each shifting in the fiducial marker is a single discretized point or index in the time series equivalent to F_s^{-1}. Since the sampling rate of the discretized signal is 128Hz, the index shifts in increments of 7.81ms. Both of these tables demonstrate that as the SNR increases, there is a greater probability of altering the fiducial marker. As a function of the type of noise applied to the signal, white noise is the most corruptive in altering the fiducial marker.

5.3.2 RQ2: Type of Entropy and Precision

Due to the fiducial point of the R-peak shifting, we address how precision (i.e., standard deviation, σ) can alter the proposed three types of entropy calculations that measure HRC.

Table 5.2: White Noise Monte Carlo Simulation Summary of Probability Distribution of the Fiducial Shift (Trials=$\sim 65,000$).

Target	**Distribution of Fiducial Shift by Milliseconds**				
SNR	**-15.62ms**	**-7.81ms**	**0ms**	**7.81ms**	**15.62ms**
2	0.01	9.74	78.9	11.21	0.10
4	0.0	8.02	82.90	9.02	0.04
6	0.0	6.52	86.29	7.16	0.01
8	0.0	5.23	89.13	5.62	0.0
10	0.0	4.20	91.35	4.44	0.0
12	0.0	3.33	93.15	3.49	0.0
14	0.0	2.65	94.57	2.79	0.0
16	0.0	2.10	95.71	2.17	0.0
18	0.0	1.66	96.60	1.73	0.0
20	0.0	1.37	97.21	1.41	0.0

Table 5.3: Pink Noise Entropy Percent Change Distribution: (Trials=$\sim 65,000$).

Time	**Target**	**Approximate Entropy**		**Permutation Entropy**		**Sample Entropy**	
Window	**SNR**	μ	σ	μ	σ	μ	σ
2	2	0.34	5.90	1.83	1.38	7.68	9.23
2	6	0.27	5.14	1.50	1.23	5.89	9.10
2	10	0.34	3.95	1.05	1.01	4.02	6.25
2	14	0.20	3.12	0.68	0.84	2.48	4.96
2	18	0.04	2.45	0.47	0.68	1.65	4.31
5	2	1.52	4.39	1.65	0.93	5.50	7.85
5	6	1.18	3.89	1.35	0.81	4.25	7.16
5	10	0.82	2.82	0.93	0.64	2.97	4.98
5	14	0.53	1.98	0.62	0.50	1.95	3.32
5	18	0.36	1.49	0.42	0.39	1.28	2.50
10	2	4.78	3.22	1.61	0.59	6.99	2.50
10	6	3.77	2.63	1.32	0.52	5.45	2.09
10	10	2.50	1.88	0.90	0.40	3.55	1.66
10	14	1.64	1.33	0.60	0.32	2.31	1.29
10	18	1.09	1.04	0.40	0.26	1.50	1.05

Tables 5.3 and 5.4 are generated from Part B of the Simulation discussed in Figure 5.2 that develops distributions based on the percent change for their respective entropy measurement, SNR, Time Segment, and type of noise. To address RQ2 we can first visually compare the standard deviations of Tables 5.3 and 5.4 where for each row in the table, all the variables (e.g., SNR, Time Segments) are held constant except for the type of entropy measurement (e.g., sample entropy). We demonstrate that the standard deviation of the permutation en-

tropy is respectively lower than either approximate or sample entropy across all SNR levels, time segments, and type of noise. Thus from the Monte Carlo simulation, we demonstrated that permutation entropy provides a more precise HRC measurement than either approximate or sample entropy. This finding is likely due to the manner in which permutation entropy is calculated versus sample and approximate entropy. Permutation entropy analyzes HRV by rank order of the R-R interval timing rather than their distance criteria like sample and approximate entropy. Outside of false positives and missed detections of the QRS wave, corruption of the fiducial shift and the timing of QRS wave indices has a stronger effect on distance than on the rank order.

For both white and pink noise, the one-way ANOVA indicated that there were significant differences in the mean log standard deviation of the percent error of the three entropy measures ($F(2, 31575) = 12440$, $p < 0.05$ and $F(2, 30957) = 11156$, $p < 0.05$ for white and pink noise respectively). Post-hoc comparisons using the Tukey-Kramer procedure revealed that all entropy methods produced significantly different mean log standard deviations of the percent error ($p < 0.05$) for both white and pink noise with permutation entropy having the lowest mean log standard deviation of the percent error followed by approximate entropy and then sample entropy.

5.3.3 RQ3: Noisy Time Segments and HRC Variance

Increased time segments provide more data which should aid in reducing the aleatory uncertainty for entropy measurements. Therefore we hypothesize that increased time segment lengths at a fixed SNR value would result in more stable and precise HRC measurements. Through the proposed Monte Carlo Simulation (Part B) in Figure 5.2, the SNR values were kept fixed but the time windows varied by 2, 5 and 10 minutes for pink and white noise. By examining Tables 5.3 and 5.4 at their respective SNR levels, we note that the standard deviation and expectation decreases as the window size increases for both white and pink

Table 5.4: White Noise Entropy Percent Change Distribution: (Trials=$\sim 65,000$).

Time Window	Target SNR	Approximate Entropy μ	Approximate Entropy σ	Permutation Entropy μ	Permutation Entropy σ	Sample Entropy μ	Sample Entropy σ
2	2	-0.34	11.21	3.45	2.07	18.84	14.63
2	6	0.21	8.35	2.75	1.74	12.92	11.75
2	10	0.40	6.29	1.98	1.46	8.65	9.84
2	14	0.32	4.78	1.39	1.19	5.71	7.84
2	18	0.26	3.77	0.94	0.98	3.63	6.64
5	2	3.31	8.26	3.21	1.65	14.24	13.38
5	6	2.46	6.41	2.57	1.32	9.50	11.54
5	10	1.72	4.96	1.88	1.04	6.42	9.03
5	14	1.15	3.73	1.33	0.80	4.24	6.78
5	18	0.85	2.52	0.89	0.62	2.97	4.20
10	2	12.7	8.07	3.23	1.09	19.20	5.24
10	6	8.54	5.58	2.59	0.86	13.10	3.87
10	10	5.75	3.80	1.89	0.66	8.50	2.82
10	14	3.80	2.69	1.32	0.51	5.49	2.18
10	18	2.48	1.90	0.90	0.40	3.54	1.66

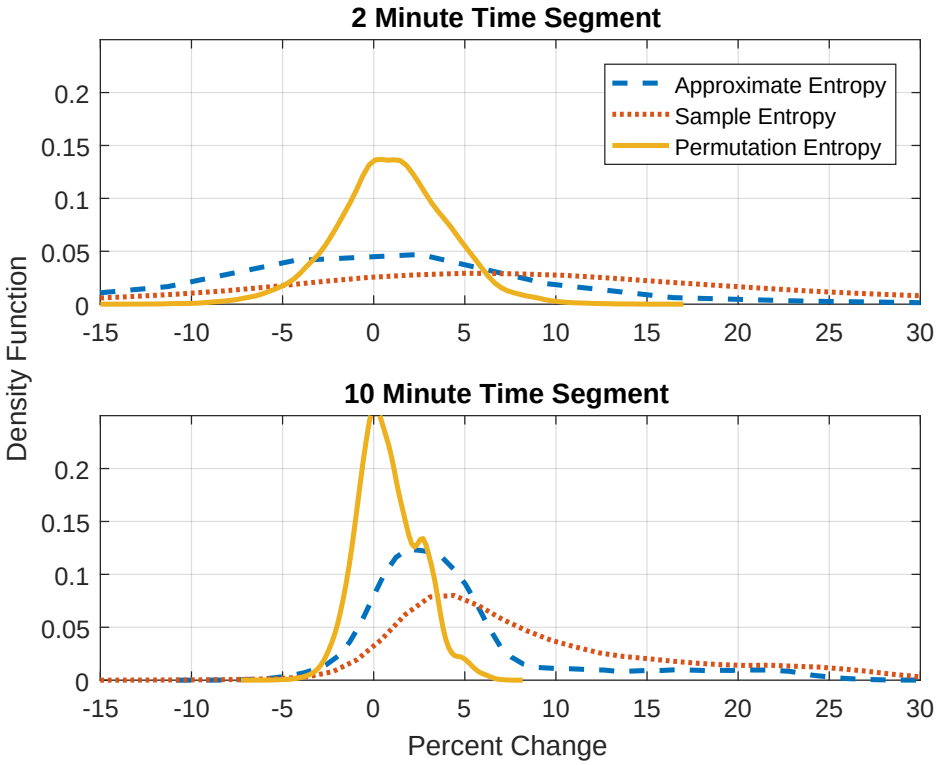


Figure 5.3: Percent Change Entropy Distribution (SNR=2db, White Noise)

noise.

Based on the Monte Carlo simulated samples, a pictorial representation of the distribution was created using kernel density estimation for white noise SNR levels of 2 and 20 at 2 and 10 minute segments shown in Figures 5.3 and 5.4. Figure 5.3 visually demonstrated Sample Entropy extending beyond a 30% change from its original entropy calculation. However, as the time segment increases we can note the mass of the distribution becomes more centralized, thus increasing the precision of the entropy measurement. This effect of increasing the windowing segment length is consistent for all three entropy calculation methods. In regards to real-time health monitoring, we are met with a trade-off between time resolution and the precision of the entropy measurement. For example, as we aim to characterize a subject's physiology within a smaller time frame, the precision of the measurement worsens. Thus, when designing these systems, one should consider the level of imprecision which can exist before unacceptable Type 1 or Type 2 errors are encountered in the system?

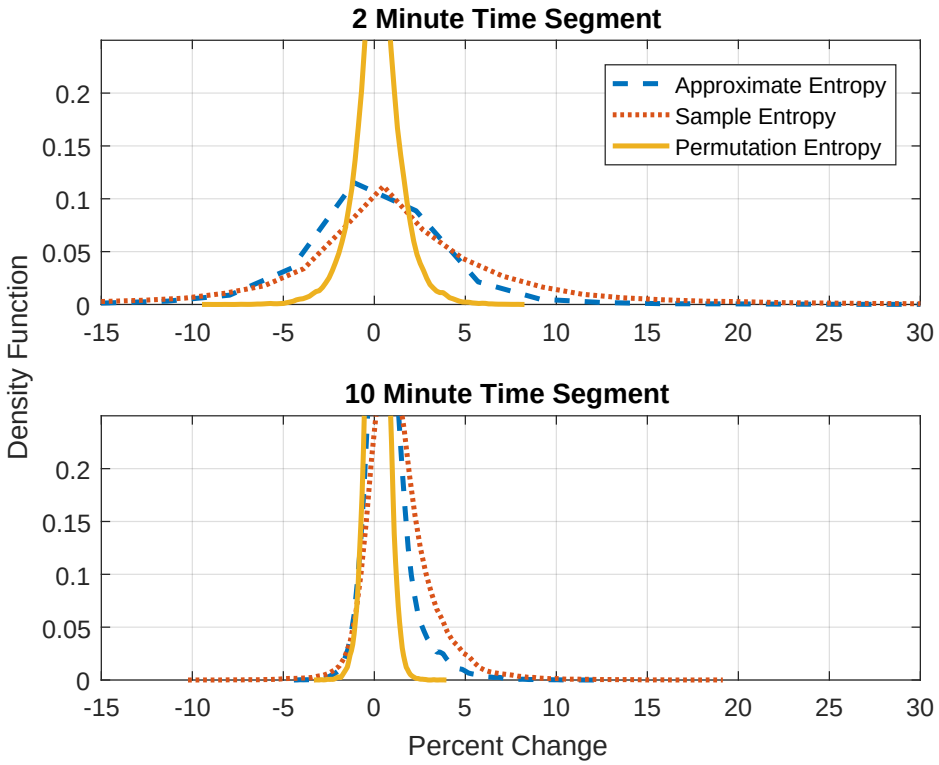


Figure 5.4: Percent Change Entropy Distribution (SNR=20db, White Noise)

5.3.4 RQ4: Insight HRC Sensitivity

Although some HRC measurements may not offer ideal precision or stability (e.g., Sample and Approximate Entropy), we hypothesize that this lack of precision allows for higher sensitivity in distinguishing features between two cohorts. That is, the increased variation in the measurement that results in a decrease in precision provides information that can aid in demonstrating independence between two classes or distributions.

Table 5.5, provides the p-value for the Wilcoxon Rank Sum Test for the three proposed entropy measurements at 2 and 5 minutes, which evaluates the hypoxic and non-hypoxic cohorts of the presented NASA experiment. In Table 5.5, we note that permutation entropy had a lower P-Valve for both sets and for all time segments than approximate and sample entropy, demonstrating independence between two classes or distributions. Based on the results and discussion of RQ3, permutation entropy was also shown to be the most precise measurement from the other proposed entropy measurements. Thus, the results of this data do not fully support our hypothesis that these imprecise measurements are more sensitive to noise but are better at distinguishing independence. Hence, there is no trade-off between increased precision and the ability to distinguish the two distributions of hypoxia and non-hypoxia, a known cause of autonomic nerve system activation which alters HRC dynamics.

Table 5.5: Significance level (p-value) of Wilcoxon Rank Sum test between hypoxic and non-hypoxic cohorts

Time	$PermEnt_O$	$ApproxEnt$	$SampEnt$
2 mins	$\mathbf{p = 0.022}$	$p = 0.55$	$p = 0.099$
5 mins	$\mathbf{p = 0.032}$	$p = 0.067$	$\mathbf{p = 0.034}$

5.3.5 RQ5: What is the Directional Change in Entropy

As noise alters the fiducial point, we aim to understand how this affects the direction of the entropy measurements. For example as noise increases and alters the fiducial point, does the measured entropy increase, decrease, or is there no clear trend apparent (symmetric)? We hypothesized that increased perturbations of the fiducial point from increased noise would make the HRV signal less predictive, thus increasing the entropy measured in the signal. There is some support of this hypothesis from the results provided in RQ2 and RQ3, where you can note a general positive shift in the reported expectation (μ) of the calculated percent change. This is also demonstrated by Figures 5.3 and 5.4, where the percent change of the distributions are skewed further to the right as SNR increases.

However, in order to properly support our hypothesis RQ5 provides the raw calculated values which were averaged over subjects and trials. More specifically Table 5.6, provides the mean value of these distributions specifically associated to either the Hypoxic (H_μ) or Non-Hypoxic (NH_μ) case across SNR levels and time windows. We can note a consistent linearly increasing mean entropy as noise increases from the initial uncorrupted dataset for all time segments and entropy measurements. This finding is critical since healthier individuals (an individual with a lower sympathetic response) have higher entropy measurement producing more complex heart patterns and higher variability. Hence this informs us that if corruption occurs and the fiducial point is altered, subjects will appear healthier than they actually are.

5.3.6 RQ6: Indistinguishable Independent Distributions

We demonstrated in RQ5 that there is a clear directional change in the measured entropy when noise is introduced into the signal. Because the noise does not alter the measured entropy distribution symmetrically (e.g., Figure 5.3) and thus shifts the entire distribution in one direction, we must then inquire about the implications of how this statistically affects

Table 5.6: White Noise Entropy Movement (Trials=$\sim$ 250).

Time	**Target**	**Approximate Entropy**		**Permutation Entropy**		**Sample Entropy**	
Window	**SNR**	H_μ	NH_μ	H_μ	NH_μ	H_μ	NH_μ
2	2	0.679	0.691	0.888	0.909	1.345	1.483
2	6	0.672	0.688	0.883	0.906	1.332	1.470
2	10	0.669	0.687	0.880	0.903	1.319	1.465
2	14	0.667	0.686	0.878	0.902	1.310	1.462
2	18	0.665	0.685	0.876	0.901	1.307	1.459
5	2	0.911	0.975	0.897	0.914	1.227	1.407
5	6	0.903	0.970	0.892	0.911	1.206	1.393
5	10	0.895	0.967	0.888	0.909	1.191	1.384
5	14	0.892	0.964	0.886	0.908	1.182	1.384
5	18	0.888	0.962	0.884	0.907	1.177	1.377

our readings and to what degree it affects statistical independence. For example if a subject has a lower entropy measurement, we perceive them to have a higher amount of physiological stress (i.e., increase sympathetic response). However if noise is introduced into the signal, causing an increase in entropy, what is the likelihood that the subject would statistically be viewed as someone not under stress?

Since hypoxia is known to cause an increase in sympathetic response, we address this research question through an approach similar to that of RQ4 using the NASA hypoxia dataset. We hypothesized that as SNR increases, it becomes increasingly difficult to discriminate ($\alpha = 0.05$) between the two cohorts of hypoxia and non-hypoxia HRC measurements. The simulation is iterated 250 times for each window time segment, cohort, type of entropy measurement, and SNR for their respective cohorts, hypoxia and non-hypoxia. In Table 5.7, (Sig_μ) denotes the simulations where noise was only added to the hypoxia cohort data. However, ($dSig_\mu$) denotes the simulations where noise was added to both the hypoxia and non-hypoxia data. The percentage of statistical tests that were significant from the 250 simulations are shown in each column of Table 5.7. This percentage is intended to provide contextual meaning of the likelihood that the two cohorts are statistically independent as a function of SNR, windowed time segment and type of entropy measurement. From Table 5.7

can observe that there is much smaller percentage of statistical independence when noise is only introduced to the hypoxia cohort. We already know from Table 5.5, when no noise is introduced, that Permutation Entropy for a 2 minute time segment is significant, ($p = 0.022$). However for the Sig_μ case of corrupting only the hypoxic cohort with an SNR of 10, we expect the two classes to be independent only 5.2%. This means that based on the 250 simulations that we ran, only 13 of the 250 generated a p-value less than 0.05. Similarly for the $dSig_\mu$ case when both the hypoxic and non-hypoxic cohorts are corrupted, 222 out of the 250 simulations generated a p-value less than 0.05. Thus, we expect to achieve significance 88.8% of the time when both signals are corrupted. This discrepancy between the Sig_μ and $dSig_\mu$ cases is due to the impact on how the noise shifts the means the distributions. Hypoxia readings have a lower entropy measurement since the subject is under stress. However if noise is present strictly for the hypoxia data it biases the distribution into making the collected data look as if the subject is not under stress and thus non-hypoxic. Therefore, it becomes more difficult to statistically distinguish between to the two cohorts. Machine learning approaches specifically designed to detect anomalies in the data, such as single class Support Vector Machines (SVMs) [32, 144], are designed to learn the decision boundaries strictly on a single class. Thus as incoming data is altered by noise and the distribution is being shifted outside classifier boundary, the algorithm will classify it as an anomaly causing a false positive. This finding is paramount if we know that hypoxia HRV signals are being corrupted, and can be combated by introducing noise into the control cohort (i.e., non-hypoxia) to appropriately adjust the results of our statistical inference (re-train our ML algorithm for more generalized classifier boundaries to incorporate various noise levels) or simply bias the distribution based on your incoming SNR.

Table 5.7: Evaluating the Uncertainty in Indistinguishable Independent: (Trials=$\sim$ 250).

Time	**Target**	**Approximate Entropy**		**Permutation Entropy**		**Sample Entropy**	
Window	**SNR**	Sig_μ	$dSig_\mu$	Sig_μ	$dSig_\mu$	Sig_μ	$dSig_\mu$
2	2	0%	0%	0%	70.0%	0%	0.4%
2	6	0%	0%	0.4%	83.2%	0%	0.8%
2	10	0%	0%	5.2%	88.8%	0%	0.2%
2	14	0%	0%	33.6%	92.4%	0%	1.2%
2	18	0%	0%	72.0%	96.4%	0%	0.4%
5	2	0%	0%	0%	24.4%	0%	18.0%
5	6	0%	0%	0%	47.2%	0%	51.6%
5	10	0%	0%	0%	69.6%	1.2%	88.4%
5	14	0%	0.8%	1.2%	78.4%	26.8%	96.4%
5	18	0%	0%	20%	85.2%	79.2%	99.6%

5.4 Conclusion

Current literature about handling HRC calculations focuses on reducing the false positive rate and false negative rates of detecting the QRS waveform in the ECG signal. The location of the R-peak is typically extrapolated from the detected QRS waveform and little attention is given to how errors regarding the location of the R-peak can alter HRC readings. In this paper we present a Monte Carlo simulation framework that evaluates the effects of ECG signal corruption on the fiducial point of the R-peak and how it effects HRC measurements when using sample, approximate, and permutation entropy.

Through the use of Monte Carlo simulations, we able to characterize PMF distributions and how the fiducial point shifts based on signal quality of the ECG. White noise was shown to cause higher perturbations in the fiducial point of the R-peak when compared to pink noise. This characterization allowed us to run additional Monte Carlo trials in order to evaluate changes in the precision of the proposed entropy measurement, in which permutation entropy demonstrated to be most precise during corruption. From these findings, we utilized a secondary data set that addresses the sensitivity of demonstrating a statistical difference between hypoxia vs non-hypoxia caused by altered heart rate dynamics from the autonomic nervous system. This analysis showed that permutation entropy not only had better preci-

sion under noisy environments but was also sensitive statistically for 2 and 5 minute time segments between the two cohorts. Approximate entropy was not significant for either 5 or 10 minutes and sample entropy was only significant for 5 minute time segments. This work then demonstrated that as perturbations of the R-peak increased as a function decreasing SNR, the entropy of the signal increased. Because of biasing of the distribution, we showed that because of its stronger precision and sensitivity, permutation entropy was still able to out perform sample and approximate entropy the likelihood of showing statistical change in heart dynamics during hypoxia for 2 minute segments.

HRC calculations are critical to the implementation of statistical modeling techniques in numerous bio-informatics domains for physiological insight. Thus we were able to address critical fundamental design questions, allowing researchers to evaluate what type of entropy measurements are best, suggestions on how to handle ECG signal corruption of the fiducial point, and ideal time window segments based on the type of environment imposed on their telemetry system.

Discovering the unexpected is more important than confirming the known.

George E. P. Box

6

Featuring Engineering to Provide Additional Model Dimensionality

Electroencephalograms are utilized for the detection of variations of brain wave patterns in order to provide information about a person's cognitive states and to detect neurological diseases. Typical methods for EEG analysis are known to examine spectral content, for which vast amounts of literature have developed spectral interpretations of the signal over the decades. More recently, non-linear analysis (e.g., sample entropy) has provided further valuable information, although, there is no non-linear method that examines isolated

spectral frequency bands tailored to EEG signals. Therefore, we develop a time-frequency intensity analysis method to extract continuous measures of intensity for key frequency bands associated with EEG. Utilizing this time-frequency intensity method, we design two novel non-linear complexity measurements referred to as Activation Complexity (AC) and Rank Order Complexity (ROC). This paradigm was then applied to a data set from a study involving hypoxia induction in 49 human test subjects, demonstrating that significant changes in EEG firing patterns can be captured. Utilizing AC, we show that we can capture anomalous changes in firing for extended 10-minute segments by examining how the timing of the peak intensity is altered during hypoxia. ROC was designed for task-related changes, which demonstrated how the rank of specific frequencies fluctuating over time is significant during tasks.

6.1 Introduction

Electroencephalograms (EEGs) detect and track variations in brain wave patterns related to the electrical activity of the brain. EEG analysis provides information about a person's cognitive states such as response inhibition, level of concentration, arousal and even diagnostic information regarding diseases such as Alzheimer's, hypoxia encephalopathies, and epilepsy [2,6,14,41,73]. In order to acquire these insights and information from EEGs, there are many types of analyses designed to extract features from the signal that examine coherence, intensity of frequency bands, signal entropy, coupling, and source localization [3,6]. These extracted EEG features are then used as the foundation for explanatory and predictive modeling. Typically, two or more of these features are utilized to generate a feature space for predictive models that can predict epilepsy, hypoxia, etc. [6]. Thus, capturing these new EEG features is paramount to uncovering latent patterns that provide further insight into the complexities of the human brain and distinguishing impairments.

