A Comprehensive Assessment of Variations in Electrocardiogram Morphology in Risk Assessment of Cardiovascular Death post-Acute Coronary Syndrome

by

Priya Parayanthal

S.B., Massachusetts Institute of Technology (2010)

Submitted to the Department of Electrical Engineering and Computer Science in Partial Fulfillment of the Requirements for the Degree of Master of Engineering in Electrical Engineering and Computer Science at the MASSACHUSETTS INSTITUTE OF TECHNOLOGY May 2011

© 2011 Massachusetts Institute of Technology. All rights reserved.

Author

Department of Electrical Engineering and Computer Science

May 20, 2011

Certified by

Collin M. Stultz
Associate Professor of Electrical Engineering and Computer Science and Associate Professor of Health Sciences Technology

Thesis Supervisor

Accepted by

Dr. Christopher J. Terman
Chairman, Masters of Engineering Thesis Committee
A Comprehensive Assessment of Variations in Electrocardiogram Morphology in Risk Assessment of Cardiovascular Death post-Acute Coronary Syndrome

by

Priya Parayanthal

Submitted to the Department of Electrical Engineering and Computer Science

May 20, 2011

In Partial Fulfillment of the Requirements for the Degree of Master of Engineering in Electrical Engineering and Computer Science

ABSTRACT

Millions of patients worldwide are hospitalized each year due to an acute coronary syndrome (ACS). Patients who have had an acute coronary syndrome are at higher risk for developing future adverse cardiovascular events such as cardiovascular death, congestive heart failure, or a repeat ACS. Currently, there have been several electrocardiographic metrics used to assess the risk of ACS patients for a future cardiovascular death including heart rate variability, heart rate turbulence, deceleration capacity, T-wave alternans, and morphologic variability.

This thesis introduces new ECG-based metrics that can be used to risk-stratify post-ACS patients for future cardiovascular death and evaluates the clinical utility of the existing electrocardiogram based metric known as morphologic variability (MV). We first analyze a metric called weighted morphologic variability (WMV) which is based on assessment of beat-to-beat morphology changes in the ECG. In addition, we introduce machine learning methods with morphology based features to separate post-ACS patients into high risk or low risk for cardiovascular death. Finally, we aim to increase the clinical utility of MV by creating a metric that can achieve good risk stratification when applied to a small amount of data. The body of this work suggests that morphologic variability is an effective metric in prognosticating post-ACS patients into high risk and low risk for cardiovascular death.

Thesis Supervisor: Collin M. Stultz
Title: Associate Professor of Electrical Engineering and Computer Science and Associate Professor of Health Sciences Technology
ACKNOWLEDGEMENTS

I would like to express my sincerest gratitude to Professor Collin Stultz for all of the support, advice, and encouragement he has given me. I would not be where I am today if it wasn’t for him. He has contributed significantly to my development and his keen insight has allowed me to evolve into a more effective researcher. His dedication, perseverance, and passion for science will always be a source of inspiration for me.

I am also indebted to Professor John Guttag and his students, Jenna Wiens, Anima Singh, and Garthee Ganeshapillai, for their continuous help and suggestions. In addition, I want to thank all my wonderful lab mates: Dr. Sophie Walker, Dr. Sarah Bowman, and the to-be doctors Charles Fisher, Orly Ullman, and Elaine Gee for making the lab so much fun.

I would like to extend a very special word of thanks to my friends. I am incredibly fortunate to have found such an amazing group of friends who have been there for me and have provided me with much laughter throughout my time at MIT. I would especially like to thank Shlee and Derka for their tremendous support in NYC, North America, Earfs, Milky Way Galaxy, Universe.

Finally, this work would not have been possible without the support of my family. I would like dedicate my efforts over the past couple of years to my parents, Asha and Padman Parayanthal. I am truly at a loss for words to express how thankful I am for all of their love, patience, and support. Their own determination and successes have inspired me to come this far and I am eternally indebted for everything that they have empowered me to achieve.
CONTENTS

ABSTRACT ........................................................................................................................... 3
ACKNOWLEDGEMENTS ........................................................................................................ 5
Chapter 1: INTRODUCTION ............................................................................................... 15
   A. Background .................................................................................................................. 16
      1. Cardiovascular Physiology ......................................................................................... 16
      2. Cardiac Electrophysiology ......................................................................................... 17
      3. Electrocardiogram ...................................................................................................... 19
      4. Atherosclerosis .......................................................................................................... 20
      5. Acute Coronary Syndromes ....................................................................................... 21
   B. Risk Stratification Measures ......................................................................................... 22
      1. Non ECG-Based, Non-Invasive Risk-Stratification Measures ...................................... 22
      2. Electrocardiographic Risk-Stratification Measures ................................................... 23
         a. Heart Rate Variability ......................................................................................... 23
         b. Heart Rate Turbulence ......................................................................................... 25
         c. Deceleration Capacity ......................................................................................... 27
         d. T-Wave Alternans ............................................................................................... 29
Chapter 2: EXTENSIONS OF THE MORPHOLGIC VARIABILITY METRIC ................. 31
   A. Morphologic Variability .............................................................................................. 31
      1. Methods .................................................................................................................... 31
         a. Morphologic Distance Time Series .................................................................... 31
b. Deriving a Morphologic Variability Measure from the MD Time Series........ 33

2. Results.................................................................................................................. 37
   a. Weighted Morphologic Variability................................................................. 38
      i. Weighted Morphologic Variability Results (n = 600)................................. 39
      ii. Weighted Morphologic Variability Results (n = 3)................................. 41

Chapter 3: MACHINE LEARNING APPLICATIONS TO MORPHOLOGIC

   VARIABILITY........................................................................................................ 45

A. Supervised Learning............................................................................................. 45
   1. Support Vector Machines.................................................................................... 46
   2. Support Vector Machines with Linear Kernels................................................... 46
   3. Support Vector Machines with Non-Linear Kernels............................................ 50

B. Linear SVMs with Morphology-Based Features.................................................... 53
   1. Normalization of the Power Spectrum............................................................... 54
   2. Validation............................................................................................................. 55
   3. Evaluation of Linear SVM's with Morphology Based Features........................... 55

Chapter 4: INCREASING THE CLINICAL UTILITY OF MORPHOLOGIC

   VARIABILITY........................................................................................................ 59

A. Morphologic Variability Performance over Time............................................... 59

B. Morphologic Variability Performance using less Data......................................... 61
   1. 1st Hour Analysis................................................................................................ 61
      a. Varied Thresholds.......................................................................................... 61
      b. Sliding 5-Minute Windows............................................................................. 66
      c. Sliding 5-Minute Windows and Varied Thresholds........................................ 69
Chapter 5: SUMMARY AND CONCLUSIONS

A. Morphologic Variability
B. Weighted Morphologic Variability
C. Linear Support Vector Machines with Morphology-Based Features
D. Increasing the Clinical Utility of Morphologic Variability

BIBLIOGRAPHY

APPENDIX
LIST OF FIGURES

Figure I-1. Physiology of the Cardiovascular System ................................................ 17
Figure I-2. Electrical Conduction Pathway of the Heart ............................................ 18
Figure I-3. Electrocardiogram .................................................................................. 20
Figure I-4. Heart Rate Turbulence Calculation ....................................................... 27
Figure I-5. Anchor and Segment Selection for Deceleration Capacity ....................... 28
Figure I-6. T-wave Power Spectrum for Electrical Alternans ................................... 30
Figure II-1. Dynamic Time Warping Beat Alignment ................................................ 32
Figure II-2. Morphologic Variability Heat Map ....................................................... 36
Figure II-3. Weights for Various Frequencies in WMV(n=600) Risk Stratification ........ 40
Figure III-1. Linear SVM Separating Positive and Negative Examples ....................... 47
Figure III-2. Linear SVM with Slack ........................................................................ 49
Figure III-3. SVM with Non-Linear Kernels ............................................................. 51
Figure III-4. SVM with Radial Basis Function Kernel .............................................. 52
Figure III-5. Linear SVM Test vs. Train Results (60 Bands) ...................................... 56
Figure III-6. MV vs. Linear SVM Results (60 Bands) ............................................. 57
Figure IV-1. MV Performance Over Time ................................................................. 60
Figure IV-2. MV-1hr using Threshold 11 vs. MV C-Statistic Results ......................... 63
Figure IV-3. MV-1hr Averaging Threshold 11 & 12 vs. MV C-Statistic Results ........... 64
Figure IV-4. MV-1hr vs. MV 90-day Hazard Ratio Results ...................................... 66
Figure IV-5. MV-1hr with Sliding 5-min Windows vs. MV C-Statistic Results ............ 67
LIST OF TABLES

Table I-1. Summary of Statistical HRV Measures ................................................................................. 25
Table II-1. MV Train vs. Test Hazard Ratio Results ................................................................................. 38
Table II-2. WMV(n = 600) Train vs. Test Hazard Ratio Results ................................................................. 41
Table II-3. WMV(n = 3) Train vs. Test C-Statistic Results ................................................................. 43
Table II-4. WMV(n = 3) vs. MV Training Results ................................................................................. 44
Table II-5. WMV(n = 3) vs. MV Test Results ....................................................................................... 44
Table IV-1. MV-1hr with Varying Thresholds vs. MV ............................................................................. 64
Table IV-2. MV-1hr with Sliding 5-min Windows vs. MV ................................................................. 68
Table IV-3. MV-1hr with Sliding 5-min Windows and Varied Thresholds vs. MV ..................................... 70
Chapter I: INTRODUCTION

Millions of Americans are hospitalized every year with an Acute Coronary Syndrome (ACS) - a clinical event in which blood supply to part of the heart muscle (myocardium) is severely reduced. Acute Coronary Syndromes are classified into two groups: unstable angina, a state in which there is no evidence that the myocardium is permanently damaged, and a heart attack in which it is. Patients who are diagnosed with an Acute Coronary Syndrome have increased risk of future adverse cardiovascular events which could include death from fatal arrhythmias, a repeat ACS, or congestive heart failure. As a result, it is very important to be able to accurately identify patients who are at high risk of developing such events. Risk assessment enables physicians to decide what type of treatments patients should receive. For example high risk patients usually benefit from more invasive treatments (e.g., coronary angiography), which may entail some risk, while low risk patients can be treated with lower risk therapies (oral medications alone).

Currently, there are several techniques that are used to predict risk for adverse cardiovascular outcomes in post-ACS patients. Some non-invasive tests used to help risk stratify patients include Cardiac Magnetic Resonance Imaging (MRI), Cardiac Computed Tomography (CT), and Cardiac Ultrasound, however these tests may not be readily available in a number of rural settings. By contrast, Electrocardiographic (ECG) –based risk metrics are relatively efficient, inexpensive, and readily available. This thesis introduces new ECG-based metrics that can be used to risk-stratify post-ACS patients for future adverse cardiovascular events and evaluates the clinical utility of the existing ECG-based metric known as morphologic variability (MV).
This thesis is organized as follows: Chapter 1 focuses on background information on the cardiovascular system, electrocardiogram, and existing risk-stratification techniques. Chapter 2 describes the MV metric and introduces a new variation to the MV metric, known as weighted morphologic variability (WMV). Chapter 3 evaluates the application of linear support vector machines (SVM) with morphology based features in post-ACS risk stratification. Chapter 4 provides an analysis of several methods used to increase the clinical utility of the MV metric. Finally, Chapter 5 concludes with a summary of the findings and conclusions in the thesis.