Literature has demonstrated that conditions like hypoxia, Alzheimer's, epilepsy and other neurological issues cause neuronal impairments that change firing patterns. Modification of firing can potentially occur at the intracellular level of an individual neuron or at the intercellular level in how neurons propagate information to each other (neuronal interactions). However, the non-linearity of the processes at both the intracellular and intercellular level are caused by dynamic behavior [3], making it difficult to capture explanatory responses. At the intracellular level, neuronal firing, or the generation of an action potential, demonstrates non-linearities in how thresholding and saturation phenomenon are governed [3]. At the intercellular level, neuronal interactions occur spatially giving a second dimension to these non-linearities [3]. These combined neuronal interactions are summed, potentially enabling subsequent neurons to meet their threshold criteria and thus firing, as well [64]. However, if a neuron is impaired, it has the potential to impede transmission and prevent subsequent neurons from firing, thus causing neurons to be functionally and electro-physiologically isolated [118].

Prior Work. Through these complex interactions of signal propagation and cellular processes, the EEG signal is characterized as a non-linear time series [3, 70, 118]. Though these complexities in neuronal firing occur, even in the case of simple cognitive changes (e.g., sleeping) signal propagation changes, which ultimately alters whether a neuron's threshold criteria is met or not [120]. Thus, this affects instances in the global measurement of electrical activity, changing the intensity/power of EEG band-limited waveforms (e.g. Alpha, Delta, Theta). Based on this rudimentary and fundamental point, it is not surprising that the first EEG patterns were noticed by Berger [16], where he observed that oscillatory patterns (alpha waves $8 - 12$ Hz) were attenuated during conscious walking when compared to rest/sleep [26].

These observations became visually apparent because the EEG alpha frequency is band-limited (8-12 Hz) and its intensity is more dominant during specific conditions (such as

sleep). Thus, spectral intensity analysis methods have been the hallmark approaches for EEG analysis [3, 70], and Fourier methods have been the typical method for analyzing the intensity of specific frequency bands (i.e. delta δ, theta θ, alpha α, etc.). However, EEG signals are non-stationary [70], making Fourier approaches problematic since they require the assumption that the signal is infinitely long and stationary [113]. We are required to make this assumption about the signal being analyzed because Fourier uses sine and cosine waves as its basis function in order to decompose the signal into its appropriate frequency components. To overcome the issue of EEG signals being non-stationary, Short-Time Fourier Analysis (STFT) is applied. STFT allows us to assume the signal is stationary by applying windowing paradigms that typically segments the time series into 5 to 30 second windows [5]. However, this type of analysis prevents us from examining how and when those frequencies and intensities change within that windowed time segment. This delimma has been previously identified as it relates to physiological signals [113] and is known as the Signal Procsesing Uncertainty Principle [52], which is related to Heisenberg's Uncertainty Principle in physics [68]. This leads to the the Heisenberg Uncertainty Principle's fundamental trade-offs related to signal processing [113]. In order to obtain an increased time resolution, one loses frequency resolution. Like-wise, in order to gain better frequency resolution one loses time resolution [113]. Furthermore, strictly examining the raw intensity leads to high variability between subjects due to the variability inevitably imposed by electrode conductance. Additional methodoloiges are required to combat this variability.

More current literature aims to determine more suitable non-linear analyses for examining EEG complexity using various types of entropy measurements to obtain informative features that can detect cognitive states and diseases [3, 6, 70, 151]. The general underlying concept of these entropy measurements is based on examining complex sequences for similarities in patterns to quantify the predictability of the sequence. If the entropy is low, there are many patterns that are similar and thus the sequence is highly predictable. On the other hand, if the entropy is high, the sequence has fewer similar patterns and is less predictable.

There are numerous ways one can quantify similarity, and hence, various methods for calculating entropy [85]. Sleigh and Abasolo both discuss EEG entropy signal analysis methods belonging to one of two possible families of entropy estimators [3, 139]. The first family consists of "phase-space embedding entropies," which are designed to estimate the signal in the time domain. Popular methods within this family consist of Approximate, Shannon, phase, sample, Kolmogorov, fuzzy, and permutation entropy [6]. However, these methods do not examine similarity with regard to the frequency content of the signal. As depicted in Table 6.1, these frequency-specific EEG waveforms (e.g. alpha, gamma) have been shown to indicate discriminable cognitive states and to have contextual meaning associated with brain damage, disease, and cognitive states based on decades of supported literature [27, 143, 156]. Thus, these temporal entropy approaches overlook a large part of the classical concepts of EEG analysis.

The second family of entropy estimators are referred to as "spectral entropy" methods, which consist of spectral, and Normalized Bispectrum Entropy methods. These methods aim to examine entropy from a frequency perspective, but at the cost of losing temporal information due to spectral windowing limitations (i.e., STFT). Furthermore, these methods typically utilize Fourier analysis methods which, as stated above, are inappropriate for EEG analysis due to the stationary assumptions Fourier analysis imposes and the lack of granularity in changes signal frequencies and intensities [113].

A unique method in the spectral entropy family is Wavelet Entropy, which uses a novel method of analyzing frequency. These entropy measurements use the wavelet transform, which imprecisely measures frequency through a scaled version of its basis function. The wavelet transform aims to overcome these traditional trade-offs imposed by Fourier analysis by using a multi-resolution analysis that provides time, scale, and magnitude. The continuous wavelet transform (CWT) uses convolution of the signal and its basis function,"the mother wavelet," to analyze the signal, similar to how one would analyze a Fast Fourier

Transform (FFT) Convolution. However, the CWT's basis function is considered to be a small, non-stationary wave composed of numerous frequencies (and therefore, imprecise), whereas the FFT approach uses stationary sine and cosine waveforms at a definitive, unique frequency. CWT's mother wavelet or basis function is scaled to where the basis function is either compressed (consisting of a range of high frequencies), or dilated (consisting of a range of low frequencies). The output of the CWT produces wavelet coefficients that provide an indicator of how well the signal correlates to the scaled mother wavelet, which can be used later to estimate local energy [127]. Thus, Wavelet analysis has gradually become an advantageous biological signal processing because of its ability to handle non-stationary processes and provide a multi-resolution analysis. The Wavelet Entropy methodology utilizes the CWT and picks only a sparse set of scaled mother wavelets. This approach is called a decimated discrete wavelet transform (DDWT) [127] and is used because the CWT scaling can be too granular, where the difference in the frequency content between adjacent mother wavelets is too similar, which causes redundant analysis. Wavelet Entropy uses these DDWT coefficients to estimate the energy of a windowed time segment, or signal intensity. However, this still causes a limitation in time resolution that prevents us from understanding how events evolve (e.g. onset, duration) over time; a consistent problem within the spectral entropy family. Furthermore, scaled mother wavelets consist of a range of frequencies and are traditionally designed in the time domain. Thus, the scaled mother wavelet that is being used to calculate energy is not designed specifically for the frequencies that incorporate multiple EEG waveform frequency bands (e.g., alpha and theta). This prevents the approach from targeting specific EEG waveforms directly, a similar problem within the family of "phase-space embedding entropies' analysis.

From an analytical standpoint, none of these Entropy methods used to characterize the nonlinearities of EEG signals capture the intensity of specific EEG waveform spectral properties continuously over time, nor do they attempt to calculate precise dynamic temporal changes. Thus, there is no complexity method that can explain both the temporal and spectral com-

plexity relationships. Analyses involving intensity changes can provide key features and a new understanding of EEG analysis by assessing the non-linear dynamics.

From a physiological standpoint, these complex neuronal dynamics revealed by a method to measure both the temporally and spectrally complex relationship of firing patterns can potentially provide new information. The notion of complex neuronal networks, which generate these fundamental neuronal oscillations, have been backed by a vast amount of literature [26, 64, 158]. In order to capture the complexity of these dynamics, we propose the use of a rudimentary example with a simulated EEG that examines a simplistic network to enable the development of these methods (see Figure 6.1).

This Figure presents two cases: column one presents a standard-band limited neuronal oscillation, and column two depicts a band-limited neuronal oscillation, where a subset of neurons are functionally isolated (i.e. a neuron is impaired and has the potential to impede transmission and prevent subsequent neurons from firing). As previously discussed, we can observe these EEG oscillatory patterns through global field potentials at localized regions on the scalp which are generated by the summation of large populations of neuronal action potentials. These populations of synchronized neuronal action potentials (in black) are are shown in Figure 6.1 in row two. In column two, the action potentials that did not fire are shown in red. These action potentials are summated at the macroscopic level and are viewed at the level of global field potentials shown in Figure 6.1, row 1 in black with the intensity in blue. It is worth noting that the intensity is attenuated via the analysis of the power spectrum (in row three) and the characterized intensity over time, both in blue.

Moreover, the chaotic nature of the intensity over time is increased. Currently, the non-linear dynamics of how and when these intensities alter over time are not captured with these methods (seen in blue in Figure 1 of the first row). Note how a band-limited frequency and it's intensity can change over time. Current entropy measurement windowing techniques do not pinpoint instantaneous changes with regards to intensity, frequency, and time. This

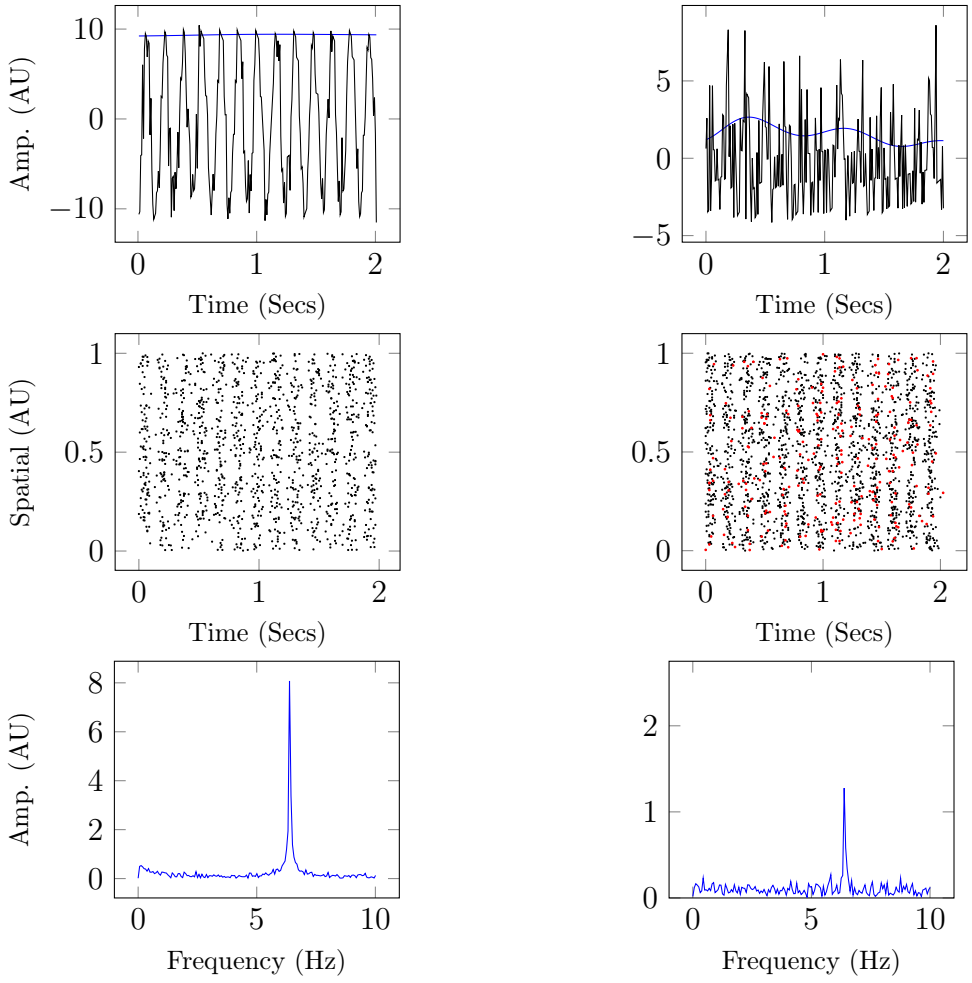
Table 6.1: EEG Waveform Interpretation

EEG Waveform	**Frequency**	**Interpretation**
Delta	0-3.5 Hz	Sleep
Theta	4-7 Hz	Idling, Inhibition
Alpha	8-12 Hz	Relaxed/Reflecting
Low Beta	13-15 Hz	Slight Concentration
Mid Beta	15-18 Hz	Active Thinking
High Beta	18-40 Hz	Over-Arousal
Gamma	32-100 Hz	Sensory Processing

limitation calls for development of techniques capable of assess whether there is additional information that could provide explanatory responses induced by neurological impairments (e.g., stroke, cancer, etc.).

Therefore, this work raises three relevant research questions: 1) Can we specifically measure EEG spectral waveforms (e.g. alpha) continuously over time to better capture changes in events? 2) How does the complexity of the intensity change for specific EEG Spectral waveforms over time? 3) How does rank of the spectral intensity change over time across competing band-limited spectra? 4) How do the proposed novel EEG entropy signal analysis methods compare to other standard measurements??

Challenges. In order to detect instantaneous changes in intensity and relate them to the specified band-limited EEG waveforms in Table 6.1, a unique signal processing method must be developed. This method would share similarities to the DDWT method in order to capture intensity continuously overtime. However, each filter would have to strictly capture the specified EEG waveforms and the adjacent filters would have to be considered in the method design to prevent redundant analysis of frequency content. Thus, if there is no redundancy, full reconstruction of the original signal can be achieved (plateau value of zero). This would mean that each filter is properly capturing only its specific intended EEG waveform. This requires intensive optimization since adjacent filter designs have dependencies on each other and are constrained by their cut-off frequencies with regard to specific EEG waveform bands. The second major consideration is how to formally apply these entropy measurements to the



(a) Neural Oscillation

(b) Functional Isolation

Figure 6.1: A rudimentary depiction of a standard neural oscillation in column 1 and a neural oscillation where functional neuronal isolation occurs, altering the local field potential column 2. These two local field potentials in black are demonstrated in columns 1 and 2 in the first row. The characterized intensity of the signal over time is in blue. The second row of figures are the simulated spikes of individual neurons, where each dot represents an action potential in space and time. The red dots in the second column are the neurons that were suppressed and did not fire. The third row is the power spectrum using the Fourier transform of the local field potential neuronal oscillation. Intensity of the waveforms in the time domain and frequency domain are both highlighted in blue in rows one and three.

proposed signal processing methodologies. Although the Wavelet Entropy method analyzes the signal in a multi-resolution approach, it does not analyze instantaneous changes in signal (granular time resolution), but instead examines entropy within the windowed time segment as a normalized sum of energy across the entire window. Thus, the resolution of any changes in the signal are generalized to the size of the window and do not examine these instantaneous changes in signals, which can inform granular complex changes in firing. Finally, one must consider how to demonstrate that the new proposed approaches provide significant explanatory features when capturing these minuscule granular intensity changes from global field potentials.

Insights. First, we test the feasibility of the proposed designed EEG algorithms using hypoxia data. Hypoxia is a state in which the body is unable to provide adeqeate levels of oxygen to its tissue. When oxygen levels are adequate, proper signal propagation between neurons can occur [51, 118]. However, when oxygen deprivation occurs, the energy substrate supplied for neurons, Adenosine Triphosphate (ATP), is depleted, preventing synaptic transmission to other neurons [98, 118]. The reduction of oxygen to tissue leads to neuronal electrophysiological isolation because of the inability to continue signal propagation [98,104], thus altering the global measurements of the EEG recordings [104, 118].

Von Tscharner developed a signal processing intensity analysis method designed for Electromyography (EMG) signals using a wavelet based analysis [154]. The design of these filters not only overcomes these time-frequency trade-offs, but the filters were designed in the frequency domain in order to minimize the plateau value of the filters to fully reconstruct the EEG signal [154]. The filter's center frequencies and bandwidths were not chosen to capture any specific frequency ranges but strictly designed to minimize the plateau value of the filter bank. Using optimization methods and these core concepts that von Tscharner presents, we can augment the current design to fit appropriate optimal bandwidths specifically for continuous EEG analysis.

Contributions We propose a filter bank approach that addresses both the aforementioned challenges by examining the intensity of band-limited frequencies relevant to EEG in continuous time. Utilizing this developed intensity approach allows one to analyze entropy as a function of both time and frequency, unlike any current method available. We coin the term "EEG Activation Complexity" to refer to the calculation of entropy as the timing between a frequency band's peak intensities and Rank-Order Complexity as the measurement of how the intensity rank alters over time.

1. Utilizing synthetic stationary and non-stationary signals, we capture a one-to-one mapping of intensity for the designed filters as a function of time.

2. We demonstrate that the timing of intensity peaks over a band-limited frequency is significantly less complex during normal oxygen conditions as compared to hypoxia conditions.

3. We demonstrate that a set of band-limited intensities have a more random fluctuation during hypoxia than during normal oxygen conditions.

4. We demonstrate that the two proposed methods provide more information and add another dimension to the analysis of EEG signal processing.

6.2 Methods

The methods section is partitioned into four sections. The first and second section discuss the data, filter design, and optimization methods applied to produce the time-frequency intensity analysis. The third and fourth section describe how we apply entropy calculations to the proposed time-frequency intensity analysis methods, where we introduce the Activation Complexity and the Rank Order Complexity, respectively.

6.2.1 Hypoxia Data Set

The dataset was collected by a research team at NASA Langley Research Center (LaRC) who subjected 49 volunteers with current hypoxia training certificates to normobaric hypoxia to study the impact of hypoxia on aircraft pilot performance [141, 142].

The goal of the study was to understand cognitive impairment due to exposure to mild hypoxia in order to develop and test psychophysiologically-based adaptive automation/autonomous systems. Subjects in the study experienced simulated altitudes of sea level (21% O2) and 15,000 feet (11.2% O2) induced by an Environics, Inc. Reduced Oxygen Breathing Device (ROBD-2). During non-hypoxic (i.e. sea level) and hypoxic exposures, each subject experienced three 10-minute bouts performing three different tasks consisting of a battery of written tests, Multiple Attribute Task Battery [131], and flight simulation tasks. In each exposure the research team collected task performance measures, a subjective self-report of workload (NASA Task Load Index) [67], and multiple physiological responses (including EEG). This article discusses only the EEG data collected during hypoxic and non-hypoxic exposures for the MATB, where the electrode configuration is provided in Figure 6.2. The specified electrode placement was used to avoid complications with the aviator's oxygen mask component of the ROBD-2 breathing device worn using straps around the subject's head.

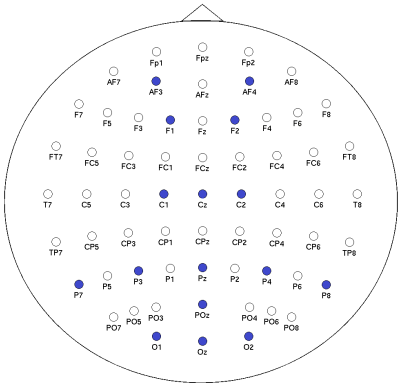


Figure 6.2: The electrode placement for the study design is highlighted in blue.

6.2.2 Filter Bank Intensity Method

The goal of our filter bank design is to develop a time-frequency intensity analysis for EEG-specific frequency bands. The general underlying concept for our filter design is to extract the intensity of the signal as a function of time and was motivated by von Tscharner [154]. In the proposed method, the filter bank comprises a collection of filters which, when summed, result in a relatively low plateau value across a range of frequencies (i.e. no one particular frequency dominates over any other).

Basis Function Definition

The spectral topology of a filter bank's basis function is important because adjacent filters in the frequency domain must be summed to obtain a reasonably stable plateau value. Von Tscharner implemented a derivative of the Paul wavelet, where the filter bank design had arbitrary cutoff frequencies for the filters. Von Tscharner's design was acceptable for analysis of EMG signals since it was only concerned with covering the entire range of possible frequencies and maintaining a low plateau value. However, this design choice is not suitable for our application, where individual EEG bands need to be extracted. Additionally, we found that the Paul wavelet could not handle the additional constraints required to capture specific EEG waveform frequencies defined in Table 6.1. We adopt a pragmatic approach in this paper whereby optimization routines are used to find filter parameters that produce a reasonably optimal plateau value, EEG cutoff frequencies, and avoid the time consuming process of manually adjusting filter bank parameters for each filter bank component.

For our filter bank, we selected the "flattened" Gaussian [21] basis function topology to balance the extraction of frequency bands of the EEG spectrum while maintaining an acceptable filter bank plateau value. In the filter bank, the i^{th} filter ($i = 1, \ldots, K$, where K is the total

number of filters) in frequencies space is defined as

$$\hat{\psi}_i(f; fc_i, a_i, b_i) = e^{-a_i(f-fc_i)^2 - b_i(f-fc_i)^4} \cdot \Theta(f) \tag{6.1}$$

and is parameterized by the center frequency fc_i and the tuning parameters a_i and b_i. The Heaviside function, $\Theta(f)$ constrains the design to only positive frequencies ($f \geq 0$). The filter bank is constructed using non-linear scaling to the basis function ($\hat{\psi}$) in the time domain by shifting each center frequency, fc_i, and tuning the parameters a_i and b_i to achieve a better filter bank design. Since these tuning parameters are not constant and altered for each i^{th} filter, we refrain from referring to this design as a wavelet implementation and instead refer to it simply as a filter bank design. However, these filters maintain the same generalized basis function (a "flattened" Gaussian) and the core concepts of design followed by [154] while still fulfilling a wavelet's admissibility criterion.

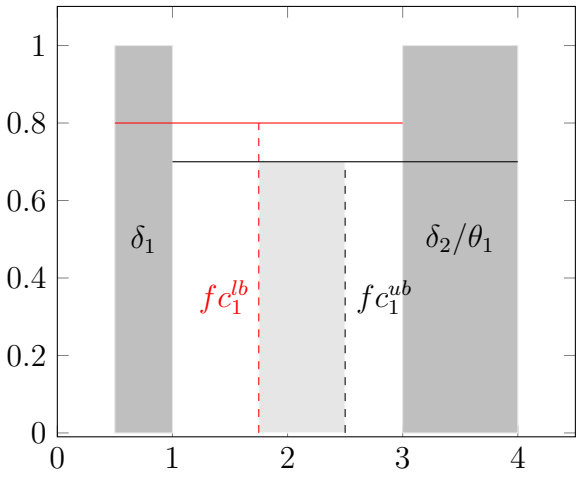


Figure 6.3: Figure depicting the calculation of center frequency acceptable range for the δ EEG band (i.e. $fc_\delta^{lb}, fc_\delta^{ub}$).

Filter Bank Optimization

The ultimate goal for our filter bank design was to find a set of filters with an acceptable (near optimal) plateau value. We attempted fitting filter parameters fc_i, a_i, and b_i by

hand, but acceptable results were time-consuming to achieve. We also tried optimizing the entire filter bank, but the results were not acceptable and were found to be computationally complex. We modified our approach to optimize sets of three filters at a time. We were able to achieve better results in less time with this approach, but we encountered difficulties in seeding the optimization routines with reasonable center frequencies. We attempted to solve this problem by optimizing only the placement of center frequencies such that the center frequencies were uniformly spaced with considerations for the boundaries imposed by EEG band cutoffs. We then used these center frequencies to seed the optimization of sets of three filters. After iterating through all filters in the filter bank, slight manual adjustments to the parameters a_i and b_i were made to better adjust for the plateau value.