A. Background

1. Cardiovascular Physiology

The main function of the cardiovascular system is to provide cells with oxygen and nutrients as well as the transport of hormones to target cells and organs. The main components of the cardiovascular system are the heart and blood vessels which can be further subdivided into two separate circulations connected in series. First there is the pulmonary circulation, which carries blood through the lungs for oxygenation, and second there is the systemic circulation which delivers oxygenated blood to the rest of the body (Figure I-1).

The heart pumps blood to various parts of the body through rhythmic contractions. The heart itself has four separate chambers: the left and right atria, and the left and right ventricles (Figure I-1). The left atrium collects oxygenated blood from the lungs and passes it through the mitral valve to the left ventricle which then passes the blood through the aortic valve to the aorta which distributes blood to the rest of the body. Then, the deoxygenated blood flows into the right atrium which delivers the blood through the tricuspid valve to the right ventricle. Finally, the right ventricle passes the blood through the pulmonic valve to the lungs for oxygenation.
The heart beat is formed by contraction of the atria followed by contraction of the ventricles.

[Error! Reference source not found.]

Figure 0-1. Physiology of the Cardiovascular System. The heart pumps blood carrying essential nutrients and oxygen throughout the body. Oxygenated blood from the lungs is sent to the left atrium and then to the left ventricle which sends the blood to the rest of the body. The deoxygenated blood is then passed to the left atrium and then to the left ventricle where is again sent to the lungs for oxygenation. Image courtesy of Daily Dose of Fitness [2].

2. Cardiac Electrophysiology

In order for the heart to pump blood through the vasculature, different parts of the heart must be electrically stimulated to contract. A depolarization front which consists of electrical impulses that stimulate muscle contraction originates at the sinoatrial (SA) node and then is conducted to the entire myocardium in a specific, timed sequence such that the atria contract before the ventricles. The electrical impulses are conducted through the heart by the
depolarization and repolarization of myocardial cells. At rest, a myocardial cell, or myocyte, remains at a negative potential relative to the outside of the cell. If the myocyte is stimulated, it becomes depolarized as positive ions flow into the cell. The cell then repolarizes and returns to its normal, resting state.

In normal sinus rhythm, the depolarization front is initiated at the sinoatrial node which serves as the pacemaker (Figure 1-2). These electrical impulses then spread to the right and left atria, then to the atroventricular (AV) node, the bundle of His, the left and right bundle branches, the smaller bundles of the Purkinje system, and finally to the myocardium itself.

[Error! Reference source not found.]

![Figure 1-2: Electrical Conduction Pathway of the Heart.](image-url)

Electrical impulses begin at the SA node and spread through the atria, stimulating them to contract and pump blood into the right and left ventricles. The impulses then enter the atroventricular (AV) node, and then spread throughout the ventricles, thus also stimulating them to contract. This conduction system determines the timing of the heart beat and causes the heart to contract in a coordinated manner. Image courtesy of Up To Date [3].
3. Electrocardiogram

The electrocardiogram measures potential differences on the surface of the body that correspond to the electrical activity of the heart. Electrocardiographic data can be collected through the use of Holter monitors. Holter monitors involve the placement of 3 to 12 leads which are electrodes, placed on various parts of the body, that measure and record voltage changes along specific axes. Holter monitors typically collect ECG data for at least 24 hours. Many risk stratification methods, including the one discussed in this thesis, only require data from a single lead.

A heart beat is typically divided into segments as shown in Figure I-3a. The P wave corresponds to right and left atrial depolarization. The PR interval is the time from initial depolarization of the atria to initial depolarization of the ventricles and within this, the PR segment is an isoelectric region that corresponds to conduction of the electrical signal from the atria to the AV node (see Figure I-2). The QRS complex corresponds to right and left ventricular depolarization. The magnitude of the QRS complex is much larger than that of the P-wave because the ventricles are much larger in size than the atria. Atrial repolarization also occurs during this time however it is not seen on a normal ECG because its amplitude is much smaller and generally “buried” in the QRS complex. Finally, the T wave indicates ventricular repolarization [1].
Figure I-3. *Electrocardiogram*. Figure I-3a depicts the parts of a single heartbeat in the electrocardiogram. The P-wave corresponds to the depolarization of the atria, the QRS complex corresponds to the depolarization of the ventricles (the repolarization of the atria are hidden within the QRS complex), and the T-wave corresponds to the repolarization of the ventricles. Image courtesy of Skipping Hearts [4]. Figure I-3b shows a sample electrocardiogram recording. Image courtesy of Malarvilli [5].

4. *Atherosclerosis*

Atherosclerosis is a condition in which there is an aggregation of lipids, cells, and other substances within the arterial wall. A discrete localization of these substances within the arterial wall is called an atherosclerotic plaque. When this build-up occurs in a patient’s coronary artery, the patient is diagnosed with coronary artery disease. If the plaques become large with respect to
the lumen of the vessel, or if they rupture leading to the formation of a clot in the vessel lumen, blood flow through the artery can be severely reduced and this results in an acute coronary syndrome [5].

5. Acute Coronary Syndromes

An acute coronary syndrome (ACS) is an event in which the blood supply to the heart is severely reduced. Some common symptoms of an ACS are angina (chest pain), nausea, dyspnea (shortness of breath), and diaphoresis (sweating). An ACS is usually caused by atherosclerotic plaque rupture.