Thus, we introduced the following generalized optimization approach for developing the proposed filter bank for the constraints of this particular EEG. A general overview of the method is as follows:

Step 1: Select a number of wavelets to represent each band of the EEG spectrum, and estimate the spacing of wavelet center frequencies by minimizing the sum of differences between separations of adjacent wavelet center frequencies.

Step 2: Use the set of center frequencies from Step 1 to approximate the optimal plateau value for the entire wavelet filter bank by optimizing sets of three wavelets.

The details of each of these steps is outlined in the following sections.

Step 1: Estimated Optimal Spacing

The goal of the Step 1 is to find a set of center frequencies ($fc \in \mathbf{R}^K$, where K is the total number of wavelets selected) to seed the optimization routine in Step 2. This set is found by minimizing the sum of differences between separations of adjacent wavelet center frequencies

as defined in the following optimization problem

$$\begin{aligned} \underset{fc}{\text{minimize}} \quad & \sum_{i=1}^{K-2}((fc_{i+1} - fc_i) - (fc_{i+2} - fc_{i+1}))^2 \\ \text{subject to} \quad & fc_i^{lb} \leq fc_i \leq fc_i^{ub},\ i = 1, \ldots, K, \end{aligned}$$

where fc_i^{lb} and fc_i^{ub} are the lower and upper constraint boundaries on the center frequencies. These boundaries can be considered either "hard" – set by the halfway point between the minimum lower and upper boundaries of the EEG band and the halfway point between the maximum lower and upper boundaries of the EEG band (as found in 6.1) or "soft" – given by the approximate acceptable regions of center frequencies that are not fully determined by the EEG band ranges in 6.1.

An example of a "hard" boundary exists for a single wavelet occupying the δ band (fc_1; starting between 0.5 Hz and 1 Hz and ending between 3 Hz and 4 Hz; see Figure 6.3). This constraint is considered "hard" because it is imposed by the adjacent θ band to the right and the undefined region of negative frequency to the left. This wavelet should have a center frequency that is between 1.75 Hz (fc_1^{lb}; halfway between 0.5 Hz and 3 Hz) and 2.5 Hz (fc_1^{ub}; halfway between 1 Hz and 4 Hz; space see Figure 6.3).

An example of a "soft" boundary exists for two wavelets occupying the α EEG band, which contains frequencies starting between 7 Hz and 8 Hz and ending between 12 Hz and 13 Hz (see Figure 6.4). The constraints on these two center frequencies are considered "soft" because the single EEG band (α in this case) is represented by two wavelets, and at least one of the boundaries for each wavelet is "artificially" imposed. For the case of the first filter in the α band, the θ EEG band is to the left (ending between 7 Hz and 8 Hz) for the first wavelet, but the upper bound is not well-defined since we have some choice as to the next α band filter.

These constraints help ensure that the cutoff frequency (fco_i) of each filter would take the

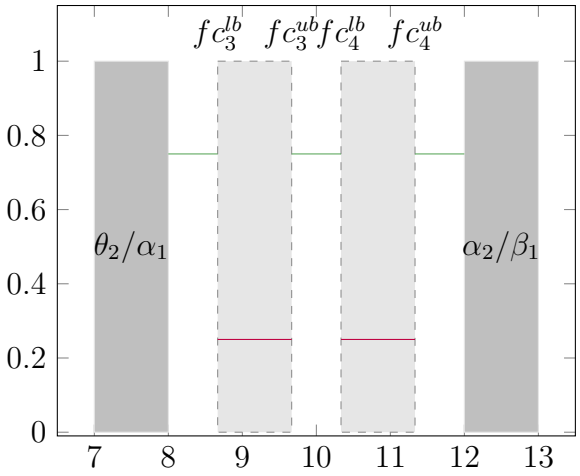


Figure 6.4: Figure depicting the calculation of multiple center frequency acceptable ranges for the α EEG band (i.e. $fc_\alpha^{lb}, fc_\alpha^{ub}$)

value of $1/e$ (i.e. $\hat{\psi}_i(f = fco_i) = 1/e$) between the frequency range specified by the boundaries of certain EEG bands during Step 2. Furthermore, center frequencies that are spaced evenly were found to produce more stable plateau values in Step 2. The aforementioned optimization problem was solved using the nonlinear constrained optimization routine (fmincon) in MATLAB.

Step 2: Optimize Basis Function Parameter Values The goal of Step 2 is to produce a filter bank that equally represents all EEG frequencies within the range of interest while providing reasonable separability of different EEG bands as defined in the scientific literature (see Tab. 6.1). One possible way to achieve such a filter bank is to find the set of filter parameters that minimize the path length integral of the sum of all filters in the filter bank. Colloquially, this amounts to finding the shortest distance between two points (a straight line in Euclidean geometry). Any deviations from a straight line result in having to "walk" a greater distance between the starting and ending frequencies. For an entire set of K filters, this objective can be operationalized as the path length integral between the first and last center frequencies,

written as

$$\underset{fc,a,b}{\text{minimize}} \quad \int_{fc_1}^{fc_K} dfL \tag{6.2a}$$

$$\text{subject to} \quad f_j^{lb} \leq fc_j \leq f_j^{ub} \tag{6.2b}$$

$$a_j^{lb} \leq a_j \leq a_j^{ub} \tag{6.2c}$$

$$b_j^{lb} \leq b_j \leq b_j^{ub} \tag{6.2d}$$

$$f_j^L \leq fco_j^L(fc_j, a_j, b_j) \leq f_j^H \tag{6.2e}$$

$$f_{j+1}^L \leq fco_j^U(fc_j, a_j, b_j) \leq f_{j+1}^H \tag{6.2f}$$

$$\hat{\psi}\left(f = fc_{j-1}; fc_j, a_j, b_j\right) \leq \epsilon \tag{6.2g}$$

$$\hat{\psi}\left(f = fc_{j+1}; fc_j, a_j, b_j\right) \leq \epsilon \tag{6.2h}$$

$$\text{where} \quad j = 1, \ldots, K \tag{6.2i}$$

where $fc, a, b \in \mathbf{R}^K$ and L is the arc length of the sum of all filters between the center frequency of the first and last filter in frequency space

$$L(f; fc, a, b) = \sqrt{1 + \left(\frac{d}{df}\left(\sum_{i=1}^{K} \hat{\psi}_i(f; fc_i, a_i, b_i)\right)\right)^2}. \tag{6.3}$$

As in Step 1, the possible ranges of center frequencies are constrained (Eq. 6.2b) as well as the parameters a and b (Eq. 6.2c and Eq. 6.2d). Additionally, Eq. 6.2e and Eq. 6.2f ensure that the lower and upper cutoff frequencies $fco_j^L(fc_j, a_j, b_j)$ and $fco_j^U(fc_j, a_j, b_j)$ for the j^{th} filter given by

$$fco_j^L(fc_j, a_j, b_j) = fc_j - \sqrt{\frac{\sqrt{a_j^2 + 4b_j} - a_j}{2b_j}}, \tag{6.4}$$

and

$$fco_j^U(fc_j, a_j, b_j) = fc_j + \sqrt{\frac{\sqrt{a_j^2 + 4b_j} - a_j}{2b_j}} \tag{6.5}$$

fall within the acceptable ranges for the associated EEG band. Finally, in order to produce a reasonably biorthogonal filter bank, the value of the j^{th} filter at the center frequencies of the $(j-1)^{th}$ and $(j+1)^{st}$ filters are constrained to be less than or equal to $\epsilon = 0.0005$ through constraints Eq. 6.2g and Eq. 6.2h.

Unfortunately, our attempts at directly optimizing Eq. 6.2a were met with poor results. However, we were able to approximate the global optimum by sequentially considering only three filters at a time until all K filters' parameters were determined (see Fig. 6.5). As such, Eq. 6.2a was modified to account for three filters ($\hat{\psi}_i$, $\hat{\psi}_{i+1}$, and $\hat{\psi}_{i+2}$) at a time as described in the following optimization problem

$$\underset{fc,a,b}{\text{minimize}} \quad \frac{\int_{fc_i}^{fc_{i+2}} df\,[L - (fc_{i+2} - fc_i)]}{(fc_{i+2} - fc_i)} \tag{6.6a}$$

$$\text{subject to} \quad fc_j^{lb} \leq fc_j \leq fc_j^{ub} \tag{6.6b}$$

$$a_j^{lb} \leq a_j \leq a_j^{ub} \tag{6.6c}$$

$$b_j^{lb} \leq b_j \leq b_j^{ub} \tag{6.6d}$$

$$f_j^L \leq fco_j^L(fc_j, a_j, b_j) \leq f_j^H \tag{6.6e}$$

$$f_{j+1}^L \leq fco_j^U(fc_j, a_j, b_j) \leq f_{j+1}^H \tag{6.6f}$$

$$\hat{\psi}\,(f = fc_{j-1}; fc_j, a_j, b_j) \leq \epsilon \tag{6.6g}$$

$$\hat{\psi}\,(f = fc_{j+1}; fc_j, a_j, b_j) \leq \epsilon \tag{6.6h}$$

$$\text{where} \quad j = i, i+1, i+2 \tag{6.6i}$$

where the arc length along the three consecutive filters L is given by

$$L(f; fc, a, b) = \sqrt{1 + \left(\frac{d}{df}\left(\hat{\psi}_i + \hat{\psi}_{i+1} + \hat{\psi}_{i+2}\right)\right)^2}, \tag{6.7}$$

and the constraints given in Eq. 6.6b to Eq. 6.6h serve the same functions as with the optimization problem in Eq. 6.2a.

For the first three filters, fc_1, a_1, b_1, fc_2, a_2, b_2, fc_3, a_3, and b_3 are found such that Eq. 6.6a is minimized. The optimal values of filter 3 are then used in the next round to find fc_4, a_4, b_4, fc_5, a_5, b_5 such that Eq. 6.6a is minimized. This process is repeated until all K filter parameters have been optimized (see Fig. 6.5).

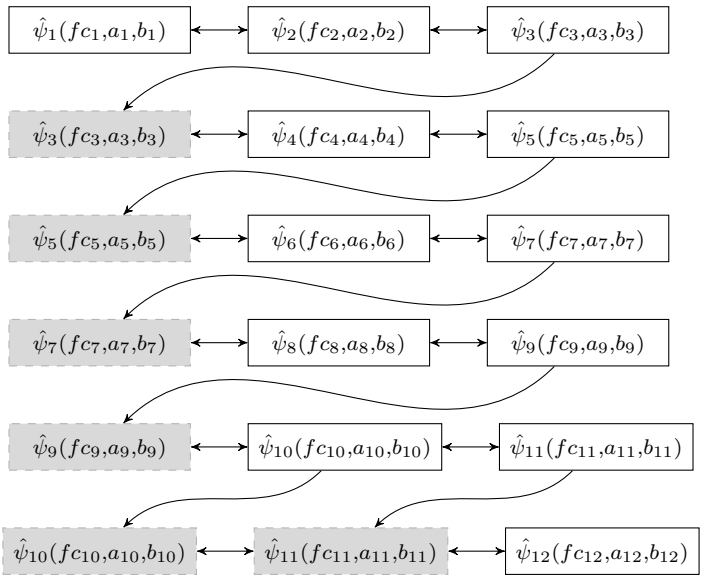


Figure 6.5: Diagram showing filter parameters optimized during each round of Step 2. Each row corresponds to an optimization round. Boxes shaded gray represent filters whose parameters were fixed during an optimization round.

As with Step 1, the method in Step 2 was implemented using the nonlinear constrained optimization (fmincon) routine in MATLAB.

Table 6.2: Filter Bank Parameters ($P_v = .0091$)

Filter i	Cf_i **(Hz)**	a_i	b_i	Fp_1 **(Hz)**	Fp_2 **(Hz)**
1	2.349	.072	.095	0.6	4.0
2	5.605	.001	.077	3.8	7.6
3	8.759	.101	.119	7.2	10.4
4	11.400	.219	.161	10.0	12.8
5	13.859	.170	.180	12.4	15.2
6	16.608	.007	.135	15.0	18.2
7	19.627	.001	.127	18.0	21.4
8	22.792	.001	.095	21.0	24.6
9	26.094	.001	.090	24.2	28.0
10	29.432	.001	.088	27.6	31.2
11	32.820	.003	.078	31.0	34.8
12	36.307	.001	.070	34.4	38.2

Optimized Filter Parameters

Utilizing the proposed methodology in which the constraints of the filtering paradigm are accounted for and optimized, we obtain Table 6.2. These parameters are then applied to Equation 6.8 and shown in Figure 6.6, where we can note the plateau vector, $\mathbf{PV}(f)$, is defined as

$$\mathbf{PV}(f) = \sum_{i=1}^{K} e^{-a_i(f-fc_i)^2 - b_i(f-fc_i)^4} \cdot \Theta(f), \tag{6.8}$$

where $\forall f \in \overline{1, S}$, $S = F_s/2$, and F_s is the sampling frequency. The plateau value, P_v, is obtained by calculating the standard deviation of the vector $\mathbf{PV}$.

EEG Filter Implementation

This discussed method deviates from Von Tscharner's classical approach of the filter implementation because of the valid concerns presented by Gabriel [53]. This discussion in [53], points out how applying the designed filters to the EEG's source signal in the frequency domain, $X_s(f)$, is inappropriate since we are applying the Fourier transform to a non-stationary

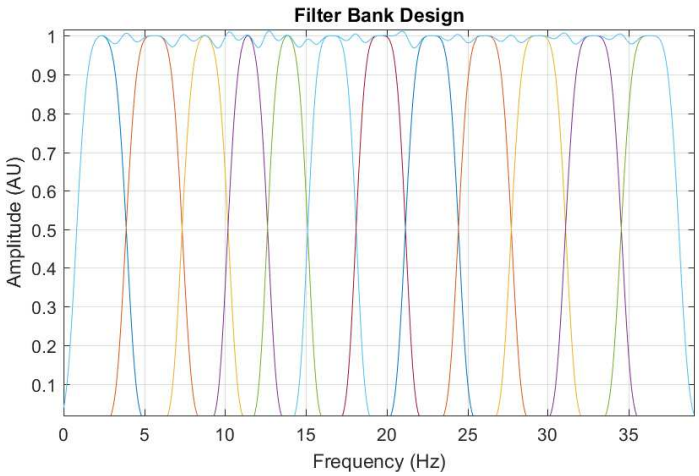


Figure 6.6: An optimized Filter Bank Version defined by the parameters provided in Table 6.2.

signal, thus defeating one of the major purposes of the novel signal processing approach. As Borg highlights [23], von Tscharner's implementation shares similarities to a basic equalizer which decomposes the EEG time domain's source signal, $x_s(n)$, into its associated intensity components, $\rho_i(n)$, with respect to each filtering process, κ_i, shown in Figure 6.7.

The presented filtering process utilizes the EEG signal in the time domain, $x_s(n)$, where we obtain a frequency band-limit intensity, $\rho_i(n)$, over time by applying convolution with the designed filters, $\hat{\psi}_i(n)$ and Gaussian smoothing methods. We define this entire process as, $\hat{\kappa}_i$, where $i \in \{1, \ldots, K\}$ filters. In order to obtain $\hat{\psi}_i(n)$, we transfer the designed respective frequency domain filter, $\hat{\psi}_i(fc_i, a_i, b_i)$, to the time domain by,

$$\hat{\psi}_i(n) = \mathcal{C}^{\mathcal{L}}\{\mathcal{F}^{-1}\{\hat{\psi}_i(fc_i, a_i, b_i)\}\}, \tag{6.9}$$

where $\mathcal{F}^{-1}$ is the inverse Fourier transform and $\mathcal{C}^{\mathcal{L}}$ is the circular shift of the numeric out of the function, where $L = \frac{N}{2}$ and N is the length of the filter in the time domain. The $\mathcal{C}^{\mathcal{L}}$ operation with $L = \frac{N}{2}$ is equivalent to performing a FFT shift, which adjusts the mirroring image in the frequency domain. However, this sequence happens to be in the time domain. Thus, the sequence $\{x(0), \ldots, x(N-1)\}$ is cyclically shifted to $\{x(N/2), x(N-1), 0, \ldots, x(N/2-1)\}$.

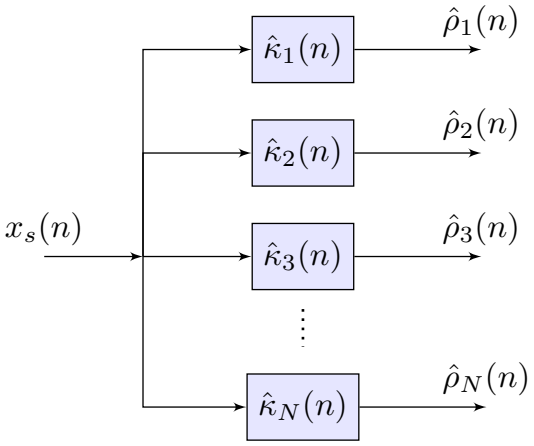


Figure 6.7: A Generalized Filter Bank design that mimics an equalizer in which the input signal $x_s(n)$ is decomposed into their respective intensity components, $\hat{\rho}_i(n)$, via the filtering paradigm, $\hat{\kappa}_i$.

By applying the equation 6.9, we are able to move the filter designed in the frequency domain to the time domain described with real and imaginary components, shown in Figure 6.8.

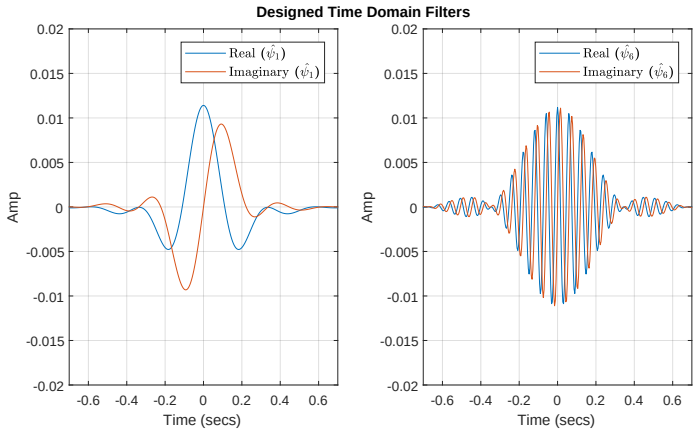


Figure 6.8: Visual representation $\hat{\psi}_i(n)$ and $\hat{\psi}_i(n)$ in the Time Domain.

Utilizing the filter in time domain, $\hat{\psi}_i(n)$, we obtain the intensity of the signal $x_s(n)$ by,

$$\rho_i(n) = 2 \mid \hat{\psi}_i * x_s, \mid \tag{6.10}$$

which is defined by the convolution of $x_s(n)$ with $\hat{\psi}_i(n)$. The intensity sequence, $\rho_i(n)$, is

then smoothed using a Gaussian filter,

$$G_f(n) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-0.5(\frac{x}{\sigma})^2} \tag{6.11}$$

where $\sigma = \frac{F_s}{2}$, F_s is the sampling frequency and $x = \in \{\frac{-3F_s}{2}, \ldots, \frac{3F_s}{2}\}$. The implementation of the smoothing is achieved by

$$\hat{\rho}_i(n) = \rho_i(n) * G_f(n), \tag{6.12}$$

where we obtain a smoothed filtered intensity EEG sequence.

6.2.3 Activation Complexity:

Activation Complexity is used to examine the predictability of how intensity in a sequence, $\hat{\rho}_i(n)$, peaks as a function of time over a specific EEG frequency range (e.g., α, δ). This is achievable with the proposed filter bank design, a peak detector, and a temporal entropy measurement (e.g. Sample Entropy). The peak detector implemented in this work utilized the function “findpeaks" from MatLab version 2017a, where the function will produce a vector of indices for the locations in time where the peaks occur, $A_i(k)$. In Figure 6.9, we depict instances in time for the peaks of the intensity of the delta waveform frequencies in the top part of the figure. Below in the Figure 6.9 is the vector$\Delta\mathbf{A}_1$, the sequence of the timing differences between all the peaks of the intensity waveform simply by

$$\Delta\mathbf{A}_i = [A_i(2) - A_i(1), \ldots, A_i(n) - A_i(n-1)]. \tag{6.13}$$

Following this, we processed the entropy of the new sequence, $\Delta\mathbf{A}_i$, by using sample entropy [85]. This reported entropy value from the sequence, $\Delta\mathbf{A}_i$ produced our Activation Complexity measurement Ac_i for each EEG lead. It is important to note that the type

of entropy measurement employed to the sequence will be sensitive to the number of data points in the sequence, thus limiting the window size that can be analyzed. Typically sample entropy and permutation entropy require a minimum of 100 samples, whereas approximate entropy requires a minimum of 1000 samples [85].

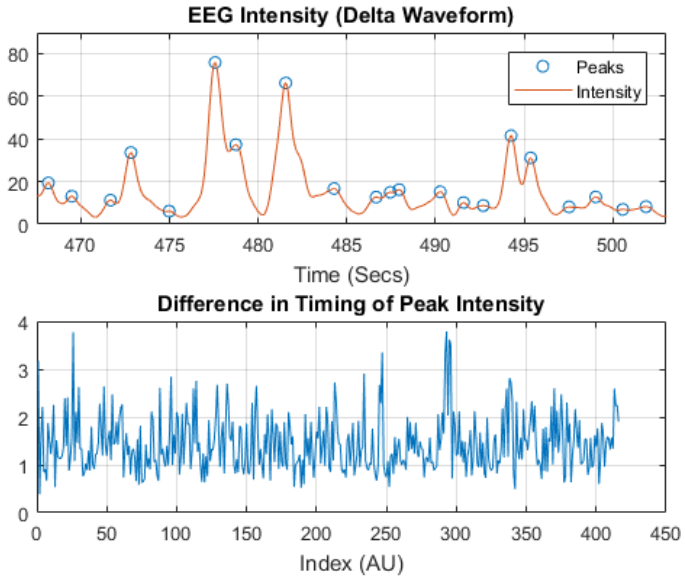


Figure 6.9: This is a pictoral representation of the process for calculating activation complexity. The top figure is the intensity of the designed delta waveform band, $\rho_1(n)$, in red. The blue circles are the indicated peaks on the intensity band, $\rho_1(n)$. The differences in time between those peaks are calculated providing $\Delta A_i(k)$, shown in the bottom image. A standard entropy measurement method can then be applied to this sequence.