ACS is generally classified into either unstable angina in which myocardium is not permanently damaged, or a myocardial infarction, in which the tissue is permanently damaged. In addition, ACS’s are further classified based on the extent of the occlusion of an artery. When the ECG shows elevation of the ST segment, this indicates complete occlusion of an artery. This is known as a ST-elevation myocardial infarction (STEMI). Non-ST elevation is less severe and indicates a partial occlusion of an artery. If necrosis also occurs, which is when the myocardium is permanently damaged, the patient is diagnosed with non-ST elevation myocardial infarction (NSTEMI) [6]. Patients diagnosed with an ACS are at higher risk of experiencing a future adverse cardiovascular event including death from fatal arrhythmias, a repeat ACS, or congestive heart failure. Thus, it is important to be able to accurately identify high risk patients so that doctors can offer them more aggressive treatments that may decrease their risk of these future adverse cardiac events.
B. Risk Stratification Measures

There are several existing techniques that can be used to determine the risk that post-ACS patients have of a future adverse cardiovascular outcome. One such technique is cardiac catheterization, in which a catheter is inserted into an artery and advanced into chambers of the heart or the coronary arteries [7]. This provides a direct visualization of the lumen of the coronary arteries as well as an assessment of the pressures in the various heart chambers and the overall cardiac function. However, it requires cannulation of the great vessels and therefore does entail some risk, thereby making it somewhat less desirable for low-risk patients.

The following section will describe a few non-invasive methods as well as electrocardiographic techniques for risk stratification of post-ACS patients.

1. Non ECG-Based, Non-Invasive Risk-Stratification Measures

The Thrombolysis in Myocardial Infarction (TIMI) risk score is a simple metric used to asses a patients risk for death and subsequent ischemic events, thereby providing doctors with a better basis for therapeutic decision making. The score is determined from seven independent factors which include: age \( \geq 65 \) years, presence of at least three risk factors for cardiac heart disease, prior coronary stenosis \( \geq 50\% \), presence of ST segment deviation, at least two anginal episodes in the preceding 24 hours, use of aspirin in the preceding 7 days, and elevated serum cardiac markers [8]. One point is given for each of these factors that a patient has, making the total TIMI risk score range from 0 to 7. Patients with a score of 0-2 are classified as low risk, 3-4 as intermediate risk, and 5-7 as high risk for a future adverse cardiovascular event. While the TIMI risk score is an effective and simple prognostication scheme that categorizes patients into high, intermediate, or low risk for death or ischemic events, the score does not provide a
quantitative statement about finer gradations of risk that exist clinically. Therefore, additional methods for accurate risk stratification are needed.

Other non-invasive tests include Cardiac Magnetic Resonance Imaging (MRI), Cardiac Computed Tomography (CT), and Cardiac Ultrasound all of which are techniques used to image the heart. The Cardiac MRI uses powerful magnetic fields and frequency pulses to produce detailed pictures of organs and can be used to evaluate the anatomy and function of the heart. The Cardiac CT involves imaging of the heart through the use of x-rays. Cardiac Ultrasound (Echocardiography) uses ultrasound techniques to image two and three-dimensional views of the heart and therefore can produce an accurate assessment of cardiac function [1]. While these techniques are useful in risk stratification, some are expensive and many are not readily available in rural settings. Electrocardiographic (ECG) -based methods have proven to be much more efficient and practical as they are routinely acquired for all patients for monitoring purposes and they are also non-invasive and inexpensive.

2. **Electrocardiographic Risk-Stratification Measures**

a. **Heart Rate Variability**

Heart rate variability (HRV) is a measure of variations in a patient’s heart rate and indirectly enables one to assess the health of the autonomic nervous system. More precisely, the heart rate is primarily modulated by the autonomic nervous system, which can be subdivided into the sympathetic and parasympathetic nervous systems. In healthy people, the body continuously compensates for changes in metabolism by varying the heart rate through regulation by the autonomic nervous system. Thus, if the heart rate is not very variable, it indicates decreased modulation by the sympathetic and parasympathetic nervous systems, meaning the heart control
systems are not appropriately responding to stimuli. Lower heart rate variability is associated with a higher risk for developing a future adverse cardiovascular event [9].

HRV is a measure of the variability in RR intervals, which are the difference in time between successive R waves in the ECG. Normal RR intervals (also called NN intervals) are the difference in time between adjacent QRS complexes in beats that begin at the sino-atrial node and follow a normal conduction path through the myocardium (Figure I-2). The NN interval corresponds to the instantaneous heart rate. Simple time domain measures of HRV include the mean NN interval, difference between the longest and shortest NN intervals, and the mean heart rate, the difference between night and day heart rate, etc. Statistical time domain measures can also be calculated when working with signals recorded over longer periods of time. The first of these statistical measures is the standard deviation of NN intervals (SDNN) which is the mean of all 5-minute standard deviations of NN intervals during a 24-hour period. However, this is not a particularly good measure because it is dependent on the length of the recording. Another commonly used statistical measure is the standard deviation of the average NN intervals (SDANN) which is the standard deviation of the mean NN interval in a five-minute window. Additional time domain measures are introduced in Table I-1.

The frequency domain metrics use the power spectral density of the series of NN intervals. The LF/HF metric is the ratio of the total power in a low frequency (LF) band to a high frequency (HF) band. This metric is defined as:

\[
HRV(LF/HF) = \frac{Power \ between \ 0.04 \ and \ 0.15Hz}{Power \ between \ 0.15 \ and \ 0.4 \ Hz}
\]  
(Equation I-1)
The ratio is computed for all 5-minute windows and the median across all windows is the LF/HF value for a particular patient. Patients with high LF/HF ratios are considered to be at high risk of death post-ACS [10]. Additional frequency domain measures are mentioned in Table I-1.