6.2.4 Rank Order Complexity:

Rank Order Complexity is used to examine how the intensity of specific frequency bands (i.e. filters) oscillate continuously over time. If the filter's rank order stays the same, the entropy of the windowed time segment is low and thus predictable. However, if the rank of the intensity of the filter is changing over time, the entropy will be higher and thus less predictable. This can be visualized in Figure 6.10. We calculate this entropy measurement in a manner very similar to permutation entropy [164]. However, instead of a single one-dimensional sequence, we examine the signal in two dimensions, where the rank is evaluated across the filter bank

for each time step. For a given filter bank design with D filters which follow the $\hat{\kappa}_i$ paradigm with a given time series $x_s(n)$, we produce D intensity sequences, $\hat{\rho}_i(n)$, where $\forall i = 1, \ldots, D$ and $\forall n = 1, \ldots, N$. This can be notated as an intensity matrix,

$$\mathbf{P} = \begin{bmatrix} \hat{\rho}_1(1) & \hat{\rho}_1(2) & \ldots & \hat{\rho}_1(N) \\ \hat{\rho}_2(1) & \hat{\rho}_2(2) & \ldots & \hat{\rho}_2(N) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\rho}_D(1) & \hat{\rho}_D(2) & \ldots & \hat{\rho}_D(N) \end{bmatrix}. \tag{6.14}$$

Similar to how all other entropy measurements examine a vector sequence with a defined template length [85] or an embedded dimension [164], our embedded dimension is defined by the number of filters in the filter bank, D. Thus we examine N column vectors composed from $\mathbf{P}$, which consist of D subsequent intensity values. Each column vector,

$$P_n = [\hat{\rho}_1(n), \hat{\rho}_2(n), \ldots, \hat{\rho}_{(D-1)}(n), \hat{\rho}_{(D)}(n)], \tag{6.15}$$

has a specific rank order pattern which is defined as unique permutation, $\pi_k = (r_1, r_2, \ldots, r_D)$ of $(1, 2, \ldots, D)$, which fulfills

$$\hat{\rho}_1(n) \leq \hat{\rho}_2(n) \leq \ldots \leq \hat{\rho}_D(n). \tag{6.16}$$

For the purposes of clarification, a numerical example starting from notation of the intensity matrix,

$$\mathbf{P} = \begin{bmatrix} 1.2 & 1.3 & 2.2 & 2.2 \\ 2.1 & 1.1 & 1.1 & 1.2 \\ 1.1 & 2.3 & 2.1 & 1.1 \end{bmatrix}, \tag{6.17}$$

where embedded dimension is $D = 3$ and the length of time series is $N = 4$. $\mathbf{P}$ produces 3 unique permutations, where $p_1 = [1.2, 2.1, 1.1]$ is $\pi = (312)$, $p_2 = [1.3, 1.1, 2.3]$ is $\pi =$

(213), $p_3 = [2.2, 1.1, 2.1]$ is $\pi = (231)$, and $p_4 = [2.2, 1.2, 1.1]$ is $\pi = (321)$. Therefore, the total number of unique permutations that can be achieved is $D! = 3! = 6$, and π_i denotes the number of occurrences for each unique permutation, where $i = 1, \ldots, D!$. Rank order complexity, R_C, can then be defined as,

$$R_c = -\frac{1}{log_2(D!)} \sum_{1}^{D!} \pi_i log_2(\pi_i). \tag{6.18}$$

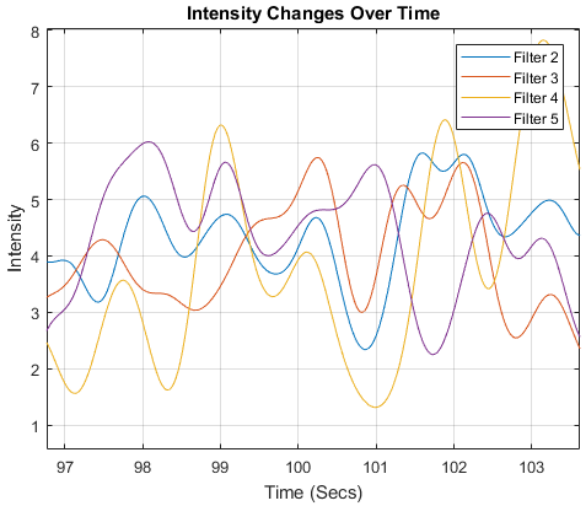


Figure 6.10: This figure provides a visual representation of the complex nature of how the intensity of the designed frequency bands alter in rank over time.

6.3 Evaluation and Discussion

In this section, we address the following research questions regarding the filter design intensity method and entropy approaches to distinguish changes in EEG firing patterns during conditions of 15,000 feet of altitude versus sea level altitude:

RQ1 Can we accurately depict the intensity of the stationary and non-stationary signals proportionally to the original time-series?

RQ2 Does the proposed Activation Complexity method demonstrate the ability to extract

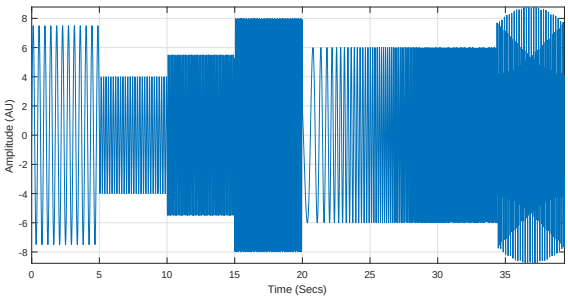
long-complexity EEG dynamic trends to distinguish between brain activity under hypoxia?

RQ3 Does the proposed Rank Order Complexity method demonstrate the ability to extract short-complexity EEG dynamic trends to distinguish between brain activity under hypoxia?

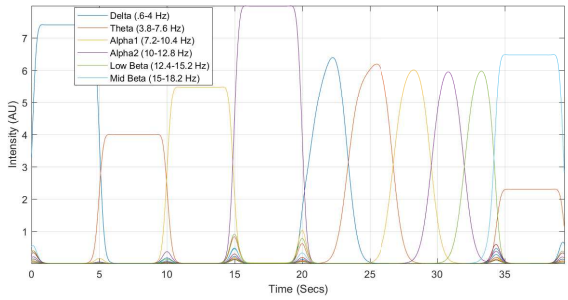
RQ4 How do classical methods perform on distinguishing differences between hypoxia and sea level changes in the brain?

6.3.1 RQ1: Simulated Continuous EEG Intensity Measurement

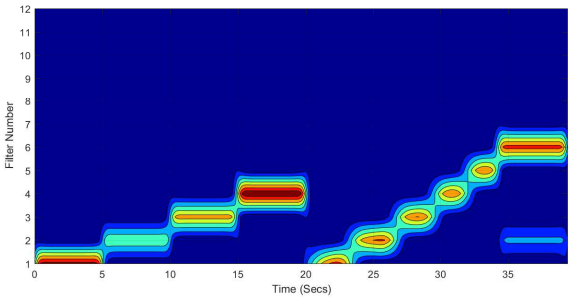
We hypothesize that using the proposed filter bank methodology, we can extract pertinent EEG frequency band intensity continuously over time. Utilizing the designed filter bank methodology, we simulate various stationary and non-stationary waveforms to evaluate intensity as function of time. In Figure 6.11a, we first modeled four stationary waveforms at frequencies of 2.3 Hz, 5.6 Hz, 8.75 Hz, and 11.4 Hz with amplitudes of 7.5, 4, 5.5 and 8, respectively. The fifth component is a non-stationary signal model using a chirp in which the frequency linearly increases from 0 to 15 Hz with an amplitude of 6. We then provide an example of a linear combination of two stationary signals at 2.3 Hz and 16.6 Hz with amplitudes of 2.3 and 6.5, respectively. All of these waveforms were concatenated together as a single time series. Thus, the transitions between waveforms were abrupt and discontinuous, causing mild perpetuations in irrelevant filters to activate. Through visual inspection in Figure 6.11b, we can obtain an equivalent proportionality to the simulated waveform generated in Figure 6.11a. Figure 6.11b can be represented in a two-dimension fashion, similar to how continuous wavelet transforms are presented using contour plots shown in Figure 6.11c. This allows a clearer depiction of time, frequency (i.e. filter number) and intensity.



(a) A Synthetically Generated Time Series

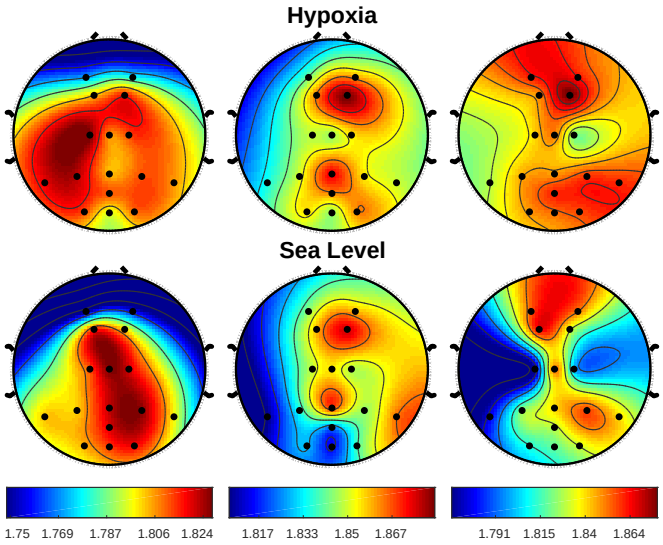


(b) 1-D Representation of the Filter Bank Output



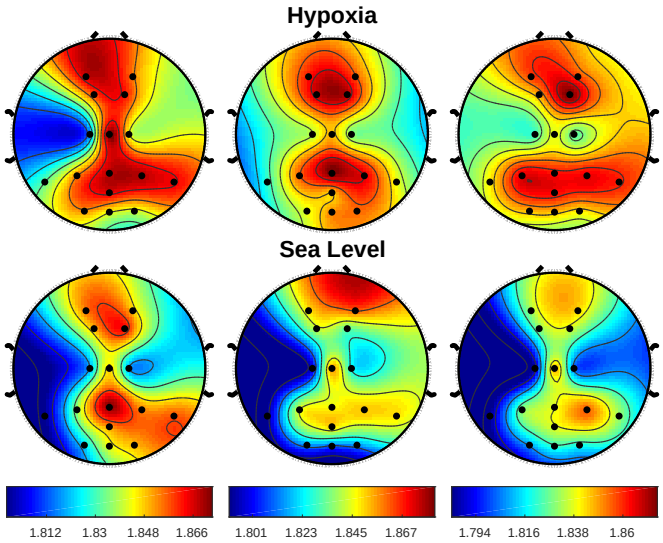
(c) Contour Plot of the Filter Bank Output

Figure 6.11: The synthetic time series consists of various stationary and non-stationary frequencies (a chirp from 20-30 seconds). We demonstrate two visual representations of the filter bank output. The 1-D representation allows for an enhanced comparison of an accurate depiction of intensity to the synthesized time series. The contour plot allows for a better global visualization of which filters are activated and their timing.



(a) Ac_1 Analysis (b) Ac_2 Analysis (c) Ac_3 Analysis

Figure 6.12: AC Analysis, Ac_i , is shown between the hypoxia and sea level cohorts for the first three intensity filter designs (delta $[0.6 - 4.0Hz]$, theta $[3.8 - 7.6Hz]$, low alpha $[7.2 - 10.4Hz]$), where i is the applied filter. Ac_1 demonstrated no significant changes across any of the EEG sites. Ac_2 demonstrated a significant increase in entropy (i.e. complexity) at the Po_z EEG site with a p-value of 0.033. Ac_3 produced an increase in complexity at the P_7, F_2, and C_1 with p-values of 0.046, 0.004, 0.0001, respectively



(a) Ac_4 Analysis (b) Ac_5 Analysis (c) Ac_6 Analysis

Figure 6.13: Ac_4 demonstrated no significant changes across any of the EEG sites. For Hypoxia, Ac_5 demonstrated a significant increase entropy (i.e. complexity) at P_7, P_z, O_z, O_2, F_1, F_2, C_2 with P-Values 0.049, 0.021, 0.006, 0.007, 0.033, 0.018, 0.028, respectively. During Hypoxia Ac_6 produced a significant increase in complexity at the P_7, O_1, F_2, C_1, P_3 with p-values 0.0003, 0.049, 0.017, 0.0464, and 0.033, respectively. Its also worth noting that O_1 had p-values of .065 and .060 for Ac_4 Ac_5, respectively

6.3.2 RQ2: Activation Complexity

Activation Complexity (AC) is applied to a hypoxia data set, where we hypothesize that this novel method can extract irregular neuronal firing patterns from global EEG recordings. This method was applied to EEG data collected from 49 subjects exposed to three 10 min bouts of normobaric hypoxic (12% O_2/15k ft) and non-hypoxic conditions (22% O_2/sea-level) at NASA Langley Research Center. One sample t-tests were used between the two cohorts for the performed computerized MATB task bout. The AC analysis using sample entropy used a template length of m=2 and a threshold value of r=.25, producing 26 different statistically significant AC measurements across filters and EEG regions, shown in Figures 6.12, 6.13, 6.14 and 6.15. We explored various other threshold values of .15, .2 and .3, which produced 17, 19, and 25 statistically different activation complexities, respectively. The AC measurements that demonstrated significant changes utilizing the other threshold values demonstrated similar patterns with regard to EEG leads and frequency band. The calculated AC complexity was normalized across all EEG leads, incorporating both the hypoxia and sea level cohorts associated to each filter bank's intensity analysis. This was done to highlight the differences in complexity produced in the colored contour map in Figures 6.12, 6.13, 6.14 and 6.15. The details of the results regarding EEG site location and p-value for the one-sample test (for $\alpha \leq .05$ with $N = 47$) are provided in the figure comments below. Additionally, Table 6.3, is provided for more details regarding the means and standard errors for instances of $\alpha \leq .1$.

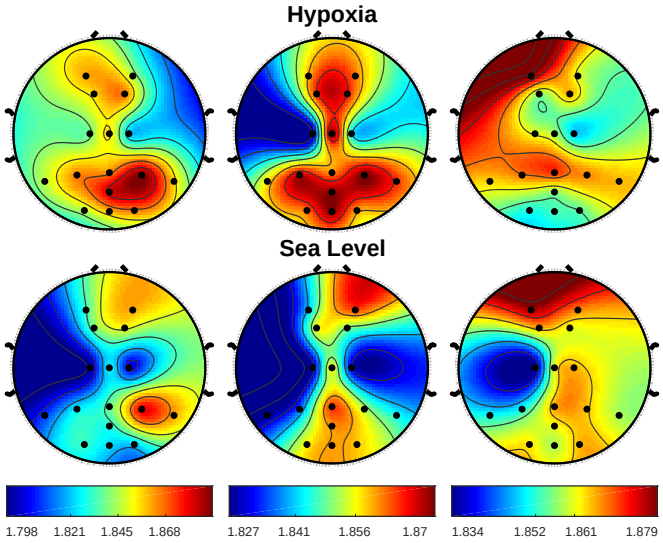
Utilizing this novel approach called *Activation Complexity*, we demonstrate that there is a significant increase in complexity during hypoxia across numerous EEG sites in the theta, alpha, and beta EEG frequency regions. However, the only significant decrease in complexity exists in the high frequencies in the gamma region. The sites that demonstrated this significant decrease in complexity never overlapped with the reported higher complexity EEG sites in the lower frequency regions. The rear left side of the brain, P_7, P_3, and C_1

had the most consistent amount of significant activation complexities across frequency bands having 5, 3, and 4, respectively. In the introduction, we discussed how hypoxia has been hypothesized to cause neuronal isolation in past literature [118]. This concept was pictorially demonstrated in Figure 6.1, where the time-frequency intensity peaks were more prominent for the functional isolation case.

We have found that this Activation Complexity method is not ideal, however, for small windowed segments of data. It is ideal for long-term trend analysis applications and has the potential to indicate small, subtle, anomalous patterns within the EEG spectral bands. Sample Entropy and other temporal entropy measurements typically require a minimum of 100 data points. The 10-minute segments that were analyzed only produced a mean of 431.2 peaks for each intensity frequency band, $\hat{\rho}_i(n)$. When analyzing simply the number of peaks in each, $\hat{\rho}_i(n)$, for each EEG lead against the hypoxia and sea level cohorts, the one t-test produced 18 significant p-values. Measuring how intensity is maximized and fluctuates as an informative measure is further supported. However, only 5 of the 18 significant p-values intersected with the 26 statistically significant different AC measurement. Thus, eluding to the fact that it is not simply the amount of peaks but the timing of these peaks, and intensity may hold a very different meaning when it comes to analysis of complex brain dynamics. Therefore, how a band-limited intensity is sustained may provide valuable information regarding neuronal firing and indicators for disease.

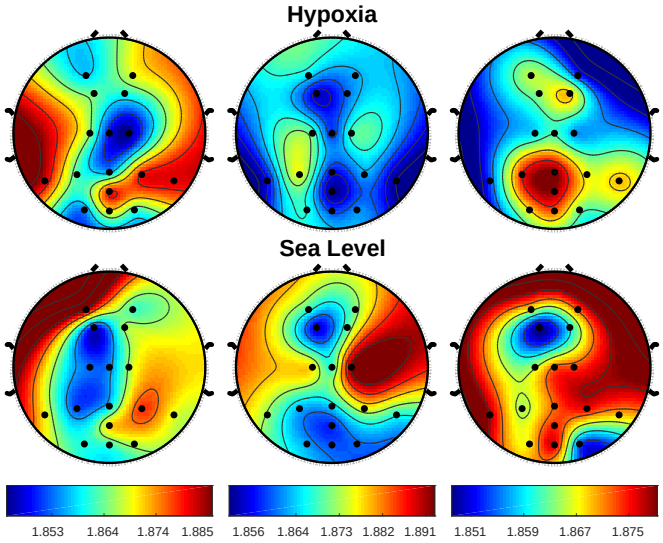
6.3.3 RQ3: Rank Order Complexity

The use of Rank Order Complexity (ROC) complexity is designed to capture short-term trends in EEG intensity dynamics. Although the data for this analysis is identical to the data used in RQ2, only specific instances when the subject was responding to multiple tasks were included (i.e., Communication and Tracking Tasks). Since the design was to examine differences in short bouts within the EEG, random sectioning was not appropriate. Thus,



(a) Ac_7 Analysis (b) Ac_8 Analysis (c) Ac_9 Analysis

Figure 6.14: The hypoxia cohort for Ac_7 produced an increase in complexity for EEG sites P_7, Po_z, P_z, 0_2, C_1, P_3 with p-values 0.036, 0.048, 0.026 0.040, 0.045, 0.035, respectively. The hypoxia cohort for Ac_8 produced an increase in complexity for EEG sites P_7, and P_3 with p-values 0.031 and 0.035, respectively. Ac_9 demonstrated no significant changes across any of the EEG sites.



(a) Ac_{10} Analysis (b) Ac_{11} Analysis (c) Ac_{12} Analysis

Figure 6.15: The hypoxia cohort for Ac_{10} produced an decrease in complexity for EEG sites AF_4 with a p-value of 0.019. Ac_{11} also produced a significant decrease in complexity at C_2 with a p-value of 0.047. Ac_{12} demonstrated no significant changes across any of the EEG sites.

Table 6.3: Activation Complexity

EEG Site	**Hypoxia** $Mean_H$	SE_H	**Sea Level** $Mean_{SL}$	SE_{SL}	**Significance** p-value
			Ac_2**:** *Frequency Range* $(3.8 - 7.6Hz)$		
$P0_z$	1.87	0.011	1.832	0.013	0.034
			Ac_3 : *Frequency Range* $(7.2 - 10.4)$		
P_7	1.844	0.018	1.801	0.025	0.046
O_1	1.861	0.017	1.830	0.020	0.061
O_z	1.865	0.012	1.836	0.013	0.083
F_2	1.888	0.010	1.848	0.009	0.004
C_1	1.855	0.020	1.791	0.025	0.0001
			Ac_4 : *Frequency Range* $(10.0 - 12.8)$		
P_7	1.846	0.0154	1.812	0.024	0.065
C_z	1.885	0.0089	1.860	0.013	0.098
O_1	1.864	0.012	1.836	0.017	0.064
			Ac_5 : *Frequency Range* $(12.4 - 15.2)$		
P_7	1.837	0.018	1.801	0.024	0.049
P_z	1.890	0.008	1.855	0.013	0.021
O_1	1.854	0.015	1.823	0.020	0.060
O_z	1.865	0.012	1.818	0.019	0.006
O_2	1.868	0.009	1.823	0.016	0.007
F_1	1.883	0.011	1.851	0.010	0.033
F_2	1.884	0.010	1.842	0.012	0.018
C_1	1.850	0.020	1.809	0.028	0.028
			Ac_6 : *Frequency Range* $(15.0 - 18.2)$		
P_7	1.849	0.016	1.794	0.022	0.0003
P_8	1.867	0.014	1.839	0.017	0.084
O_1	1.854	0.018	1.818	0.021	0.049
F_1	1.881	0.011	1.844	0.012	0.017
C_1	1.832	0.023	1.798	0.030	0.046
P_3	1.872	0.013	1.837	0.017	0.033
			Ac_7 : *Frequency Range* $(18.0 - 21.4)$		
P_7	1.850	0.017	1.814	0.025	0.022
C_z	1.858	0.010	1.832	0.014	0.068
PO_z	1.883	0.008	1.835	0.013	0.001
P_z	1.867	0.010	1.841	0.013	0.027
O_z	1.861	0.012	1.829	0.016	0.067
O_2	1.867	0.012	1.828	0.019	0.032
C_1	1.843	0.022	1.798	0.030	0.041
P_3	1.867	0.011	1.832	0.018	0.049
			Ac_8 : *Frequency Range* $(21.0 - 24.6)$		
P_7	1.859	0.016	1.827	0.021	0.037
P_8	1.868	0.011	1.846	0.011	0.094
P_3	1.873	0.012	1.838	0.013	0.029
P_4	1.877	0.009	1.854	0.010	0.052
			Ac_9 : *Frequency Range* $(24.2 - 28.0)$		
C_1	1.866	0.014	1.834	0.020	0.061
			Ac_{10} : *Frequency Range* $(27.6 - 31.2)$		
AF_4	1.867	0.009	1.895	0.009	0.019
			Ac_{11} : *Frequency Range* $(31.0 - 34.8)$		
C_2	1.874	0.009	1.90	0.010	0.047
			Ac_{12} : *Frequency Range* $(34.4 - 38.2)$		
P_7	1.857	0.009	1.882	0.010	0.072

we controlled for which bouts to sample from when subjects were prompted to perform dual tasks. However, for the task instances that were randomized within the experiment (where

some subjects performed more tasks than others and were independent of the different levels of oxygen saturation in the blood), a two sample t-test was performed. In addition, since we are not examining long-term EEG trends to extract how EEG dynamics are altered, we aim to control these instances by thresholding the data by the level of oxygen saturation at the time of the task. The hypoxia bouts had an oxygen saturation of less than 85%, where the mean was 81.42% and the sea level bouts required an oxygen saturation of 95% with a mean of 97.24%. This produced 215 hypoxia samples and 641 sea level samples. Since ROC incorporates all the intensities of various bandwidths into one ROC value the results are more concise, presented in Figure 6.16 and Table 6.4. These results demonstrate a significant increase in ROC for the parietal and occipital poles of the brain during the hypoxic MATB dual tasks for tracking and communicating. The interruption of an increase in ROC means that the intensity of these specific frequency bands are changing in a more chaotic fashion. We hypothesize that this reduced ROC on the front and central poles may be caused by a need for increased energy resources (i.e. sugar and oxygen) during hypoxic tasks, where engagement in these areas is not as demanding. We already know through literature that task difficulty increases as hypoxia is imposed on subjects, specifically in the parietal region in which demand is higher during tasks. Therefore, for the parietal and occipital areas to maintain a more ideal functional state, resources may be required at a high rate to maintain the subjects' engagement. It is additionally worth noting that this method is almost identical to calculating permutation entropy, which requires approximately 100 samples per measurement. Our analysis windowed 10 seconds of data with a sampling frequency of 256. This produced 2560 samples for ROC. Based on permutation entropy theoretical requirements, much smaller window sizes and their physiological implications should be explored.