<table>
<thead>
<tr>
<th>Statistical Measure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Time Domain Measures:</strong></td>
<td></td>
</tr>
<tr>
<td>SDNN index: standard deviation of the NN intervals</td>
<td>- mean of all 5-minute standard deviations of NN intervals in a 24 hour period</td>
</tr>
<tr>
<td>SDANN index: standard deviation of the average NN intervals</td>
<td>- mean of all 5-minute standard deviations of average NN intervals calculated over 24 hours</td>
</tr>
<tr>
<td>r-MSSD: root mean squared successive differences</td>
<td>- square root of the mean of the squared differences between successive NN intervals over 24 hours</td>
</tr>
<tr>
<td>pNN50</td>
<td>- percentage of differences between successive NN intervals over 24 hours that are greater than 50ms</td>
</tr>
<tr>
<td><strong>Frequency Domain Measures:</strong></td>
<td></td>
</tr>
<tr>
<td>HF: high frequency</td>
<td>- sum of power between 0.15Hz to 0.4Hz in power spectrum</td>
</tr>
<tr>
<td>LF: low frequency</td>
<td>- sum of power between 0.04Hz to 0.15Hz in power spectrum</td>
</tr>
<tr>
<td>VLF: very low frequency</td>
<td>- sum of power between 0.0033Hz to 0.04Hz in power spectrum</td>
</tr>
<tr>
<td>TP: total power</td>
<td>- sum of power spectrum across all frequencies in 5-min recordings (net effect of all possible physiological mechanisms contributing to variability in heart rate)</td>
</tr>
<tr>
<td>LF/HF Ratio</td>
<td>- (sum of power between 0.04-0.15Hz) / (sum of power between 0.15-0.4Hz)</td>
</tr>
</tbody>
</table>

Table I-1. Summary of Statistical HRV Measures.

**b. Heart Rate Turbulence**

Heart rate turbulence (HRT) is another metric assessing the risk of patient’s status post-ACS. This metric is based on changes in the heart rate in that it assesses the response of the heart rate following a premature ventricular contraction (PVC). A PVC is an abnormal heart
rhythm in which a beat originates from the ventricles rather than the sino-atrial node. The premature beat results in a reduction in the amount of blood ejected from the heart (a lower ejection fraction) leading to a lower blood pressure than expected. As a result, the autonomous nervous system increases the heart rate in an attempt to raise blood pressure. Following this phase, the heart returns to the baseline heart rate. If the heart takes too long to return to this homeostatic level, it may indicate some problem with baroreflex (mechanism for controlling blood pressure) sensitivity [11].

HRT is first characterized by the spontaneous initial acceleration or the turbulence onset (TO), which is the relative change of RR intervals immediately preceding and following a PVC. More specifically, the TO is the difference between the mean of the last two normal RR intervals before the PVC (RR-2 and RR-1) and the first two normal RR intervals following the PVC (RR1 and RR2):

$$TO = \frac{(RR_1 + RR_2) - (RR_{-2} + RR_{-1})}{(RR_{-2} + RR_{-1})} \times 100$$

(Equation I-2)

In addition, HRT is characterized by the Turbulence Slope (TS) which is a measure of the slowing of the heart rate as it returns to its baseline value after a PVC. It is quantified as the maximum slope of a regression line over any five consecutive normal RR intervals following a PVC. If a patient has a lower TS, this means that the patient’s heart takes a longer period of time to return to its baseline level and may indicate an unhealthy autonomic nervous system [12].
Figure 1-4. Heart Rate Turbulence Calculation. Heart rate turbulence (HRT) is characterized by the turbulence onset (TO) and turbulence slope (TS). The TO is quantified as the relative change of the RR intervals from before to after a PVC. In this example, the average of the two RR intervals preceding the PVC is 1000 ms and the average of the 2 RR intervals following the PVC is 960 ms. Thus, the TO = (960-1000)/1000 × 100 = -4%. The TS is the maximum regression slope of 5 consecutive normal RR intervals within the first 15 RR intervals immediately following a PVC. In this example, the regression line that is fit to beats 3-7 is labeled slope 3 and the line that is fit to beats 6-10 is labeled slope 6. The TS = 36.4 ms/beat because the regression line that is fit to beats 3-7 has the largest slope among all possible regression lines. Image courtesy of Watanabe [13].

Patients are given a final HRT score of 0 if both the TO and TS are normal, 1 if either the TS or the TO is abnormal, and 2 if both the TS and TO are abnormal. Post-ACS patients who were given a high HRT score of 2 were at much higher risk for future cardiovascular death [14].

c. Deceleration Capacity

Deceleration capacity (DC) is an extension of HRT and focuses on regions where the heart rate slows down; i.e. regions that correspond to vagal modulation. First, anchors, which are defined as a longer RR interval following a shorter RR interval, are identified along an ECG
recording. Next, segments of data around the anchors, equivalent in size, are selected (Figure I-5).

Step 1: Definition of anchors

Step 2: Definition of segments

**Figure I-5. Anchor and Segment Selection for Deceleration Capacity.** Anchors are defined by a heartbeat interval that is longer than the preceding interval. These correspond to black circles in the figure. The bars in the figure correspond to segments of data surrounding the anchors. These segments are the same size. The deceleration capacity is defined as the average of the RR interval of the anchor with the RR interval immediately following the anchor and the last two RR intervals immediately preceding the anchor. Image courtesy of Bauer [1515].

The lengths of the two RR intervals immediately preceding the anchor are X(-2) and X(-1), the length of the anchor is X(0), and the length of the RR interval immediately following the anchor is X(1). The resulting DC metric is defined as:
\[ DC = \frac{X(0) + X(1) - X(-1) - X(-2)}{4} \]

(Equation I-3)

It has been found that diminished deceleration capacity is an important prognostic marker for death after myocardial infarction [15].