Table 6.4: Rank Order Complexity

EEG	**Hypoxia**		**Sea Level**		**Significance**
Site	$Mean_H$	SE_H	$Mean_{SL}$	SE_{SL}	p-value
C_z	0.2670	$8.46x10^{-4}$	0.2676	$4.56x10^{-4}$	0.0002
PO_z	0.2740	$7.11x10^{-4}$	0.2734	$3.76x10^{-4}$	0.013
P_z	0.2698	$3.69x10^{-4}$	0.2688	$4.07x10^{-4}$	0.058
O_z	0.2753	$7.82x10^{-4}$	0.2748	$4.24x10^{-4}$	0.051
AF_3	0.2682	$6.73x10^{-4}$	0.2683	$3.68x10^{-4}$	0.015
AF_4	0.2684	$7.07x10^{-4}$	0.2677	$4.03x10^{-4}$	0.013
F_2	0.2660	$1.09x10^{-4}$	0.2666	$4.84x10^{-4}$	0.094
C_1	0.2642	$1.01x10^{-4}$	0.2669	$4.97x10^{-4}$	0.024
C_2	0.2623	$9.91x10^{-4}$	0.2652	$4.55x10^{-4}$	0.014
P_3	0.2732	$7.06x10^{-4}$	0.2722	$3.81x10^{-4}$	0.062
P_4	0.2730	$6.33x10^{-4}$	0.2729	$3.36x10^{-4}$	0.015

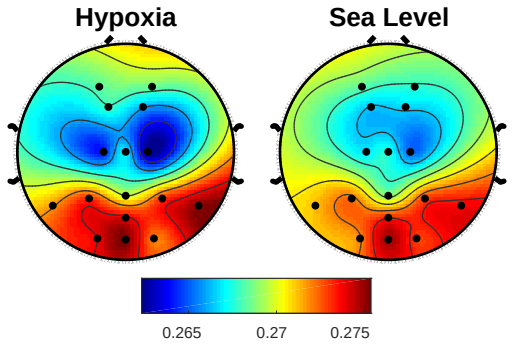


Figure 6.16: The Rank Order Complexity Method demonstrates significant decreases the mean ROC during hypoxic tasks in the Frontal and Central poles (C_z, C_1, C_2, and F_2), except for AF_4. However, a significant increase in mean ROC during hypoxic tasks in the Parietal and Occipital poles (PO_z, P_z, O_z, P_3, and P_4)

6.3.4 RQ4: Classical Methods

We hypothesize that classical EEG methods such as intensity (i.e power) analysis can still produce significant results, but that they provide an incomplete picture. Moreover, we specifically use spectral intensity analysis (SIA) since this method is the only approach in which we can compare effects on both time, spectral bands, and intensity. Currently, there is no complexity measurement that analyzes all of these components and thus a one-one comparison would not be possible. Since AC was designed for long trend analysis and specific frequency analysis across electrodes, whereas ROC is designed for short trend analysis for generalized global site data, we partition this discussion into parts.

Comparison for Activation Complexity: We discuss this hypothesis by first analyzing the changes in EEG intensity caused by hypoxia. Utilizing the filter banks' summed intensity values for each 10-min segment, we performed statistical one sample t-tests. Due to the fact that electrical conductance can change from subject to subject and thereby altering the intensity values, each subject's individual filter's intensity values were normalized. When individual subject normalization was not applied, no significant difference was found. However, with proper normalization to account for electrical conductance changes, we found 26 filter intensity values across the 16 leads that were significantly different. Since our aim is not to directly discuss the effects of intensity and its relationship to hypoxia, but rather to determine the significant indicators that an intensity analysis provides, for brevity, we will only provide the results of significant EEG sites and their filters rather than the 576 means and standard errors associated with the intensity data.

Spectra of No Change : The filters 3, 4, 10, and 11 which are associated to frequency bands (7.2-10.4 Hz), (10.0-12.8 Hz), (27.6-31.2 Hz), and (31.0-34.8 Hz), demonstrated no significant intensity change across any of the EEG electrodes for SIA. However, AC demonstrated significance for all four frequency bands, shown in Table 6.3. This demonstrates additional features that AC has extracted from the EEG signal.

Spectra of Significant Increases: SIA did demonstrate significant increases in intensity only for Filters 1, 5, 6, 7, and 12 for the hypoxia cohort for a variety of EEG leads. More specifically, Filter 1 showed a significant increase in intensity for the O_2 EEG site. Filters 1 and 5 showed an insignificant increase in intensity during Hypoxia for the O_2 EEG site with p-values of .020 and .038, respectively. AC demonstrated no significance for Filter 1. However, AC for Filter 5 demonstrated increases for the entire occipital region (O_1, O_z, and O_2) as well as other EEG sites. SIA for Filter 6 showed the largest change across EEG leads, exhibiting a significant increase in intensity during hypoxia for PO_z, P_z, P_8, O_1,O_z, O_2, AF_4, F_2, C_1, and P_4, with p-values of 0.0006, 0.007, 0.018, 0.015, 0.002, 0.024, 0.040, 0.039, 0.041,

and 0.009, respectively. All of these sites, except for 0_1 and C_1 are located on the right hemisphere. On the other hand, the AC measurement only reports significant increases on the left hemisphere (6 significant sites), demonstrating divergent findings across methods. SIA for Filter 7 also demonstrates significant increases in intensity for hypoxia for PO_z, F_1, F_2, and P_4 with p-values of 0.045, 0.030, 0.024, and 0.027. These regions are essentially the frontal lobe and left parietal region. AC provided 8 significant sites, but only overlaps with PO_z. The majority of AC are located in the posterior right part of the brain in the parietal and occipital region. SIA for Filter 12 exhibits a significant increase during hypoxia for O_z, with a p-value of 0.035, and AC reports an increase for P_7 during hypoxia. Overall, we can note that although both methods unanimously demonstrate significant increases with their respective method, there is very little to no overlap with regard to EEG regions of the brain and sites.

Spectra of Significant Decreases: With regards to decreases in intensity, only Filters 2, 8, and 9 showed a decrease in intensity during hypoxia across various EEG locations. Filter 2 showed a decrease in intensity for PO_z, P_z, O_z, and P_z with p-values of 0.038, 0.013, 0.0009, and 0.018, respectively. AC only reports PO_z as increased intensity during hypoxia. Filter 8 demonstrated a significant decrease during hypoxia for O_2, AF_3, and AF_4 with respective p-values of 0.035, 0.047, and 0.018. We can also note an increased AC for the parietal regions during hypoxia. Filter 9 has also AF_3, and AF_4, which exhibited a decrease in intensity with p-values of 0.029 and 0.026, respectively. AC only reported C_1 as a increase during hypoxia.

Results Summary: In summary we note that when comparing AC to SIA the two methods are very divergent in their reported findings, specifically on the direction of the measurement (increase vs decreasing), region of the brain, and spectral properties. These divergent results support the hypothesis that AC adds an additional dimensionality to the analysis. These SIA results share a resemblance to Papedelis's work with hypoxia, where they reported a

Table 6.5: Engagement Index

EEG	**Hypoxia**		**Sea Level**		**Significance**
Site	$Mean_H$	SE_H	$Mean_{SL}$	SE_{SL}	p-value
P_7	0.7109	$1.51x10^{-2}$	0.6701	$8.46x10^{-3}$	0.021
P_8	0.700	$1.66x10^{-2}$	0.6635	$7.80x10^{-3}$	0.032
0_1	0.6541	$1.46x10^{-2}$	0.6215	$7.75x10^{-3}$	0.047
C_1	0.5378	$1.38x10^{-2}$	0.5731	$7.77x10^{-3}$	0.030
C_2	0.4982	$1.32x10^{-2}$	0.5368	$7.28x10^{-3}$	0.012

increase in spectral power [118]. One caveat was that the majority of the spectral intensity findings were on the right hemisphere, whereas Papadelis reported left hemisphere dominance [118]. However, our subjects utilized their left hands for the MATB tracking tasks whereas the subjects in Papadelis's study used their right hand [118]. EEG asymmetries and cerebral lateralization in literature is well known and may explain the discrepancy between our results [25, 118].

Comparison for Rank Order Complexity: Next, we examined a typical way to measure bouts of engagement using the Engagement Index (EI) [122] and compare the proposed ROC method. The EI has been used in literature to monitor human engagement during tasks, which has been related to workload as task load increases [123]. EI is calculated using the proportion of beta over the sum of alpha and theta waves. Utilizing the proposed filter design, which proportionally captures intensity over time, we define EI as

$$EI = \frac{\sum_{n=1}^{\tau} \hat{\rho}_5(n) + \hat{\rho}_6(n) + \hat{\rho}_7(n)}{\sum_{n=1}^{\tau} \hat{\rho}_2(n) + \hat{\rho}_3(n) + \hat{\rho}_4(n)} \tag{6.19}$$

where $\hat{\rho}_i(n)$ is the intensity of the spectral band over time, n is the discrete time step, and τ is the length of the 10 second task bout. The results of comparing hypoxia and sea level changes during dual tasks for EI are presented in Figure 6.17 and Table 6.5, where any results with an alpha criteria of less that .1 was provided. The controls (i.e. tasks, oxygen saturation levels) and statistical tests were identical to the proposed ROC method. We note the similarity between ROC and EI, where there is a decrease in the central poles and

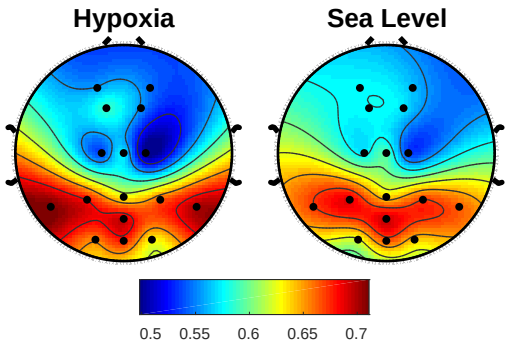


Figure 6.17: During hypoxic conditions, the EI demonstrates a significant decreases in the Central region (C_1, and C_2). Additionally, we note a significant increase in EI in the parietal and occipital poles (P_7, P_8, and O_1) during hypoxia.

a decrease in the parietal and occipital regions. However, the specific EEG sites are not identical, except for C_1 and C_2. EI and ROC further share similarities by which all the EEG sites are positively and linearly correlated using the pearson's correlation shown in Table 6.6 with all sites having a P value less than .001 . However, ROC has demonstrated 11 EEG sites to be significantly different versus EI only demonstrating 5 EEG sites to be significantly different. On the other hand, the only intersection of these sites are C_1 and C_2. It is not fully clear whether ROC is a more sensitive metric for engagement. These more chaotic intensity fluctuations could be strictly related to changes caused by neuronal firing associated with hypoxia (neuronal isolation) due to the increased workload or a combination of both. However, EI, mathematically, has the potential to capture these dynamic changes in intensity because of the proportion it sets between frequency bands, seen in Equation 6.19. Hence, if neuronal isolation did occur due to hypoxia, these changes in intensity for specific neuronal oscillations would alter the reported EI. However, ROC quantifies these dynamic intensity changes more granularly (each discretized instance) and across a larger set of spectral bands using entropy. In addition, ROC examines rank rather than intensity directly. This allows ROC to be more robust against large changes in intensity caused by noise.

Table 6.6: EI vs ROC Correlation

EEG Site	**Hypoxia** (ρ)	**Sea Level** (ρ)
P_7	0.7847	0.7184
C_z	0.7343	0.7815
PO_z	0.7544	0.7426
P_z	0.6741	0.7260
P_8	0.7261	0.7300
O_1	0.7982	0.7581
O_z	0.7274	0.7462
O_2	0.8001	0.8136
AF_3	0.5971	0.6500
AF_4	0.6505	0.6910
F_1	0.6368	0.5910
F_2	0.7153	0.7021
C_1	0.7881	0.7379
C_2	0.7414	0.7217
P_3	0.6579	0.6680
P_4	0.6274	0.7149

6.4 Conclusions

Our two methods are the only nonlinear dynamic analysis methods for isolated EEG frequency bands and intensity analysis. More specifically, this work demonstrates how we can recover a continuous reading of intensity for EEG signals and thus extract how intensity alters over time for specific frequency bands. This demonstrates a new predictive EEG feature for hypoxia and opens up a novel avenue for analysis of diseases and physiological sensory changes.

These two methods showed significant differences in hypoxic vs non-hypoxic states, facilitating future analyses (specifically for diseases related to stroke, ischemia, and cognitive changes). We hypothesize that these methods are extracting various forms of neuro-isolation but to fully support this hypothesis we require additional study designs.

This work ultimately opens an additional dimension of spectral and complexity analysis, opening the exploration of EEG signal for further explanatory analysis in the area of neuro- and cognitive science. The ability to isolate the intensity of neuro-oscillations as a function of time allows for further explorations into not only the timing of peak intensity, but addi-

tional multi-variate features such as their troughs and width dimensions. These additional complexity analyses can potentially address how these band-limited neuro-oscillations are sustained.

All models are wrong.

George E. P. Box

7

Naive Adaptive Probabilistic Sensor Fusion

In the area of human-machine interaction and human-cognitive research, machine learning (ML) problems become increasingly complex due to limitations in the experimental design, resulting in poor decision-making processes. More specifically, experimental study designs for avionic systems produce very few data instances, have large class imbalances, and generate wide data sets due the diverse amount of sensors. These problems are further exacerbated in anomaly detection cases where class imbalances occur and there are almost always more features than samples. Typically, dimensionality reduction methods (e.g., PCA, autoencoders) are utilized to handle these issues from wide data sets. However, these dimensionality reduc-

tion methods do not always map to a lower dimensional space appropriately and they capture noise or irrelevant information. In addition, when new sensor modalities are incorporated, the entire ML paradigm has to be remodeled because of new dependencies introduced by the new information. Remodeling these ML paradigms is time-consuming and costly due to lack of modularity in their design, which is not ideal. Furthermore, human cognitive research experiments, at times, creates ambiguous class labels because the ground truth data can not be agreed upon by subject matter experts, making ML paradigm even more challenging.

This work pulls insights from Dempster-Shafer (D.S.) Theory to address uncertainty for multi-classification ML problems caused by low samples, class imbalances, and wide data sets. We propose a probabilistic model fusion approach which combines (ML) paradigms built around bagging algorithms to overcome these experimental data concerns while maintaining a modular design for future sensor (new feature integration) and conflicting ground truth data. We demonstrate significant overall performance improvements using NAPS (an accuracy of 95%) in detecting human errors (a four class problem) caused by impaired cognitive states, when compared to other methodologies (an accuracy of 65%). This work potentially sets the foundation for other human-machine interaction systems.

7.1 Introduction

The impairment of cognitive states has significant downstream implications on human performance and the type of errors that are produced. Future human-machine interactions aim to close the gap between humans and machines, but one major hurdle to achieve this is for machines to understand the human. This understanding can come from the detection of human emotions, cognitive states, and the errors humans will produce before they happen. The ability to understand the human from a machine perspective would allow the machine to aid and compensate for task-related activities required by the human. However, from

the perspective of impaired cognitive states, a task related error can potentially come from various cognitive impairments or a combination of them. In order for a machine to directly adjust and assist the human, we must be able to predict the exact error caused by the cognitive impairment for the machine to intervene and remediate the event.

Under the umbrella of predictive cognitive research, a large amount of cognitive prediction literature examines human brain function with functional magnetic resonance imaging (fMRI) and generates reasonable machine learning (ML) results [50]. However, fMRI is not a feasible approach for highly granular, longitudinal studies or for evaluating people in open environments performing everyday tasks. Thus, real time monitoring multi-modal psychophysiological data is necessary to capture subtle changes in cognitive trajectories over time to detect task performance caused by cognitive impairments. Although analysis of a single modality can predict a specific cognitive state using traditional statistical methods [69]. This approach is insufficient for classifying multiple cognitive detection problem and overall lacks dimensionality to predict multiple classes [12, 128]. This is because different cognitive changes are detectable in patterns of varied combinations of physiological subsystems [12, 42]. This is because in real time open-world environments, cognition is multi-dimensional, requiring insight from numerous physiological subsystems, where data corruption is inevitable and subject-to-subject variability is prominent. This requires ML methods to handle these complex interactions and properly predict multiple classes of task relate errors. Thus, by utilizing ML methods we can determine a generalized approach to predicting psychological states.

Prior Work. Despite the considerable amount of literature that discusses the predictive cognitive psychophysiological research, some papers do not actually use "predictive modeling" but rather use "explanatory modeling," where only statistical significance is examined. These two terms are frequently conflated making the literature a bit ambiguous at first glance [137]. The majority of the other cognitive state prediction literature focuses on examining a single

physiological system to infer a cognitive state [69]. Thus the research literature for predicting task related errors caused by cognitive impairment is very sparse, especially when we it multi-class prediction problems. Even from the standpoint of attention-related human performance limiting states (AHPLS), which assesses engagement, fatigue, emotion, and workload, there is scant research for real time classification performance [10] and no accepted generalized approach to predict psychological states [114]. One major reason there is no generalized predictive approach for cognitive psychophysiological research is many of the published work for multi-class problems have not demonstrate exceptional performance on any particular dataset.

Challenges. For Human-Machine Interaction, large sets of annotated data are required in order to properly capture the variance of multiple cognitive changes and subject variability for proper classification. Thus, detecting task performance changes is a complex ML task which requires numerous modalities of streaming sensors (e.g., Electrocardiogram, Electroencephalogram, etc.) in order to capture the systemic psychophysiological changes [42]. These sensors create hundreds of features (predictive variables) which can potentially outnumber the quantity of tasks performed (response variables) in the experiment. The more modalities used to gain knowledge about underlying cognitive states, the more features we obtain, but task-related sample size can still limit the usable information. These problems are further exacerbated in anomaly detection cases (e.g., exposure to rare conditions) where class imbalances occur and there are almost more features collected than there are samples (e.g. wide data). Furthermore, in some cognitive experimental designs subject matter experts tend to disagree with data annotations, causing ambiguity for the ground truth annotated datasets [103]. All of these design problems caused by uncertainty are challenging to address due to a lack of training data instances, class imbalances, wide data sets (numerous modalities), subject-to-subject variability, and the inability to modularly scale to new sensors from previous ML paradigms. Thus these issues place uncertainty on the optimal decision boundary for the ML model and blunts its predictive power. So how can we overcome these ML

challenges?

Insights. Dempster-Shafer (DS) Theory is a framework equipped to deal with little to no prior knowledge, ignorance, and uncertainty [18, 157]. DS theory was developed for its ability to handle imperfect data in an effective and more intuitive manner [92]. DS Theory has three major caveats when compared to probability theory but can still be considered a generalization of probability [161]. The first main caveat between the two methods is that Bayesian approaches assume that the distribution, otherwise known as probability mass function (p.m.f), is fully defined. Therefore, Bayesian approaches conform strictly to axioms of probability [49, 56]. DS theory relaxes the axiom of additivity, by stating that if evidence is not provided or there is conflict, the "support" is assigned to the full set of propositions as uncertainty. Thus, DS Theory handles and quantifies uncertainty for distributions that are not fully defined due to incomplete information. The second caveat is that probability theory only assigns probabilities to "singletons" (in the example, $\{X\}$, $\{Y\}$ and $\{Z\}$). In order to determine probabilities associated with other propositions (e.g., $Pr(\{X \cup Y\}) = Pr(\{X\}) + Pr(\{Y\})$), we examine the union of the probabilities. DS theory, on the other hand, allows "supports" to be assigned to the complete power set of possibilities, meaning you can set support to various permutations within the set (i.e. doubleton, $\{X, Y\}$). Thus, DS Theory allows us to model ignorance by the way we set "supports" to propositions, because we are uncertain of which singleton proposition to support. Thirdly, DS approaches are conveniently designed to combine B.o.E., through fusion paradigms such as Dempster's Combination Rule (DCR) [112, 134]. These B.o.E. act as independent sources of information (i.e. A sensor Modality), but are combined to update the set of 'supports' for the propositions and uncertainty.

Although our approach does not need to be restricted to applying DS Theory, the core concepts of implementing similar frameworks is what potential can address the challenges previously discussed. Despite D.S. Theory's designed to handle uncertainty, the three major

part of the framework that can potentially overcome these challenges are: 1) the concept of combining independent sources of information (i.e., Combining multiple models) 2) the concept of setting support to various permutations of the proposition (i.e., augmenting the response variable).

First, this concept of fusing information from multiple Bodies of evidence (B.o.E.) together allows use to combining multiple subspaces of information in order to capture the full feature space of the data. This can be achieved by making a single B.o.E. represented a sensor or a small subset of features within the vector space. We can simply build numerous small models that are associated to a sensor or subset of sensors which are then fused together to expand our to a larger feature space. Thus the most important take away from this framework concept is our model can cover a larger feature space using fusion but our constraints caused by high dimensionality and small sample sizes are loosed allowing our model uncertainty is decreased. This enable us to avoid the application of dimensionality reductions to our data. These dimensionality reductions methods would potential fail to capture relevant information caused by the low amount of samples and class imbalances in the data. Furthermore, this also allows the addition of other B.o.E. (i.e. a new sensor modality) to be included in the detection paradigm for increased modularity and downstream iterative design.

Secondly, the concept of the framework allows the bodies of evidence (B.o.E.) to avoid being constrained to specifically support any single hypothesis (i.e. Fussy Classification). In real world cognitive performance problems subject matter experts do not always agree on the same cognitive state label. This framework allows the paradigm to give support to a label that contains multiple class within it (i.e., a doubleton, $\{X, Y\}$). This allows us to account for ignorance within the data. These labels with multiple classes essential merge both response variable together and their data. Thus the most important takeaway is that when we merge a response variable together we can increase sample size and simplify boundaries of the model's classification.

7.2 Methods

7.2.1 Experimental Design and Data Collection

The primary goal of the study was to initial verify and validate the cognitive capacity of human subjects undergoing normobaric hypoxia induction. Symptoms of hypoxia are shown to cause cognitive impairment causing lapses of attention, loss of situational awareness in operational contexts, temporary mental deficits and even complete incapacitation, all of which threaten safety of flight. It is evident that Hypoxia is a complicated phenomenon that can result in disastrous consequences in aviation in which Human-Machine interaction frameworks could pose as a vital solution. Thus, this work provides strong ML design case for predicting task performance errors caused cognitive impairments either during normal or hypoxic conditions.

The data was collected by a research team at NASA LaRC who subjected 56 volunteers with current hypoxia training certificates to normobaric hypoxia to study the impact on aircraft pilot performance. The dataset was later reduced down to 49 subject due to experimental obstacles and data attrition. The experimental session lasted approximately 4 hours. Subjects completed informed consent documentation. Subjects were briefed on the operation of the ROBD-2 and connected to physiological recording equipment. Subjects completed training sessions for each experimental task. Subjects sat quietly breathing room air while wearing masks to establish a physiological baseline, shown in Figure 7.1. Subjects performed each task three times under the following conditions: 1) breathing room air while wearing a mask, 2) breathing sea level gas mixture through a mask, and 3) breathing 15,000 feet gas mixture through a mask. Subjects recovered from hypoxia exposure by breathing 100% O2 for 2 minutes following 15,000 feet exposure. Subjects completed self-reported workload measures (NASA-Task Load Index, NASA-TLX) after each trial. After completing all trials, subjects were debriefed regarding the study purpose. Data from the NASA PEC study was

used to investigate the relationship between pilot physiology and performance under both hypoxic and non-hypoxic conditions. Subjects in the study experienced simulated altitudes of sea level (21% O2) and 15,000 feet (11.2% O2) induced by an Environics, Inc. Reduced Oxygen Breathing Device (ROBD-2). In the study, the pilots were administered a multitude of tasks comprising of the Cognitive Function Test (CFT) or the CogScreen-HE, the Multi-Attribute Task Battery (MATB), and X-Plane 10 flight simulator sessions (SIM). During all trials, multiple sensors collected pilot physiological responses including electrocardiogram (EKG), electroencephalogram (EEG), electrical dermal activity (EDA), oxygen concentration (O2), oxygen saturation (O2Sat), photoplethysmogram (PPG), and respiration. Since hypoxia has been found to cause cognitive and psychomotor deficits, we expected induced hypoxia to affect a pilot's ability to perform these tasks. Cognitive modeling of MATB performance captures the behavioral impacts of changes in workload, operator stress level, or fatigue levels caused by hypoxia and characterizes the high-level strategies engaged during continuous multi-tasking.