\textit{d. T-Wave Alternans}

T-wave alternans (TWA) is a measure of beat-to-beat variations in the amplitude of the T-wave in the electrocardiogram. These variations are very subtle and are often only discovered at the microvolt level. It has been found that electrical alternans affecting the T-wave and even the ST-segment are associated with increased risk for ventricular arrhythmias as well as sudden cardiac death. Alternation between successive beats in the ECG morphology is most simply alteration in time aligned sampled values of successive beats of the ECG. However, variation between time aligned beats of the ECG is often confounded by noise, respiratory modulation, muscle artifact, etc., making it difficult to identify every-other-beat variations. As a result, the signal is converted into a power spectrum which is in the frequency domain, to separate out relevant frequency components. This power spectrum is calculated by taking the discrete Fourier transform of the Hanning-windowed sample autocorrelation function of the T-wave and ST segments (Figure I-6) [16].
Figure 1-6. *T-wave Power Spectrum for Electrical Alternans.* The power spectrum of the T-wave is shown above. T-wave alternans is the amplitude of the peak of the power spectrum at the 0.5 frequency which corresponds to every-other-beat variations. Because alternans are typically subtle changes on the microvolt level, they are very difficult to detect by simply looking at the electrocardiogram. This is why the T-wave power spectrum is created. The spectrum at 0.5 cycles per beat in this figure shows a clear peak and thus the presence of alternans variation. Image courtesy of Rosenbaum [17].

A spectral peak at the 0.5 cycle per beat frequency corresponds to beat-to-beat variations and the magnitude at this peak is a measure of electrical alternans. The alternans ratio is calculated as:

\[
\text{alternans ratio} = \frac{\text{alternans peak} - \text{mean}(\text{noise})}{\sigma_{\text{noise}}}
\]

(Equation I-4)

Patients with alternans ratios greater than 2.5 in either the T-wave or ST segment are found to be vulnerable to arrhythmias. While there is high correlation between electrical alternans and risk for arrhythmias, electrical alternans require sensitive equipment to detect because they operate at the micro-volt level [17].
Chapter II:
EXTENSIONS OF THE MORPHOLOGIC VARIABILITY METRIC

A. Morphologic Variability

Morphologic variability (MV) quantifies beat-to-beat changes in the shape of the heartbeat. Because damaged myocardial tissue does not conduct an electrical signal the same way that undamaged tissue does, changes in the morphology of the ECG waveform may indicate some underlying problem with electrical conduction through the heart. In order to calculate the MV of a patient, individual beats of the patients ECG recording must be identified. Analogous to the NN time series used for the HRV metric, MV is calculated from an intermediate time series called the morphologic distance (MD) time series which captures differences between two successive beats. Similar to what is done for the HRV frequency measures, this time series is then converted into the frequency domain, yielding a power spectrum from which the MV metric is quantified [18].

1. Methods

a. Morphologic Distance Time Series

To calculate the morphologic distance (MD) time series, a technique called dynamic time warping (DTW) is used to quantify subtle differences between the shapes of successive beats. The first step in comparing successive beats is to align them to avoid matching segments of one waveform that correspond to a different segment of another waveform. As shown in Figure II-1, if the red and blue beats were simply subtracted, we would be comparing the middle of the T-wave in the red beat to the end of the T-wave in the blue beat. The DTW algorithm is used to
match parts of the ECG that correspond to the same physiologic processes so that relevant conduction phases are compared.

Figure II-1. Dynamic Time Warping Beat Alignment. In the figure on the left, samples in the red beat are directly compared to samples that occur at the same time in the blue beat. However, when beats are directly compared in this manner, regions of one beat might be compared to regions of another beat which correspond to completely different physiological processes. In this example, the middle of the T-wave in the red beat is compared to the end of the T-wave in the blue beat. The figure on the right shows that DTW fixes this issue and aligns the beats in such a way the same points in the conduction path of the two different beats are being compared. Image courtesy of Syed [18].

DTW is performed on successive beats and involves creating an m x n matrix, where m is the length of the first beat and n is the length of the second beat. Each element in the matrix corresponds to the variability, or Euclidean distance $(A(i)-B(j))^2$ between the $i^{th}$ sample of the first beat and the $j^{th}$ sample of the second beat. Any particular alignment corresponds to a path, $\varphi$, of length $K$ is defined as:

$$\varphi(k) = (\varphi_A(k), \varphi_B(k)), 1 \leq k \leq K$$

(Equation II-1)

Where $\varphi_A$ represents the row index and $\varphi_B$ represents the column index of the distance matrix. The optimal dynamic time warping alignment is the path through this matrix with the
minimum associated cost. The cost is defined as the sum of squares of the differences between pairs of matched elements under all allowable alignments. Given two beats, $x_A$ and $x_B$, the cost of the alignment path is defined as:

$$C_{\phi}(x_A, x_B) = \sum_{k=1}^{k} d(x_A[\phi_A(k)], x_B[\phi_B(k)])$$

(Equation II-2)

Thus, the DTW between these two beats is:

$$DTW(x_A, x_B) = \min_{\phi} C_{\phi}(x_A, x_B)$$

(Equation II-3)

The final DTW energy difference captures both amplitude changes as well as the length $K$ of the alignment path, thus also capturing timing differences between two beats [19].

The MD time series is formed by computing the DTW distance of each beat with the previous beat in a pair-wise manner. First, successive beats of the ECG are aligned using dynamic time warping and then the MD is calculated by taking the sum of squares energy difference, or dynamic time warping cost, between aligned beats. To smooth this sequence a median filter of length 8 is then applied, and the resulting sequence is known as the MD time series [18].

**b. Deriving a Morphologic Variability Measure from the MD Time Series**

As in HRV, frequency domain based measures of morphology differences can also characterize the variability of successive beats. It is believed that frequency-based metrics are more robust because high frequency noise in ECG recordings does not interfere with measurements of the low frequency components that are physiologically relevant [6]. Power spectra are calculated for a given patient and these power spectra are used to quantify MV for
that particular patient. The final MV metric is calculated as the sum of powers across a
particular frequency band in a patient’s power spectrum.

Similar to HRV, a patient’s MD time series is broken up into 5-minute intervals and a
power spectrum is calculated for each of these intervals. The power spectrum is the energy in
the MD time series for each of these 5-minute intervals and in order to characterize the
morphologic changes in each of these intervals, the power in a particular diagnostic frequency
band is summed:

\[ MV_\theta = \sum_{v=LF}^{HF} Power_\theta(v) \]

(Equation II-4)

Where \( \theta \) represents a particular 5-minute interval and \( v \) is the frequency which ranges from a low
frequency (LF) to a high frequency (HF) cutoff. The final MV value for each patient is taken as
the 90\(^{th}\) percentile value across the sums for all 5-minute intervals, or the 90\(^{th}\) percentile value of
the MV\(\theta\)'s for each patient. A threshold of 90\% was found to have improved risk stratification
and discrimination quality while avoiding the effects of noise.

In order to find the optimal diagnostic frequency band, all combinations of low frequency
and high frequency thresholds between 0.10Hz and 0.60Hz in intervals of 0.01Hz were used to
calculate the MV value for each patient. The combination of thresholds that resulted in the
highest correlation between MV and cardiovascular death as represented by the c-statistic, was
taken to be the optimal frequency band. The c-statistic is the area under the receiver operating
characteristic (ROC) curve. The ROC curve measures the true positive rate, or sensitivity,
versus the false positive rate, or 1-specificity. In this case, the curve would be the proportion of
correctly identified patients at high risk of future adverse cardiovascular outcomes versus the rate
of incorrectly identifying patients at high risk. Thus, the goal is to maximize both sensitivity and
specificity. Ultimately, the area under the ROC curve, or the c-statistic, is a measure of the predictive capability of a metric. A larger c-statistic indicates that a metric holds more predictive power and generally, a c-statistic of 0.7 or greater is adequate to distinguish between two outcomes [18].

To determine the frequency band with the largest c-statistic for MV-based risk assessment, we used data from the DISPERSE-2 (TIMI33) trial. This trial enrolled patients who were admitted to the hospital following a non-ST segment elevation acute coronary syndrome (NSTEMI). Patients who were entered into the trial were hospitalized within 48 hours of the NSTEMI incident, experienced ischemic symptoms greater than 10 minutes in duration at rest, and had evidence of myocardial infarction (MI) or ischemia. Patients were followed up for the endpoints of death and MI and there were 15 deaths in this group during the follow up period [20]. A total of 990 patients were enrolled in this trial and after excluding patients with less than 24 hours of ECG data, 764 patients remained. To carry out noise removal, first the ECG signal was median filtered in order to make an estimate of baseline wander, and this wander was subtracted out from the original signal [21]. Next, a wavelet de-noising filter with a soft threshold was applied to the ECG signal to remove any additional noise [22]. Segments of the ECG signal where the signal to noise ratio was significantly low after noise removal were discarded. Parts of the signal with a low signal quality index as well as ectopic beats were removed using the Physionet Signal Quality Index (SQI) package [23]. Finally, the remaining data was segmented into 30-minute intervals and the standard deviation of R-wave amplitudes was calculated. If this standard deviation was greater than 0.2887, the 30 minute interval was discarded. A standard deviation greater than 0.2887 corresponds to the R-wave amplitude
changing uniformly by more than 50% of its mean value and these outliers are thrown out because this is physiologically unlikely [18].

For each combination of thresholds, the c-statistic was calculated and the results were combined to form a heat map (Figure II-2) which graphically shows which combination of low and high frequency thresholds results in best predictive value.

![Morphologic Variability Heat Map](image)

Figure II-2. Morphologic Variability Heat Map. The MV heat map is created by calculating the c-statistic for all combinations of low frequency and high frequency cutoffs. The optimal diagnostic frequency range was found to be 0.3 Hz - 0.55 Hz.

The optimal frequency band for MV in DISPERSE-2 was found to be 0.30Hz to 0.55Hz which resulted in a c-statistic of 0.771. The quartile cutoff of MV was found to be 52.5, and was used to dichotomize patients into a high risk (MV>52.5) or low risk (MV<52.5) group. Patients
in the top quartile of morphologic variability had a 90-day hazard ratio of 8.46 and a 30-day hazard ratio of 12.30, indicating that a high MV value is significantly associated with cardiovascular death in post-ACS patients [18].

2. Results

Morphologic variability methods were tested on data from the MERLIN (TIMI 36) trial which, like the DISPERSE-2 trial, comprised of patients who were admitted to the hospital following a non-ST segment elevation acute coronary syndrome (NSTEACS). The goal of the MERLIN trial was to determine the safety and efficacy of ranolazine in patients with NSTEACS. Patients admitted to this trial were at least 18 years of age, experienced ischemic symptoms greater than 10 minutes at rest, and had either indication of risk for death or ischemic events, evidence of necrosis, at least 1mV of ST-segment depression, diabetes mellitus, or a TIMI risk score of at least 3. There were 6,560 patients in the MERLIN trial who were followed up for a period of two years. After excluding the patients with too little data, 2301 patients remained in the MERLIN Placebo group. There were 102 deaths in the MERLIN Placebo group during the follow up period [2424].