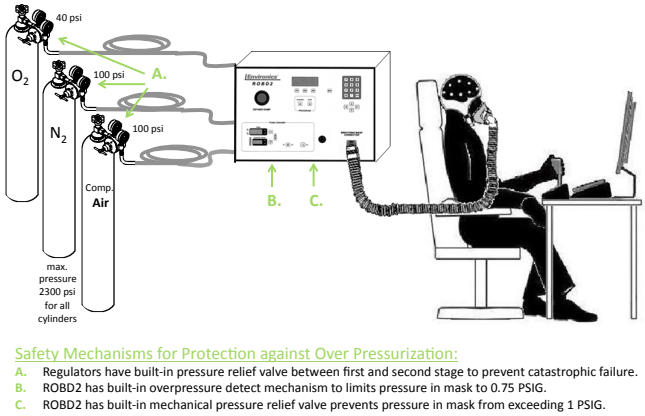


Figure 7.1: Environics ROBD2 System designed for inducing hypoxia, without changing atmospheric pressure

Multiple Attribute Task Battery (MATB)

The Multiple Attribute Task Battery (MATB) was use to provide important insight into the applied effects of performance. MATB was first developed in 1990 as a test designed

to evaluate operator performance and workload via a set of aviation-related tasks. Tasks consist of monitoring, tracking, communication, and resource management, as demonstrated in Figure 7.2. The three tasks imposed on pilots using the MATB were tracking, resource management, and communications. The tracking is located in the upper middle window and requires the test subject to keep the circular target in the center of the window using a joystick. This task is a compensatory task, thus increased reaction time resulting from a hypoxic state could affect a subject's ability to compensate or cause them to overcompensate. Resource management requires subjects to maintain fuel tanks at a level of 2500 units each, which can be achieved by transferring fuel from tank to tank. However, since hypoxia can cause impaired mental arithmetic and decision-making skills, maintaining appropriate fuel levels may be difficult under hypoxic conditions. The subject must also listen for audio messages addressed to their communications callsign, which is displayed at the top of the communications window, and ignore messages directed at other callsigns. The audio message directs the subjects to change the frequency of one of the radios listed on the screen, but because hypoxia can negatively impact the subject's ability to learn and memorize their callsign as well as their ability to pay attention to the audio signal, performance on this task may decrease with the onset of a hypoxic state.

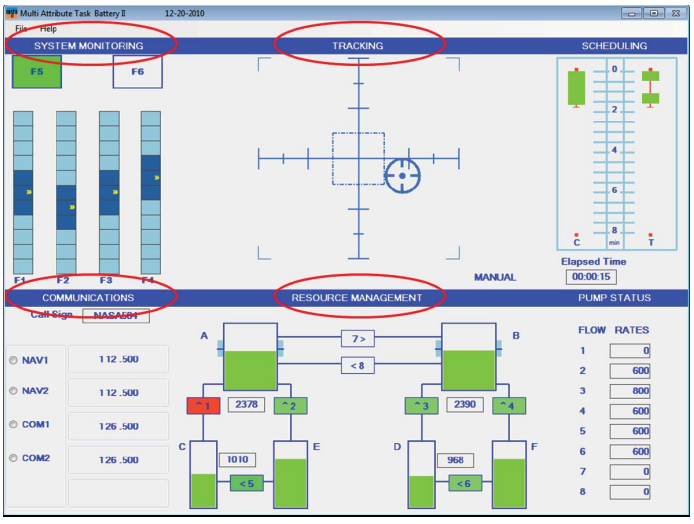


Figure 7.2: This illustration demonstrates the Multiple Attribute Task Battery

Table 7.1: Features Type per Sensing Modality

Modality	Feature Type	Number of Features
ECG	Summary Stats	2
	HR Complexity (HRC)	2
EEG	Power Spectrum	15 x 16 Channels
	Engagement Index	1 x 16 Channels
	Rank Order Complexity	1 x 16 Channels
Respiration	Rate	1
	Complexity	2
	Interactions (e.g. HRC)	7
0^2 Saturation	mean	1
	Interactions (e.g., Age)	7
Demographics	Anatomical (e.g., BMI)	2
	Flight Info.	2
Total		**261**

Physio Data and MATB Task Performance Error and

The MATB performance variables are updated and reported every 5 seconds and oxygen saturation is sampled at 256 Hz. Our design will therefore focus around a window size with a minimum of 5 seconds and a maximum of 15 seconds. Because this design is multimodal, not all physiological features can provide information within such a narrow window. Thus, not all the time series data can be placed on a a single time scaled window. It is simply not feasible. Ultimately, this will allow for us to analyze the data in a multi-scale approach where we specifically acquire features that provide longitudinal trend information on the autonomic system (longer time windows) and instantaneous features of the autonomic system (shorter time windows). This paints an enhanced picture of both the entire forest and of the individual trees in the forest.

7.2.2 Naive Adaptive Probabilistic Sensor (NAPS) Fusion

Naive Adaptive Probabilistic Sensor (NAPS) Fusion is a proposed framework that utilizes numerous, random, small (a sub-space of the feature space), bagged models that have augmented response variables that randomly span the full feature space of our dataset. The

purpose of NAPS Fusion is to:

1. Reduce model uncertainty caused by the data structure
   (a) Small sample sizes
   (b) Class imbalances
2. To Better Handle Subject-to-Subject Variability (i.e., Better Model Flexibility)
3. Detect Anomalies
4. Maintain modularity for future modalities and features
5. Handle ambiguous annotations of the ground truth data.

However, the crux of the NAPS Fusion design is to improve predictive performance through reducing uncertainty in models by employing five specific mechanisms in the following order: 1) smaller, random, but organized, feature sets (i.e. Sensors), 2) augmenting the response variable, 3) utilizing bagging algorithms, 4) quantifying model uncertainty, and 5) spanning the feature space through model fusion or combination. Figure **??** provides an overview of the NAPS Fusion Method, highlighting these five mechanisms which are discussed in more detail in the following subsections.

Development of "Sensors"

The development of "sensors" refers to the concept of sampling the feature space into organized subsets of the feature space. These "sensors" or feature sets are the foundation of the various model permutations formed (as discussed in the following Subsection 7.2.2). This means that we do not completely select any random vectors, p, within the vector space but it is randomly structured. For this study, each sensor (i.e., feature set), S_m, $(m = 1, \ldots, M)$ where M is the total number of sensors created from the full vector space, has 11 features. These 11 features for each sensor are randomly structured by choosing two random EEG

leads where, within the each lead, 4 EEG features are randomly taken. These produces 8 features total (2 EEG leads x 4 EEG features). The other three features are randomly taken from EKG, Respiration, O2 Saturation or Demographics modalities, producing a total of 11 features for each S_m. We produced a total of 175 of these sensors ($M = 175$), thus there is obvious overlap within the feature space between various sensors. The reduction in the amount of features now allows us to avoid model sparsity (issues of uncertainty) caused by too many features and not enough samples of data.

Augmented Response Variables and their Model Permutations

D.S. Theory's Framework for modeling ignorance and uncertainty inspired us to test the notation to augment the response variable. This approach is a way to help with model sparsity and thus aid in the training of the model by reducing the amount of response variables. From our insight into D.S theory (discussed in Subsection **??**), each individual class we want to predict, θ_i, is considered a singleton proposition (i.e., Delay in Error, Tracking Error). The DS Framework contains a full set of mutually exclusive and exhaustive propositions of interest which is referred to as the *frame of discernment (FOD).* Thus, for our specific case, the FOD Ω is defined with the four hypotheses,

$$\Omega = \{\theta_1, \theta_2, \theta_3, \theta_4\}, \tag{7.1}$$

based on our four-class ML detection problem. The FOD is a set of all subsets of Ω, which creates the power set 2^{Ω}. The power set consists of the combinatorial set that make up Ω. This combinatorial sets allows us to model ignorance related to the ML class label, where

the ML model can be trained around an augmented response variables consisting of

$$\begin{aligned} 2^{\Omega} =& \{\emptyset,\ \{\theta_1\}, \{\theta_2\}, \{\theta_3\}, \{\theta_4\}, \{\theta_1, \theta_2\}, \\ & \{\theta_3, \theta_4\}, \{\theta_2, \theta_3\}, \{\theta_1, \theta_4\}, \{\theta_1, \theta_3\} \\ & \{\theta_2, \theta_4\}, \{\theta_1, \theta_2, \theta_3\}, \{\theta_2, \theta_3, \theta_4\}, \\ & \{\theta_1, \theta_3, \theta_4\}, \{\theta_1, \theta_2, \theta_4\}, \Omega\}, \end{aligned} \tag{7.2}$$

permutations of the response variable. The power set in Equation 7.2, demonstrates there are 16 different propositions that can be created. These propositions can be augmented and mixed to produce different variable ML models, for which specific classification boundaries can now be relaxed or simply neglected. The permutations of the response variables that were chosen are presented below in Table 7.2. Depending on the permutations that are

Table 7.2: Permutations of the Response Variable (P=8)

Response Permutations	Class / Proposition Assignments	
1	$\{\theta_1\}$ $\{\theta_2\}$	$\{\theta_3\}$ $\{\theta_4\}$
2	$\{\theta_1\}$	$\{\theta_2, \theta_3, \theta_4, \}$
3	$\{\theta_2\}$	$\{\theta_1, \theta_3, \theta_4, \}$
4	$\{\theta_3\}$	$\{\theta_1, \theta_2, \theta_4, \}$
5	$\{\theta_4\}$	$\{\theta_2, \theta_3, \theta_4, \}$
6	$\{\theta_1, \theta_2\}$	$\{\theta_3, \theta_4\}$
7	$\{\theta_1, \theta_3\}$	$\{\theta_2, \theta_4\}$
8	$\{\theta_1, \theta_4\}$	$\{\theta_2, \theta_3\}$

chosen as your response variables ($\{\{\theta_1, \theta_3, \theta_4\}, \{\theta_2\}\}$), the class imbalances within the ML paradigm will improve or worsen. In order to combat these issues and reduce model bias as much as possible, SMOTE was dynamically implemented for each individualized selected sensor and augmented response. Therefore, this produces n feature sets (i.e. sensors) by P permutations of the response variables (i.e. augment reponses), producing nxP individual models, M_{pn}. The variety of these models can therefore be expressed as a matrix of randomly

structured models defined as

$$\begin{bmatrix} M_{11} & M_{12} & \dots & M_{1n} \\ M_{21} & M_{22} & \dots & M_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ M_{d1} & M_{d2} & \dots & M_{np} \end{bmatrix}. \tag{7.3}$$

Sensor ML Bagging

A *basic probability assignment (BPA),* otherwise referred to as a *basic belief assignment* is a function $m : 2^{\Omega} \rightarrow [0, 1]$, where 2^{Ω} is the power set, such that

$$m(\emptyset) = 0; \quad \sum_{A_i \subseteq 2^{\Omega}} m(A_i) = 1, \tag{7.4}$$

which requires probabilistic "supports" for its propositions. We aim to obtain these probabilistic "supports" from the outputs of each ML model's, M_{pn}, propositions. However, not all models can be generalized to a probability easily and thus we would be constrained to specific types of models that may not generalize well with specific data sets. This issue can be avoided by utilizing a bagging approach, also known as bootstrap aggregation. A bagging approach not only improves the stability and accuracy of ML model (reducing high variance/uncertainty) for the statistical classification, but can be thought of as a probability structure. The typical bagging approach sub-samples the observations of M_{pn}'s vector space, T times, using sampling with replacement (typically around 60% of the data). Therefore for each M_{pn}, we essentially create T "micro-models", since M_{pn} is already a sub-space of the original dataset. The T models are trained and produce a frequency for each time a class was chosen as the predicted class and could be thought of as a simplistic probability structure.

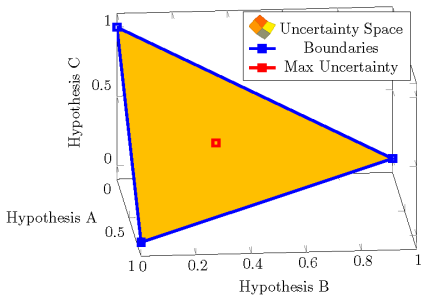


Figure 7.3: A visualization of the Uncertainty Space

Qualification of model uncertainty

In order to fully develop a BPA from a BOE (i.e. M_{pn}), an uncertainty assignment is required to utilize the DS framework approach. Although the proposed NAPS Fusion framework is not strictly subject to using DS Theory, we can utilize the uncertainty assignment as a means for determining which models to utilize. The uncertainty assignment is obtained by accounting for two factors: 1) uncertainty on the contextual meaning of the decision 2) uncertainty of model bias

Uncertainty on the Contextual Meaning of the Decision: On a data set that has absolutely no class imbalances, we can obtain contextual meaning around the frequency of each class. This provides informative context for predicting a specific class based on which T models were reported as the predicted class. For example, if you have four classes and 100 models, and we received 25 votes for each each class, we become very uncertain about the appropriate class to pick. Thus the model for this specific instance is unreliable. Ensembling this model,M_{pn}, with another bagged model, $M_{p(n+1)}$, that handles a different set of sensors may have very little conflict, thus, fusing these two ensemble methods together can be thought of as a way to easily perform a pseudo method of model averaging which weights each model differently according to the uncertainty assignment. Standard ensemble methods do not contextualize the model within the voting and strictly count the amount of votes [43]. Thus, we aim to model this uncertainty through the normalized bagging voting space, where T is the total number of votes (bags in the model) and V_c is the number of votes received for the c^{th} class

where C is the total number of classes or hypotheses. Therefore, we can define a bounded vector space defined by the number of propositions (hypotheses) and the amount of votes each hypothesis received by

$$\Theta_1 = 1 - \left(\frac{\sqrt{(\frac{V_1}{T} - \frac{1}{C})^2 + \cdots + (\frac{V_C}{T} - \frac{1}{C})^2}}{\sqrt{\frac{C-1}{C}}}\right). \tag{7.5}$$

Equation 7.5 is visually depicted in Figure 7.5, where, if all the proportions of the votes are approximately equal, the uncertainty assignment operates near the maximum uncertainty region in red. Conversely, if the votes are highly disproportional and there is a dominate proposition, the uncertainty operates near the edges of the vector space where the minimum uncertainty resides.

Uncertainty of the Biased Model: Model uncertainty is also altered when class imbalances are imposed on the model by the data. When classes are imbalanced, the model will tend to favor the majority class more (the prior). Therefore, similar to the previous uncertainty assignment, this adds additional doubt as to the model's ability to perform. Although SMOTE was applied to address the class imbalance, the approach is not always able to fully rectify the class imbalance, and the samples are still synthetic instances. Therefore, we use the original data sets for our uncertainty assignment to address the class imbalance issues. The imbalance of the class, θ_i, depends on the proportion of the amount of instances that occur for the class, I_c, over the total instances in the data set, S. Thus, quantifying the uncertainty associated to an imbalance for a multi-class problem is defined as,

$$\Theta_2 = \frac{\sqrt{(\frac{I_1}{S} - \frac{1}{C})^2 + \cdots + (\frac{I_C}{T} - \frac{1}{C})^2}}{\sqrt{\frac{C-1}{C}}}. \tag{7.6}$$

Total Uncertainty The total uncertainty, Θ, for the BOE is calculated by simply taking the

mean of the two methods that account for the uncertainty in the BOE through,

$$\Theta = \frac{\Theta_1}{2} + \frac{\Theta_2}{2} \tag{7.7}$$

Model Section

Due to the nature of D.S. theory's framework and cognitive state experiments, developing a model requires synergy for the model to work well with the psychophysiology data and for us to fuse models together. Since the sensors (subsets of features) were generated randomly, not all the sensors will capture the appropriate amount of variance for them to be a strong predictor. In addition, a top performing sensor may no longer be the sensor for a different response permutation (a different augment class). Similar to how the uncertainty adapts and varies across the samples, the sensors used for each permutation of the model should also adapt. In order to accomplish this, we took a simplistic approach where we utilized the 6 sensors with lowest uncertainty for each response permutation. Therefore, if we implement an all-vs-one method (Response Permutations 2-5), each response permutation has 6 sensors that are combined, producing 24 different models for the decision process.

Model Fusion

All the models that were selected are combined together to account for a large feature space, thus acting like a larger complex model.

An All-vs-one Approach An all-vs-one approach is a simplistic and contemporary approach that uses a binarization strategy, where we build a binary response model for each class, which is referred to as an one-vs-all ML framework [54]. The generated combinatorial responses for M_2n to M_5n are utilized $\{C_1, C_2, C_3\}, \{C_4\}$, $\{C_1, C_2, C_4\}, \{C_3\}$, or $\{C_1, C_3, C_4\}, \{C_2\}$, or $\{C_2, C_3, C_4\}, \{C_1\}$. Then, all the models are evaluated and the one with the highest probability is the predicted class. This addresses the problem with ignorance but the prob-

Table 7.3: RF ($A_c = 68.37\%$)

		Predicted			
		θ_1	θ_2	θ_3	θ_4
Actual	θ_1	**632**	24	6	1
	θ_2	93	**24**	1	0
	θ_3	124	5	**3**	0
	θ_4	43	2	0	**0**

lem of feature importance and dimensionality reduction still exists. For instance, when you have a four-class, classic singleton case we would generate $\{C_1\}$, $\{C_2\}$,$\{C_3\}$,$\{C_4\}$. However for a non-traditional one-vs-all ML framework design, we would generate combinatorial responses such as $\{\{C_1, C_2, C_3\}, \{C_4\}\}$, or $\{\{C_1, C_2, C_4\}, \{C_3\}\}$, or $\{\{C_1, C_3, C_4\}, \{C_2\}\}$, or $\{\{C_2, C_3, C_4\}, \{C_1\}\}$. We then evaluated which key features for the model will be analyzed and how models perform over these combinational patterns.

A D.S. Theory Approach Building upon this body of evidence can be formed at the sensor level, we aim to combine other sources of evidence to analyze performance using DCR.

Dempster's combination rule (DCR) allows one to combine or fuse evidence represented as DST models [135]:

$$m(A) = \frac{\sum_{B \cap C = A} m_1(B)\, m_2(C)}{1 - \sum_{B \cap C = \emptyset} m_1(B)\, m_2(C)}, \tag{7.8}$$

7.3 Evaluations and Discussion

In this section we address the following research questions regarding the use case of implementing Naive Adaptive Probabilistic Sensor Fusion for a more generalized ML framework to handle experimental study data for machine computer interaction:

RQ1 How do typical dimensionality reduction methods perform on our multi-class problems with small samples, class imbalances, and large features?

Table 7.4: RF & SMOTE ($A_c = 61.37\%$)

		Predicted			
		θ_1	θ_2	θ_3	θ_4
Actual	θ_1	**512**	71	68	12
	θ_2	59	**52**	5	2
	θ_3	95	13	**21**	3
	θ_4	32	6	4	**3**

RQ2 How does dimensionality reduction combined with augmented response perform?

RQ3 As we address these issues caused by uncertainty, how does performance for these ML paradigms change utilizing a NAPS framework (dimensionality reduction through the use of sensor fusion and augmented response variables)?

7.3.1 RQ1: Dimensionality Reduction for Multi-Class Problems

In this subsection, we implement an assortment of various dimensionality reduction methods and frameworks in order to evaluate if adequate lift in performance is evident to justify a particular framework. Since the NAPS framework utilized SMOTE to improve the class imbalances, SMOTE was also implemented to these higher dimensionality problems as well. Dimensionality reduction was examined by implementing various autoencoder frameworks (i.e. number of nodes and layers) and activation functions. Typical methods such as PCA, were deemed unnecessary to implement since a single layer (SL) linear activation function is strongly comparable to PCA.

Baseline Performance. Prior to implementing the dimensional reduction experiments, we obtained a baseline by simply examining the performance before we implemented any dimensionality reduction. Considering our concerns about how dimensionality reduction methods may actually hurt performance by capturing irrelevant information, we took two baselines: one experiment with strictly the raw data (SMOTE was not implemented), and the implementation of SMOTE with the Random Forest Algorithm (RF) with 150 bags. Table 7.3, is

Table 7.5: SL Autoencoders

Autoencoder Size	**Function**	**Accuracy**	**Function**	**Accuracy**
70	logSig	55.63%	SatLin	54.07%
60	logSig	55.94%	SatLin	53.03%
50	logSig	57.51%	SatLin	51.57%
40	logSig	55.21%	SatLin	54.28%
32	logSig	55.63%	SatLin	50.42%
24	logSig	56.47%	SatLin	50.51%
16	logSig	56.36%	SatLin	50.93%
8	logSig	52.29%	SatLin	48.84%
5	logSig	48.23%	SatLin	42.49%

the confusion matrix and accuracy ($A_c = 68.37\%$) for applying a RF algorithm to the raw data set. Utilizing RF and SMOTE together, Table 7.4, demonstrates a drop in performance accuracy. However if you examine the confusion matrix in Table 7.3, the precision and recall is very poor for the other classes (θ_2,θ_3,θ_4). Thus, we can note how SMOTE provides lift in performance for minority classes by adjusting the class imbalances.

Single Layer Autoencoders. To address this research question, we first started with a single layer autocoder framework where we ran experiments using two different activation functions and 9 different neural network (NN) node architectures. The random forest algorithm with 150 bags was implemented on these reduced feature sets from the autoencoders. Table 7.5, provides the averaged accuracy over a 5-fold cross-validation for each NN architecture and activation function. These two activation functions essentially allowed us to compare how linear and non-linear manifolds may increase accuracy. We can note that across all the NN architectures (autoencoder size/nodes), the non-linear activation function (logsig) provides an increase in performance downstream. In Table 7.6, we provide the confusion matrix for autocoder with 50 nodes that utilized the "logsig" activation function which was then stacked with a random forest providing an accuracy of 57.51% for the four class problem. However, when compared to Table 7.4, the most accurate autoencoder, which used the same RF paradigm, still performs the worst across all classes.

Deep Networks. The dimensionality reduction experiments were then extended for designing

Table 7.6: SL Autoencoder LogSig ($A_c = 57.51\%$)

		Predicted			
		θ_1	θ_2	θ_3	θ_4
Actual	θ_1	**465**	79	82	37
	θ_2	49	**52**	12	5
	θ_3	74	16	**33**	9
	θ_4	27	3	14	**1**

Table 7.7: Deep Network Architectures

Layer 1 Size	**Layer 2 Size**	**Function**	**Accuracy**
220	8	logSig	51.77%
180	16	logSig	53.13%
100	32	logSig	54.27%
85	45	logSig	52.82%
120	64	logSig	53.44%

two layer autocoders with a third soft layer used to train on the feature set. A simulation was done to test various permutations of the NN architecture, where permutation of the first layer consisted of 220, 180, 120, 100, or 85 nodes and the second layer consisted of permutations of 8, 16, 32, 45, or 64. This produced 25 different architectures. In Table 7.7, we provide the most accuracy architecture for each second layer permutation, where Table 7.8 is the confusion matrix for the highest performing architecture. As you can note, the current deep architectures are one of the lowest performing models.

RQ1 Takeaway. There was not a single dimensionality reduction method that provided better lift in performance when compared to the baseline approach. We hypothesize that because of subject-to-subject variability within the experimental data, this causes anomalous patterns in the feature space which can not be generalized or clustered as relevant information

Table 7.8: Deep Network ($A_c = 54.27\%$)

		Predicted			
		θ_1	θ_2	θ_3	θ_4
Actual	θ_1	**453**	72	107	31
	θ_2	58	**38**	16	6
	θ_3	73	22	**28**	9
	θ_4	29	8	7	**1**

Table 7.9: SL Autocoders & RF All-vs-One

Autoencoder Size	Function	Accuracy	Function	Accuracy
185	logSig	61.17%	SatLin	55.63%
128	logSig	62.84%	SatLin	55.94%
64	logSig	62.84%	SatLin	57.51%
32	logSig	60.55%	SatLin	55.21%
16	logSig	57.82%	SatLin	55.63%
8	logSig	54.17%	SatLin	56.47%

through an unsupervised fashion [94]. Thus, it is innate for these unsupervised dimensionality reduction methods to capture irrelevant information for their new projections of the feature space.

7.3.2 RQ2: Dimensionality Reduction Combined with Augmented Response Variables

In this subsection, we aim to address if the combination of a simplified response variable would allow us to better handle the sparsity of the data for increased performance. We addressed this issue by applying upsampling using SMOTE, augmenting the response variable, and reducing the dimensionality (feature space). The augmented response variable approach we utilized is the all-vs-one approach. Thus four-model binary models are designed, and we took the highest singleton probability as the predicted class. In Table 7.9, we once again examine the different NN encoder architectures and activation functions, where we achieved a tie for the best performance for a 128 and 64 node autoencoder using the nonlinear activation function. Note there is an approximate 5% increase in performance compared to the previous method, which did not augment the response variable. In addition, the nonlinear activation function out-performed the linear activation. In Table 7.10, we provide the confusion matrix for the single layer 64 nodes autoencoder using the "logsig" activation function.

However, based on our findings and hypothesis from *RQ1* that dimensionality reduction is capturing irrelevant information, it behooved us to also remove the autoencoder from the

ML design. Table 7.11, demonstrates an increase in performance accuracy when compared to the other all-vs-one methods, Table 7.10. In addition, there is also a slight performance increase from the original baseline model using SMOTE, Table 7.4.

RQ2 Takeaway. Once again, we demonstrated that unsupervised dimensionality reduction decreased the performance. However, when strictly comparing the approach of augmenting the response variable (i.e All-vs-One) to the original baseline in Table 7.4, we can note a slight increase in overall accuracy. Thus, by augmenting the response variable, we can aid in the uncertainty of the model by reducing the sparsity of the data set. However, the model performance is not optimal because of the large feature space the model is forced to cover.

Table 7.10: SL Encoder (64 Vars.) & RF All-vs-One ($A_c = 62.84\%$)

		Predicted			
		θ_1	θ_2	θ_3	θ_4
Actual	θ_1	**522**	54	65	22
	θ_2	57	**44**	15	2
	θ_3	85	11	**33**	3
	θ_4	30	4	8	**3**

Table 7.11: Raw Data Set & RF All-vs-One ($A_c = 64.91\%$)

		Predicted			
		θ_1	θ_2	θ_3	θ_4
Actual	θ_1	**543**	57	53	10
	θ_2	59	**54**	2	3
	θ_3	96	14	**21**	1
	θ_4	34	4	6	**1**

7.3.3 RQ3: NAPS Fusion Approach

Baseline. From *QR2*, we stated that we can improve the performance of the model if the feature space was reduced. We first examined a framework that uses six different sensors (6 different sub-spaces in the feature space) to build 6 different RF models that span the feature

Table 7.12: 4 Class Committee Voting Model ($A_c = 83.04\%$)

		Predicted			
		θ_1	θ_2	θ_3	θ_4
Actual	θ_1	**659**	1	0	0
	θ_2	74	**43**	1	0
	θ_3	49	6	**77**	0
	θ_4	31	3	9	**2**

space. The predicted classes of each model are reported and tallied for a committee vote to determine the predicted class. In Table 7.12, we can note a significant increase in performance when compared to the models in *QR1* and *QR2*. Simply through deductive reasoning, we can attribute this performance increase to the small feature set and combination of features sets (sensors). This claim can be supported by examining how high dimensionality and low samples negatively alters classification performance and the SMOTE Algorithm. In order to increase the dimensionality of a classification problem, the number of training samples must increase as well. Otherwise, the classifier performance will decrease. When your training samples are fixed and your dimensionality increases, the density of the training sample within the vector space will exponentially decrease. This lack of density in the feature space or sparsity imposes uncertainty on the model, since we become uncertain as to whether the classification boundary set is correct. In addition, this affects how we attempt to increase the minor training samples by using SMOTE. SMOTE works on a nearest neighbor algorithm, and if our vector space is too sparse, the nearest neighbors algorithms begin to break down since the distances between the adjacent neighbors are pushed to new hyperplanes as the dimensionality increases with a fixed sample size. Therefore, we are able to properly apply SMOTE to increase the sample size. However, utilizing a small sub-space within the feature space (a sensor) should enhance SMOTES capabilities.

A Committee All-vs-One Voting Approach. As previously discussed, the all-vs-one approach augments the response variable, allowing for a more generalizable model to fit the data. Possibly more importantly for this case of having high dimensionality, low samples and class

Table 7.13: Committee All-vs-One Voting Model ($A_c = 95.12$)

		Predicted			
		θ_1	θ_2	θ_3	θ_4
Actual	θ_1	**658**	1	1	0
	θ_2	0	**118**	0	0
	θ_3	0	6	**126**	0
	θ_4	23	3	12	**7**

imbalances, we can instantaneously create more data by merging the responses together and reducing the sparsity vector space which then allows us to better use SMOTE as well. In Table 7.13, we can note these significant increases in performance of the model as we handle this issue of model uncertainty. For the all-vs-one model approach for this application, we handle uncertainty in four ways: 1) by reducing the feature set into groups of sensors 2) by merging the response variable 3) utilizing SMOTE after these adjustment were made 4) the implementation of the sensor selection paradigm which aims to quantify model uncertainty. The sensor selection paradigm is critical since we randomly generated 175 models for each permutation of responses (a total of 8). Thus, we had 1400 models to choose from, but we only used 24 of them for this approach.

A DS Fusion Approach. Since our all-vs-one model, a maximum likelihood approach, also utilizes a committee voting method to combine the other 6 sensors, we explored other approaches to fuse the models together. In addition, although all-vs-one is revered as an approach to provide lift in performance, we are unable to handle ignorance when we have conflicting class labels. However, our proposed DS framework can handle this conflict and ambiguous labels. We first compared a D.S approach design to the same binarization strategies of response variables as the Committee All-vs-One Voting Approach above, shown in Table 7.14. There is a very minute increase in performance when compared to the Committee All-vs-One Voting (shown in Table 7.13), and it is essentially the same in performance. However, for the cases where ignorance was required to be modeled, an all-vs-one response variable was unable to be used. In Table 7.15, we demonstrate still a very high perfor-

mance accuracy of $A_c = 93.93\%$ when multiple class doubleton class labels are set to an ML paradigm. This is an extremely critical aspect of human-machine interaction research and class labels can not be agreed upon. Thus, if one subject matter expert labels the event as θ_1 and another subject matter expert labels the event as θ_2, we can set the class label to train as a doubleton set of $\{\theta_1, \theta_2\}$.

RQ3 Takeaway. The NAPS framework provides superior lift in performance compared to other approaches. This framework opens up the discussion of various methods that could be implemented to improve on the model selection and the fusion of models. When there is no conflict within the ground truth labels, we see no need to implement the Dempster-Shafer framework unless the exploration of a new fusion paradigm or an improved uncertainty assignment could demonstrate an improved detection performance. The computational complexity of the DS framework is much higher than the all-vs-one approach and can be quite overwhelming to implement. However, the need to implement a DS fusion framework is because we have these ambiguities between the class label.

Table 7.14: NAPS Fusion Confusion Matrix ($A_c = 95.29\%$)
Tripleton Set (Comparable to All-vs-One)

		Predicted			
		θ_1	θ_2	θ_3	θ_4
Actual	θ_1	**658**	1	1	0
	θ_2	0	**118**	0	0
	θ_3	0	5	**127**	0
	θ_4	23	3	12	**7**

Table 7.15: NAPS Fusion Confusion Matrix (($A_c = 93.93\%$)
Doubleton Set)

		Predicted			
		θ_1	θ_2	θ_3	θ_4
Actual	θ_1	**659**	1	1	0
	θ_2	1	**117**	0	0
	θ_3	5	10	**116**	1
	θ_4	23	10	7	**5**

7.4 Conclusion

Many cognitive performance experiments for human-machine interaction produce a small sample size, have large class imbalances, and have a high dimensionality feature space. Typically, we are faced with machine learning problems that have one or two of these problems simultaneously. However, when all three of these problems occur, the standard approaches for handing each problem individually essentially fails. For instance, class imbalances could be handled by down-sampling our data to adjust for imbalance in the classes, but we do not have enough data for down-sampling. Conversely, we can up-sample by using SMOTE, but as we discussed and demonstrated, the benefits of SMOTE degrade when we have a low amount of samples and high dimensionality. In addition, we can try to address the problem by applying dimensionality reduction methods, but they are unable to capture relevant information most likely due to the small sample size and class imbalances.

The NAPS Fusion framework addresses how we can overcome these experimental data issues and ML challenges. NAPS accomplishes this through appropriate model selection, uncertainty assignment, augmentation of response variables, and fusion of sensors that span the feature space that allow SMOTE not only to adjust for its original class imbalances but the imbalances created by the augmented response variables. The work presented utilizes Dempster Combintation Rule (DCR) as a means of fusion and is one of the most rudimentary fusion algorithms, which opens up research areas for exploring methods for more optimal model fusion paradigms, and model selection methods. This DS approach for combining models under NAPS opens the door for cognitive performance research to move forward into open world environments, where conflict of class labels occur frequently. Moreover, this framework provides an extremely modular design, where new models and hardware modalities can be interchanged easily without tuning and training from ground zero.

If you hear a voice within you say, "You cannot paint," then by all means paint, and that voice will be silenced"

Vincent Van Gogh

8

Conclusion

Part I Contributions. This work, "Handling Uncertainty for Signature Detection," demonstrates how simplistic ML signature detection applications can be improved by stacking an additional uncertainty framework upon the pre-existing matching paradigm. This is achieved by appropriately quantifying the uncertainty within the sets and utilizing D.S. Theory to handle conflicting information with the uncertainty assigned to aid in deciphering the winning match. Due to the additional computational complexity imposed by this work, we demonstrate that by utilizing Dempster Combination Rule (a fusion method that is commutative and associative) we can approximately linearly scale on a distributive computing

framework (i.e., MapReduce) to enhance the computational performance speed and handle larger-scale data sets. This work also has demonstrated the advantages of handling information that is corrupted which provides weak support for its proposition (i.e. hypothesis) but can be enhanced through fusion methods. This work was then applied to a two lead electrocardiogram (ECG) data set, where we showed a novel implementation method for the uncertainty framework. Moreover, we demonstrated the importance of how both ECG leads do not require strong agreement in order to converge to a consensus for detecting the appropriate signature. Therefore, if one lead of the ECG signal is corrupted and the other is mildly corrupted, we can overcome these detection limitations and still see adequate performance. The most powerful aspect of this method, and most heavily weighted utility for this work, is how this approach requires little to no prior data. The only data requirement is a single template for each source. *Part I Future work.* This work has opened up avenues in the area of DS theory for time series and image processing applications. We are currently using these matching paradigms for Interferometric Synthetic Aperture Radar (InSAR) for tunnel detection by developing various templates that indicate signatures concavity in the ground. This paradigm is ideal since it takes 6 months to one full year to obtain a single template and is highly costly to collect imaging of changes in the ground using InSARs. The current theoretical framework has demonstrated enhanced detection performance by assuming each proposition is strictly independent from each of the other propositions (i.e., singleton). An extension to this theoretical framework would be to address proposition matches that share similarities within their template signatures (e.g., doubletons, tripletons). This allows us to model ignorance within the set of propositions.

Part II Contributions. "Handling uncertainty for predictive problems for systems with diverse information sources" stresses the importance of how downstream work-flow should be considered when developing a comprehensive predictive system. Through the use of Monte Carlo Simulations and signal corruption, we demonstrate how simple noise within the input features can statistically alter integral results that would lead downstream into the model

development. In addition, we demonstrate the importance of how new feature engineering can produce additional dimensionality (i.e. new features for class separation) to potentially aid in the predictive model. In terms of model development, we demonstrate that these design problems due to uncertainty in open world settings are challenging due to a lack of training data instances, noise, class imbalances, wide data sets, numerous modalities, subject-to-subject variability, and the inability to modularly scale to new bio-sensors from previous ML paradigms. However, we show that utilizing and developing small individual models (a small subset of the feature space) which are then fused with other small individual models (enabling us to expand our feature space) provides a significant increase in classification performance. This also enables a highly modular design as the system is updated with additional sensor modalities over time. *Part II Future work.* The new developed electroencephalogram (EEG) features have demonstrated great value in the clinical and human performance domain. Extensions of the work would be to evaluate the newly developed method, rank order complexity (ROC), to prior methods for human task engagement such as the engagement index. In addition, we hypothesize that the other newly formed EEG Method, activation complexity (AC), is linked to neuronal isolation. The ability to capture these effects and quantify them would be invaluable in the clinical domains in the areas of hypoxia, epilepsy, and stroke research. In terms of the model development, we demonstrated that small individual models which utilize a committee paradigm are ideal for human performance research. However, the main finding that will lead to future work is the ability to handle data sets that have conflicting or ambiguously annotated ground truth data. This problem occurs consistently for cognitive state human performance research, where the subject matter experts cannot agree on a specific class label (type of cognitive state). The developed model uncertainty methodology, naive probabilistic sensor fusion (NAPS), allows us to model ignorance (the ambiguity of the class label) with the ML paradigm but still refine our support to a single class proposition.

Bibliography

[1] Apache Hadoop. `https://hadoop.apache.org`.

[2] *Prediction of Cognitive States during Flight Simulation using Multimodal Psychophysiological Sensing* (2017), AIAA Information Systems-AIAA Infotech.

[3] Abasolo, D., Hornero, R., Espino, P., Alvarez, D., and Poza, J. Entropy analysis of the eeg background activity in alzheimer's disease patients. *Physiological Measurement 27*, 3 (2005).

[4] Abbound, S., and Sadeh, D. The use of cross-correlation function for the alignment of ECG waveforms and rejection of extrasystoles. *Computer and Biomedical Research 17* (1984), 258–266.

[5] Abdullah, H., and Cvetkovic, D. Electrophysiological signals segmentation for eeg frequency bands and heart rate variability analysis. *The 15th International Conference on Biomedical Engineering 43* (2014).

[6] Acharya, U. R., Fujita, H., Sudarshan, V. K., Bhat, S., and Koh, J. E. Application of entropies for automated diagnosis of epilepsy using {EEG} signals: A review. *Knowledge-Based Systems 88* (2015), 85 – 96.

[7] Addair, T., Dodge, D., Walter, W., and Ruppert, S. Large-scale seismic signal analysis with hadoop. *Computers and Geosciences 66* (2014), 145 – 154.

[8] Afonso, V. X. *Biomedical digital signal processing: C-language Examples and Laboratory Experiments for the IBM PC.* Prentice-Hall, Upper Saddle River, NJ, USA, 1993, pp. 236–264.

[9] Aljarrah, I. A., Ghorab, A. S., and Khater, I. M. Object recognition system using template matching based on signature and principal component analysis. *International Journal of Digital Information and Wireless Communication 2*, 2 (2012), 156–163.

[10] Angela R. Harrivel, C. L. S., Milletich, R. J., Heinich, C. M., Last, M. C., Napoli, N. J., Abraham, N., Prinzel, L. J., Motter, M. A., and Pope, A. T. Prediction of cognitive states during flight simulation using multimodal psychophysiological sensing. 1–8.

[11] Backs, R. W. Going beyond heart rate: Autonomic space and cardiovascular assessment of mental workload. *The International Journal of Aviation Psychology 5*, 1 (1995), 25–48.

[12] Bailenson, J. N., Pontikakis, E. D., Mauss, I. B., Gross, J. J., Jabon, M. E., Hutcherson, C. A., Nass, C., and John, O. Real-time classification of evoked emotions using facial feature tracking and physiological responses. *International Journal of Human-Computer Studies 66*, 5 (2008), 303 – 317.

[13] Bandt, C., and Pompe, B. Permutation Entropy: A Natural Complexity Measure for Time Series. *Phys. Review Letters 88*, 17 (Apr 2002).

[14] Bauer, G., Trinka, E., and Kaplan, P. Eeg patterns in hypoxic encephalopathies (post-cardiac arrest syndrome): fluctuations, transitions, and reactions. *J Clin Neurophysiol 30*, 5 (2013), 477–89.

[15] Benavoli, A., Chisci, L., Farina, A., and Ristic, B. Modelling uncertain implication rules in evidence theory. In *Proc. International Conference on Information Fusion (FUSION)* (Cologne, Germany, June/July 2008), pp. 1–7.

[16] Berger, H. Nervenkr. *Arch. Psychiatry 87*, 527 (1929).

[17] Berntson, G., and Stowell, J. Ecg artifacts and heart period variability: don't miss a beat! *Psychophysiology 35*, 1 (1998), 127–32.

[18] Bhalla, D., Bansal, R. K., and Gupta, H. O. Integrating ai based dga fault diagnosis using dempster–shafer theory. *Inter. Journal of Electrical Power and Energy Systems 48* (2013), 31 – 38.

[19] Black, P. K. Is shafer general bayes. *Int. J. Approx. Reasoning* (1988).

[20] Blackman, S., and Popoli, R. *Design and Analyis of Modern Tracking Systems*. Artech House, Norwood, MA, 1999.

[21] Blazquez, J., Garcia-Berrocal, A., Montalvo, C., and Balbas, M. The coverage factor in a flatten-gaussian distribution. *Metrologia 45*, 5 (2008).

[22] Bobrov, D., Kitov, I., and Zerbo, L. Perspectives of cross-correlation in seismic monitoring at the international data centre. *Pure Applied Geophys 171* (2014), 439???468.

[23] Borg, F. Filter banks and the "intensity analysis" of emg. *http://arxiv.org/abs/1005.0696* (2003).

[24] Briechle, K., and Hanebeck, U. D. Template matching using fast normalized cross correlation. In *Aerospace/Defense Sensing, Simulation, and Controls* (2001), International Society for Optics and Photonics, pp. 95–102.

[25] Bullmore, E., M, B., Harvey, I., Murray, R., and Ron, M. Cerebral hemispheric asymmetry revisited: effects of handedness, gender and schizophrenia measured by radius of gyration in magnetic resonance images. *Psychol Med 25*, 2 (1995), 349–63.

[26] Buzsaki, G., and Draguhn, A. Neuronal oscillations in cortical networks. *Science 304*, 527 (2004).

[27] Cacioppo, J., Tassinary, L., and Berntson, G. *Handbook of Pyschophysiology*, 3 ed. Cambridge University Press, Cambridge, UK, 2007.

[28] Cain, B. A Review of the Mental Workload Literature. Tech. rep., Defence Research and Development Canada Toronto, Toronto, Ontario, 2007.

[29] Capell, A.-S., Colot, O., and Fernandez-Maloigne, C. 3d segmentation of mr brain images into white matter, gray matter and cerebro-spinal fluid by means of evidence theory. In *Artifical Intelligence in Medicine 9th Conference of Article Intelligence in Medicine in Europe* (2003), vol. 9, AIME, p. 5.

[30] Chan, R. H., Ho, C.-W., and Nikolova, M. Salt-and-pepper noise removal by median-type noise detectors and detail-preserving regularization. *Image Processing, IEEE Transactions on 14*, 10 (2005), 1479–1485.

[31] Charlton, S. G. Measurement of cognitive states in test and evaluation, 2001.

[32] Chen, Y., Qian, J., and Saligrama, V. A new one-class svm for anomaly detection. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (May 2013), pp. 3567–3571.

[33] Chew, C., and Zahedi, E. Limb cardiovasulature system identification using adaptive filtering. *International Federation for Medical and Biological Proceedings* (2006), 406–409.

[34] Chin, F. J., Fang, Q., Zhang, T., and Cosic, I. A fast critical arrhythmic ECG waveform identification method using cross-correlation and multiple template matching. *International Conference of the IEEE EMBS 32* (August 2010), 1922–1925.

[35] Cinaz, B., La Marca, R., Arnrich, B., and Tröster, G. Monitoring of mental workload levels. *Proceedings of the IADIS International Conference e-Health 2010, EH, Part of the IADIS Multi Conference on Computer Science and Information Systems 2010, MCCSIS 2010* (2010), 189–193.

[36] Ciresan, D. C., Meier, U., and Schmidhuber, J. Multi-column deep neural networks for image classification. *CoRR abs/1202.2745* (2012).

[37] Clifford, G., Long, W., Moody, G., and Szolovits, P. Robust parameter extraction for decision support using multimodal intensive care data. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 367*, 1887 (2009), 411–429.

[38] Clifford, G. D., Azuaje, F., and McSharry, P. Ecg statistics, noise, artifacts, and missing data. In *Advanced Methods And Tools for ECG Data Analysis*. Artech House, Inc., Norwood, MA, USA, 2006, ch. 3, pp. 70–71.

[39] Costa, M., Goldberger, A. L., and Peng, C.-K. Multiscale entropy analysis of complex physiologic time series. *Phys. Rev. Lett. 89* (Jul 2002).

[40] Costa, M., Peng, C.-K., and Goldberger, A. Multiscale analysis of heart rate dynamics: Entropy and time irreversibility measures. *Journal of Cardio Eng 8*, 2 (2008), 88 – 93.

[41] Courtiol, J., Perdikis, D., Petkoski, S., Müller, V., Huys, R., Sleimen-Malkoun, R., and Jirsa, V. K. The multiscale entropy: Guidelines for use and interpretation in brain signal analysis. *Journal of Neuroscience Methods 273* (2016), 175 – 190.

[42] Cowley, B., Filetti, M., Lukander, K., Torniainen, J., Henelius, A., Ahonen, L., Barral, O., Kosunen, I., Valtonen, T., Huotilainen, M., Ravaja, N., and Jacucci, G. vol. 9. 2016.

[43] Davidson, I., and Fan, W. *When Efficient Model Averaging Out-Performs Boosting and Bagging*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006, pp. 478–486.

[44] Dawes, B., Abrahams, D., and Rivera, R. Boost C++ libraries. `http://www.boost.org`.

[45] Dean, J., and Ghemawat, S. MapReduce: simplified data processing on large clusters. *Communications of the ACM 51*, 1 (2008), 107–113.

[46] Deng, L., and Yu, D. Deep convex net: A scalable architecture for speech pattern classification. In *Proceedings of Interspeech* (2011).

[47] Dobbs, S., Schmitt, N., and Ozemek, H. QRS detection by template matching using real-time correlation on a microcomputer. *Journal of Clinical Engineering 9* (1984), 197–212.

[48] Ellis, R. J., Zhu, B., Koenig, J., Thayer, J. F., and Wang, Y. A careful look at ecg sampling frequency and r-peak interpolation on short-term measures of heart rate variability. *Physiological Measurement 36*, 9 (2015), 1827.

[49] Fioretti, G. A mathematical theory of evidence for g.l.s. shackle. *Mind and Society 2* (2001), 77–97.

[50] Frolich, L., and et al. Incremental value of biomarker combinations to predict progression of mild cognitive impairment to alzheimer's dementia. *Alzheimer's Research and Therapy 9*, 1 (Oct 2017), 84.

[51] Fusheng, Y., Bo, H., and Qingyu, T. *Approximate Entropy and Its Application to Biosignal Analysis*. John Wiley and Sons, Inc., 2012, pp. 72–91.

[52] Gabor, D. Theory of communication. *IEEE 93* (1946), 429–459.

[53] Gabriel, D. A., and Kamen, G. Point:counterpoint comments. *J Applied Physiology 105* (2008).

[54] Galar, M., Fernández, A., Barrenechea, E., Bustince, H., and Herrera, F. An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes. *Pattern Recognition 44*, 8 (2011), 1761 – 1776.

[55] Ganeshapillai, G., and Guttag, J. Real time reconstruction of quasiperiodic multi parameter physiological signals. *EURASIP Journal on Advances in Signal Processing 173*, 1 (2012), 1–15.

[56] Gelman, A. The boxer, the wreslter and the coin flip: A paradox of robust bayesian inference and belief functions. *The American Statistician 60*, 2 (May 2006), 146–150.

[57] Goldberger, A., Amaral, L., Glass, Leon Hausdorff, J., Ivanov, P., Mark, R., Mietus, J., Moody, G., Peng, C., and Stanley, H. Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals. *Circulation 101* (2000), 230–236.

[58] Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C.-K., and Stanley, H. E. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation 101*, 23 (June 2000), e215.

[59] Goldberger, J., and Ng, J. *Practical Signal and Image Processing in Clinical Cardiology*. Springer, 2010.

[60] Google. MapReduce4C. `https://github.com/google/mr4c`.

[61] Gradwell, D. Hypoxia and hyperventilation. In *Ernstings's Aviation Medicine*, D. Gradwell and D. J. Rainford, Eds., 4nd ed. CRC Press, 2006, ch. 3, p. 41–56.

[62] Gustafsson, F. Determining the initial states in forward-backward filtering. *IEEE Transactions on Signal Processing 44*, 4 (Apr 1996), 988–992.

[63] Hall, D. L., and Llinas, J., Eds. *Handbook of Multisensor Data Fusion*. CRC Press, Boca Raton, FL, 2001.

[64] Ham, F. M., and Kostanic, I. *Principles of Neurocomputing for Science and Engineering*. McGraw-Hill, 2001.

[65] Haralick, R. M., and Shapiro, L. G. *Computer and Robot Vision*, vol. 2. Addison-Wesley, 1992.

[66] Harrison, J., Izzetoglu, K., Ayaz, H., Willems, B., Hah, S., Woo, H., Shewokis, P. A., Bunce, S. C., and Onaral, B. Human Performance Assessment Study in Aviation Using Functional Near Infrared Spectroscopy BT - Foundations of Augmented Cognition: 7th International Conference, AC 2013, Held as Part of HCI International 2013, Las Vegas, NV, USA, July 21-26, 2013. Procee. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 433–442.

[67] Hart, S. G., and Staveland, L. E. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Human Mental Workload*, P. A. Hancock and N. Meshkati, Eds. North Holland Press, Amsterdam, 1988.

[68] Heisenberg, W. *The Physical Principles of the Quantum Theory*. Courier Corporation, Chicago, IL, 1949.

[69] Henderson, J. M., Shinkareva, S. V., Wang, J., Luke, S. G., and Olejarczyk, J. Predicting cognitive state from eye movements. *PLOS ONE 8* (05 2013), 1–6.

[70] Ignaccolo, M., Latka, Mirek Jernajczyk, W., Grigolini, P., and West, B. J. The dynamics of eeg entropy. *J Biology Physics 36* (2010), 185–196.

[71] Itao, K., Umeda, T., Lopez, G., and Kinjo, M. Human recorder system development for sensing the autonomic nervous system. In *2008 IEEE Sensors* (Oct 2008), pp. 423–426.

[72] Javorka, M., Zila, I., Balharek, T., and Javorka, K. Heart rate recovery after exercise: relations to heart rate variability and complexity. *Brazilian Journal of Medical and Biological Research 35*, 8 (2002), 991–1000.

[73] Kanda, P. A., Oliveira, E. F., and Fraga, F. J. {EEG} epochs with less alpha rhythm improve discrimination of mild alzheimer's. *Computer Methods and Programs in Biomedicine 138* (2017), 13 – 22.

[74] Kasdin, N. J. Discrete simulation of colored noise and stochastic processes and 1/f alpha; power law noise generation. *Proceedings of the IEEE 83*, 5 (May 1995), 802–827.

[75] Kay, R. U. Fundamentals of the Dempster-Shafer theory and its applications to system safety and reliability modelling. *Reliab. Theory Appl 3* (2007), 173–185.

[76] Khandoker, A. H., Jelinek, H. F., and Palaniswami, M. Identifying diabetic patients with cardiac autonomic neuropathy by heart rate complexity analysis. *Biomedical engineering online 8* (2009), 3.

[77] Kim, B.-H., Noh, Y.-H., and Jeong, D.-U. Implementation of template matching based ecg compression algorithm for mobile application. In *IT Convergence and Security (ICITCS), 2013 International Conference on* (Dec 2013), pp. 1–4.

[78] Kim, K., Baek, H., Lim, Y., and Park, K. Effect of missing rr-interval data on nonlinear heart rate variability analysis. *Comp. Methods and Programs in Biomed. 106*, 3 (2012), 210 – 218.

[79] Kim, K., Kim, J., Lim, Y., and Park, K. The effect of missing rr-interval data on heart rate variability analysis in the frequency domain. *Physiological Measurement. 30*, 10 (2009), 210 – 218.

[80] Kim, K. K., Lim, Y. G., Kim, J. S., and Park, K. S. Effect of missing rr-interval data on heart rate variability analysis in the time domain. *Physiological Measurement 28*, 12 (2007), 1485.

[81] Kothalkar, S. S., and Manjusha, D. Comparative study of image registration methods. *International Journal of Image Processing (IJIP) 8*, 3 (2014), 125–147.

[82] Krasteva, V., and Jekova, I. QRS template matching for recognition of ventricular ectopic beats. *Annals of Biomedical Engineering 35* (December 2007), 2065–2076.

[83] Kumar, A., Joshi, A., Kumar, A., Mittal, A., and Gangodkar, D. Template matching application in geo-referencing of remote sensing temporal image. *International Journal of Signal Processing, Image Processing and Pattern Recognition 7*, 2 (2014), 201–210.

[84] Laborde, S., Mosley, E., and Thayer, J. F. Heart rate variability and cardiac vagal tone in psychophysiological research – recommendations for experiment planning, data analysis, and data reporting. *Frontiers in Psychology 8* (2017), 213.

[85] Lake, D., Richman, J., Griffin, M., and Moorman, R. Sample entropy analysis of neonatal heart rate variability. *American Journal of Physiology 283*, 3 (2002), R789–R797.

[86] Last, T., Nugent, C. D., and Owens, F. J. Multi-component based cross correlation beat detection in electrocardiogram analysis. *Biomedical Engineering Online 3*, 26 (July 2004), 1–14.

[87] Leavline, E., and Singh, D. A. G. Salt and pepper noise detection and removal in gray scale images: An experimental analysis. *International Journal of Signal Processing, Image Processing and Pattern Recognition 6*, 5 (2013), 343–352.

[88] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE 86*, 11 (1998), 2278–2324.

[89] LeCun, Y., Cortes, C., and Burges, C. The MNIST Database of Handwritten Digits. `http://yann.lecun.com/exdb/mnist/`.

[90] Lee, S., Jo, S., and Kim, J. MRDataCube: Data cube computation using MapReduce. In *International Conference on Big Data and Smart Computing* (2015), pp. 95–102.

[91] Lefevre, E., Colot, O., and Vannoorenberghe, P. Belief function combination and conflict management. *Information Fusion* (2002), 149–162.

[92] Lefevre, E., Vannorenberghe, P., and Colot, O. Using information criteria in Dempster-Shafer basic belief assignment. *IEEE International Fuzzy Systems Conference Proceedings*, I (August 1999), 173–178.

[93] Leistedt, S. J.-J., Linkowski, P., Lanquart, J., Mietus, J. E., Davis, R., Goldberger, A. L., and Costa, M. Decreased neuroautonomic complexity in men during an acute major depressive episode: analysis of heart rate dynamics. *Translational Psychiatry 8* (July 2011), 1–7.

[94] Lenz, M., Muller, F.-J., Zenke, M., and Schuppert, A. Pca and the reported low intrinsic dimensionality of gene expression microarray data. *Nature Scientific Reports* (2016), 1–11.

[95] Leon-Garcia, A. *Probability, Statistics, and Random Processes for Electrical Engineering*, 3rd ed. Person Prentice Hall, 2008.

[96] Li, Q., Mark, R. G., and Clifford, G. D. Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter. *Physiology Measurement 29*, 1 (2008), 15–32.

[97] Limaye, H., and Deshmukh, V. Ecg noise sources and various noise removal techniques: A survey. *International Journal of Application or Innovation in Engineering and Management 5* (2016), 406–409.

[98] Lipton, P. Ischemic cell death in brain neurons. *Physiological Review 79*, 4 (1999), 1432–1568.

[99] Liu, N. T., Batchinsky, A. I., Cancio, L. C., and Salinas, J. The impact of noise on the reliability of heart-rate variability and complexity analysis in trauma patient. *Comp. in Biology and Medicine 43*, 11 (2013), 1955 – 1964.

[100] Liu, N. T., Cancio, L. C., Salinas, J., and Batchinsky, A. I. Reliable real-time calculation of heart-rate complexity in critically ill patients using multiple noisy waveform sources. *J Clin Monit Comput 28*, 2 (Apr 2014), 123–131.

[101] Liu, Y.-S., Duh, D.-J., Chen, S.-Y., Liu, R.-S., and Hsieh, J.-W. Scale and skew-invariant road sign recognition. *International Journal of Imaging Systems and Technology 17* (2007), 28–39.

[102] Mahbuba, U., Imtiaza, H., Roya, T., Rahmana, S., and Ahadb, A. R. A template matching approach of one-shot-learning gesture recognition. *Pattern Recognition Letters 34*, 15 (November 2013), 1780–1788.

[103] Mandryk, R. L., and Atkins, M. S. A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. *International Journal of Human-Computer Studies 65*, 4 (2007), 329 – 347.

[104] Martin, R., Lloyd, H., and Cowan, A. The early events of oxygen and glucose deprivation: setting the scene for neuronal death? *Trends Neuroscience 17*, 6 (1994), 251–7.

[105] Mercado, J. E., Reinerman-Jones, L., Barber, D., and Leis, R. Investigating Workload Measures in the Nuclear Domain. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (2014), vol. 58, pp. 205–209.

[106] Moon, S., Lee, J.-G., and Kang, M. Scalable community detection from networks by computing edge betweenness on MapReduce. In *International Conference on Big Data and Smart Computing* (Jan 2014), pp. 145–148.

[107] Moore, D. S. *The Basic Practice of Statistics*, 3rd ed. W.H Freeman and Company, 2004.

[108] Napoli, N., and Barnes, L. A Dempster-Shafer Approach for Corrupted Electrocardiograms Signals. *Twenty-Ninth International Florida Artificial Intelligence Research Society Conference* (2016).

[109] Napoli, N. J. The detection of analytes using spectroscopy: A Dempster-Shafer approach. Thesis, University of Miami, Coral Gables, FL, August 2014.

[110] Napoli, N. J., Barnes, L., and Premaratne, K. Correlation coefficients based template matching: Accounting for uncertainty in selecting the winner. In *International Conference on Information Fusion* (July 2015), pp. 1–8.

[111] Napoli, N. J., Barnes, L. E., and Premaratne, K. Correlation coefficient based template matching: Accounting for uncertainty in selecting the winner. In *FUSION* (2015), IEEE, pp. 311–318.

[112] Napoli, N. J., Leach, K., Barnes, L. E., and Weimer, W. A mapreduce framework to improve template matching uncertainty. In *Big Data and Smart Computing* (2016).

[113] Napoli, N. J., Mixco, A. R., Bohorquez, J. E., and Signorile, J. F. An emg comparative analysis of quadriceps during isoinertial strength training using nonlinear scaled wavelets. *Human movement science 40* (April 2015), 134—153.

[114] Novak, D., Mihelj, M., and Munih, M. A survey of methods for data fusion and system adaptation using autonomic nervous system responses in physiological computing. *Interacting with Comp. 24* (2012).

[115] Nugent, C., Webb, J., Black, N., and Wright, G. Electrocardiogram 1: Classification. *Automedica 17* (1999), 281–306.

[116] Pan, J., and Tompkins, W. J. A real-time qrs detection algorithm. *IEEE Transactions on Biomedical Engineering 32*, 3 (1985), 230–236.

[117] Pan, J., and Tompkins, W. J. A Real-Time QRS Detection Algorithm. *IEEE Trans. Biomed. Engin. BME-32*, 3 (1985), 230–236.

[118] Papadelis, C., Kourtidou-Papadeli, C., Bamidis, P. D., Maglaveras, N., and Pappas, K. The effect of hypobaric hypoxia on multichannel eeg signal complexity. *Clinical neurophysiology official journal of the International Federation of Clinical Neurophysiology 118*, 1 (January 2007), 31—52.

[119] Pincus, S. M. Approximate entropy as a measure of system complexity. *PNAS 88*, 6 (1991), 2297–2301.

[120] Pizzagalli, D. A. Electroencephalography and high- density electrophysiological source localization. In *Handbook of psychophysiology*, J. Cacioppo, L. Tassinary, and G. Berntson, Eds. Cambridge University Press, Cambridge, UK, 2007.

[121] Plews, D. J., Laursen, P. B., Stanley, J., Kilding, A. E., and Buchheit, M. Training adaptation and heart rate variability in elite endurance athletes: Opening the door to effective monitoring. *Sports Medicine 43*, 9 (2013), 773–781.

[122] Pope, A. T., Bogart, E. H., and Bartolome, D. S. Biocybernetic system evaluates indices of operator engagement in automated task. *Biological Psychology 40*, 1 (1995), 187 – 195.

[123] Prinzel, L. J., Freeman, F. G., Scerbo, M. W., Mikulka, P. J., and Pope, A. T. A closed-loop system for examining psychophysiological measures for adaptive task allocation. *The International Journal of Aviation Psychology 10*, 4 (2000), 393–410.

[124] Ranger, C., Raghuraman, R., Penmetsa, A., Bradski, G., and Kozyrakis, C. Evaluating mapreduce for multi-core and multiprocessor systems. In *High Performance Computer Architecture, 2007. HPCA 2007. IEEE 13th International Symposium on* (2007), Ieee, pp. 13–24.

[125] Richman, J. S., and Moorman, J. R. Physiological time-series analysis using approximate entropy and sample entropy. *Amer. Jour. of Physio. - Heart and Circ. Physio. 278*, 6 (2000), 2039–2049.

[126] Riedl, M., Müller, A., and Wessel, N. Practical considerations of permutation entropy. *The Euro. Phys. Jour. Special Topics 222*, 2 (2013), 249–262.

[127] Rosso, O. A., Blanco, S., Yordanova, J., Kolev, V., Figliola, A., Schürmann, M., and Başar, E. Wavelet entropy: a new tool for analysis of short duration brain electrical signals. *Journal of Neuroscience Methods 105*, 1 (2001), 65 – 75.

[128] Russell, J. A. A circumplex model of affect. *Journal of Personality and Social Psychology 39*, 6 (1980), 1161–1178.

[129] S. Wang and, I. P., Johnson, D., Ibrahim Emam, F. G., Oehmichen, A., and Guo, Y. Optimising parallel r correlation matrix calculations on gene expression data using mapreduce. *BMC Bioinformatics 15*, 351 (2014).

[130] Salahuddin, L., Cho, J., Jeong, M. G., and Kim, D. Ultra short term analysis of heart rate variability for monitoring mental stress in mobile settings. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Aug 2007), pp. 4656–4659.

[131] Santiago-Espada, Y., Myer, R. R., Latorella, K. A., and Comstock, J. R. The multi-attribute task battery ii (matb-ii) soft- ware for human performance and workload research: A user's guide., 2011. An optional note.

[132] Schubert, C., Lambertz, M., Nelesen, R., Bardwell, W., Choi, J.-B., and Dimsdale, J. Effects of stress on heart rate complexity a comparison between short-term and chronic stress. 325–32.

[133] Sentz, K., and Ferson, S. Combination of evidence in Dempster-Shafer theory. Tech. Rep. SAND2002-0835, Sandia National Laboratories, Albuquerque, NM, Apr. 2002.

[134] Sentz, K., and Ferson, S. Combination of evidence in Dempster-Shafter theory. Thesis, Binghamton University, 2002.

[135] Shafer, G. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ, 1976.

[136] Shankland, S. Google spotlights data center inner workings. http://www.cnet.com/news/google-spotlights-data-center-inner-workings/.

[137] Shmueli, G. To Explain or to Predict? *Statistical Science*, 3 (2010).

[138] Shvachko, K., Kuang, H., Radia, S., and Chansler, R. The hadoop distributed file system. In *IEEE Symposium on Mass Storage Systems and Technologies* (2010), pp. 1–10.

[139] Sleigh, J. W., Steyn-Ross, D. A., Steyn-Ross, M. L., Grant, C., and Ludbrook, G. Cortical entropy changes with general anaesthesia: theory and experiment. *Physiological Measurement 25*, 4 (2004), 921.

[140] Smets, P. Practical uses of belief functions. In *Proc. Conference on Uncertainty in Artificial Intelligence (UAI)*, K. B. Laskey and H. Prade, Eds. Morgan Kaufmann, San Francisco, CA, 1999, pp. 612–621.

[141] Stephens, C., Kennedy, K., Crook, B., Williams, R., Last, M., and Schutte, P. Mild normobaric hypoxia exposure for human-autonomy system testing. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 61* (2017).

[142] Stephens, C., Kennedy, K., Napoli, N., Demas, M., Crook, B., Williams, R., Last, M., and Schutte, P. Effects on task performance and psychophysiological measures of performance during normobaric hypoxia exposure. *Proceedings of the International Symposium on Aviation Psychology Dayton, OH* (2017).

[143] Stern, R., Ray, W., and Quigley, K. *Psychophysiological Recording*, 3 ed. Cambridge University Press, Cambridge, UK, 2000, ch. 7.

[144] Su, J., Long, Y., Qiu, X., Li, S., and Liu, D. *Anomaly Detection of Single Sensors Using OCSVM_KNN*. Springer International Publishing, Cham, 2015, pp. 217–230.

[145] Swiler, L. P., Paex, T. L., and Mayes, R. L. Epistemic uncertainity quantification tutorial. *Proceedings of the IMAC-XXVII* (2009).

[146] Sztajzel, J. Heart rate variability: A noninvasive electrocardiographic method to measure the autonomic nervous system. *Swiss Medical Weekly 134*, 35-36 (2004), 514–522.

[147] T. E. de Campos. The Chars74K dataset: Character recognition in natural images, Oct. 2012.

[148] Teixeira, Renata Roland and, M. M., Santos, T. V. d. S., Bernardes, J. T. M., Peixoto, L. G., Bocanegra, O. L., Neto, M. B., and Espindola, F. S. Chronic stress induces a hyporeactivity of the autonomic nervous system in response to acute mental stressor and impairs cognitive performance in business executives. *PLoS ONE 10*, 3 (2015).

[149] Thomas, C., and Balakrishnan, N. Modified evidence theory for performance enhancement of intrusion detection systems. In *Proc. International Conference on Information Fusion (ICIF)* (Cologne, Germany, June/July 2008), pp. 1–8.

[150] Thompkins, W. J. *Biomedical Digital Signal Processing*. Prentice Hall, 2000.

[151] Thul, A., Lechinger, J., Donis, J., Michitsch, G., Pichler, G., Kochs, E. F., Jordan, D., Ilg, R., and Schabus, M. {EEG} entropy measures indicate decrease of cortical information processing in disorders of consciousness. *Clinical Neurophysiology 127*, 2 (2016), 1419 – 1427.

[152] Ukil, A., Bandyopadhyay, S., Puri, C., and Pal, A. Heart-trend: An affordable heart condition monitoring system exploiting morphological pattern. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (March 2016), pp. 6260–6264.

[153] Vavilapalli, V. K., Murthy, A. C., Douglas, C., Agarwal, S., Konar, M., Evans, R., Graves, T., Lowe, J., Shah, H., Seth, S., Saha, B., Curino, C., O'Malley, O., Radia, S., Reed, B., and Baldeschwieler, E. Apache hadoop yarn: Yet another resource negotiator, 2013.

[154] Von Tscharner, V. Intensity analysis in time-frequency space of surface myoelectric signals by wavelets of specified resolution. *Journal of Kinesiology and Electromyography 6*, 10 (2000), 433–45.

[155] Wafa, M., and Zagrouba, E. Estimation of mass function in evidence theory for fusion of gray level based images. *The International Conference on Signal and Electronic Systems* (September 2010).

[156] Walter, W. *The Living Brain*. Norton, New York, NY, 1953.

[157] Wang, J., Hu, Y., Xiao, F., Deng, X., and Deng, Y. A novel method to use fuzzy soft sets in decision making based on ambiguity measure and dempster–shafer theory of evidence: An application in medical diagnosis. *Artificial Intelligence in Medicine 69* (2016), 1 – 11.

[158] Wang, X.-J. Neurophysiological and computational principles of cortical rhythms in cognition. *Physiological Reviews 90*, 3 (2010), 1195–1268.

[159] Wilson, G. F. Psychophysiological Test Methods and Procedures. In *Handbook of Human Factors Testing and Evaluation*, S. G. Charlton and T. G. O'Brien, Eds., 2nd ed. LAWRENCE ERLBAUM ASSOCIATES, Mahwah, NJ, 2001, ch. 7, pp. 127–156.

[160] Yager, R. R., Fedrizzi, M., and Kacprzyk, J., Eds. *Advances in the Dempster-Shafer Theory of Evidence*. John Wiley and Sons, New York, NY, 1994.

[161] Yager, R. R., and Liu, L. *Classic works of Dempster-Shafer of Belief Functions. Studies in Fuzziness and Soft Computing*. Springer, 2008.

[162] Yager, R. R., and Liu, L., Eds. *Classic Works of the Dempster-Shafer Theory of Belief Functions*. Studies in Fuzziness and Soft Computing. Springer-Verlag Berlin Heidelberg, 2008.

[163] Yoo, J., and Ahn, C. Image matching using peak signal-to-noise ratio-based occlusion detection. *IET Image Process 6*, 5 (2012), 483–495.

[164] Zanin, M., Zunino, L., A. Rosso, O., and Papo, D. Permutation entropy and its main biomedical and econophysics applications: A review. *Entropy 13* (2012), 1553–1577.

[165] Zheng, Q., Chen, C., Li, Z., Huang, A., Jiao, B., Duan, X., and Xie, L. A novel multi-resolution svm (mr-svm) algorithm to detect ecg signal anomaly in wecare project. In *2013 ISSNIP Biosignals and Biorobotics Conference: Biosignals and Robotics for Better and Safer Living (BRC)* (Feb 2013), pp. 1–6.

[166] Zhong, Y., Jan, K.-M., Ju, K. H., and Chon, K. H. Quantifying cardiac sympathetic and parasympathetic nervous activities using principal dynamic modes analysis of heart rate variability. *American journal of physiology. Heart and circulatory physiology 291*, 3 (2006), H1475–83.

[167] Zhu, X., Vondrick, C., Ramanan, D., and Fowlkes, C. Do we need more training data or better models for object detection?. In *BMVC* (2012), vol. 3, Citeseer, p. 5.