Hazard ratio results for MV when trained on the DISPERSE-2 data and tested on the MERLIN Placebo data are presented in Table II-1. While the morphologic variability metric is found to be highly correlated with the prediction of death in post-ACS patients, it only uses a single frequency band to estimate each patient’s risk. It may be that there is information in the other frequencies that could aid in estimating risk and thus the weighted morphologic variability (WMV) measure was created. This measures uses information in all frequencies of the power spectrum to determine a patient’s risk.
<table>
<thead>
<tr>
<th>Dataset</th>
<th>30-Day Hazard Ratio</th>
<th>90-Day Hazard Ratio</th>
<th>1-Year Hazard Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Results (DISPERSE -2 data)</td>
<td>12.30</td>
<td>8.46</td>
<td>8.46</td>
</tr>
<tr>
<td>Test Results (MERLIN Placebo data)</td>
<td>4.84</td>
<td>5.16</td>
<td>3.29</td>
</tr>
</tbody>
</table>

Table II-1. *MV Train vs. Test Hazard Ratio Results*. The MV metric was trained on the DISPERSE-2 data and tested on the MERLIN Placebo data.

**a. Weighted Morphologic Variability**

WMV is computed in a similar manner to MV however, rather than selecting a single diagnostic frequency band, WMV allows for multiple bands in the power spectrum and applies weights to each of these bands. To quantify WMV, the MD time series is calculated and divided into 5-minute intervals and converted into the power spectrum. The power spectrum, which ranges from 0.001Hz to 0.60Hz, is then divided into $n$ intervals. The WMV value for each of the $n$ intervals, $WMV_i$, is the 90$^{th}$ percentile value of the sum of the power in that particular interval.

$$WMV_i = \sum_{v=LF_i}^{HF_i} Power_g(v)$$

(Equation II-5)

The final WMV value is the sum of the weighted $WMV_i$'s:

$$WMV = \sum_{i=1}^{n} \alpha_i \cdot WMV_i = \sum_{i=1}^{n} \alpha_i \sum_{v=LF_i}^{HF_i} Power_g(v)$$

(Equation II-6)

where $n$ is the number of intervals that the power spectrum is divided into, $\alpha_i$ is the weight applied to each band $i$, $LF_i$ and $HF_i$ are the low frequency and high frequency thresholds.
associated with the particular band $i$, and $Power_i$ is the sum of powers for a particular 5-minute interval. In addition, the weights that multiply each interval must sum to 1.

$$\sum_{i=1}^{n} \alpha_i = 1$$

(Equation II-7)

i. **Weighted Morphologic Variability Results ($n = 600$)**

WMV was first investigated using the smallest possible frequency bands in the power spectrum. The power spectral density of the MD time series ranges from 0.001Hz to 0.60 Hz with a resolution of 0.001Hz. Thus, if we consider each single frequency as a band, we are left with $n = 600$ different single frequency bands. The morphologic variability for a single frequency ($MVSF_i$) is the 90th percentile power from the MD time series for that particular frequency.

$$WMV(n = 600) = \sum_{i=1}^{600} \alpha_i \cdot MVSF_i$$

$$MVSF_i = Power_i(\nu) \text{ where } \nu = i \cdot 10^{-3} \text{ Hz}$$

(Equation II-8)

In the WMV with $n = 600$ measure, an optimal set of weights for each of the 600 single frequencies must be chosen to maximize the c-statistic. This is an NP-hard optimization problem, and thus, simulated annealing can be used to determine the set of weights that best predict cardiovascular death. Simulated annealing is a method of optimization analogous to crystal optimization in which there is a control parameter called temperature that is used to heat and cool an energy function which, in this problem, is the negative of the c-statistic. The
algorithm ultimately finds the global minimum which corresponds to the combination of weights that gives the maximum c-statistic [25].

Figure II-3. Weights for Various Frequencies in WMV(n = 600) Risk Stratification. The optimum combination of weights is determined using the simulated annealing algorithm. The weights in this figure are derived from the MERLIN Placebo dataset and are smoothed using a window size of 0.007 Hz. Image courtesy of Sarker [25].

The WMV metric with $n = 600$ was derived from the MERLIN Placebo data set and its performance was assessed using data from the DISPERSE-2 data set. The cutoff used to dichotomize patients into high risk and low risk groups was determined by finding the point which maximized specificity (rate of correctly identifying patients at low risk) and sensitivity (rate at correctly identifying patients at high risk) on the MERLIN Placebo training data. This
cutoff value was determined to be 1.356 where patients were dichotomized into low-risk (WMV(n=600) < 1.356) and high risk (WMV(n=600) ≥ 1.356).

The training and test results are shown in Table II-2. As shown in these tables, WMV with n = 600 performs better than MV on the training results but fails to perform as well as MV on the test results. The problem with WMV when n = 600 is that all the data in the power spectrum is used and a weight is assigned to every single frequency in the spectrum making it very easy for metric to over-fit to the training data. The metric is not very generalizable as it does not perform as well on other data sets from which it was not developed from. To counter this problem and find a balance between performance and generalizability, we decrease the value of n (number of bands) in the WMV metric as described in the next section.

<table>
<thead>
<tr>
<th>Risk Stratification Metric</th>
<th>Training Results: 90-day Hazard Ratio (MERLIN Placebo)</th>
<th>Test Results: 90-day Hazard Ratio (DISPERSE-2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WMV(n = 600)</td>
<td>6.24</td>
<td>7.15</td>
</tr>
<tr>
<td>MV</td>
<td>5.12</td>
<td>8.46</td>
</tr>
</tbody>
</table>

Table II-2. WMV(n = 600) Train vs. Test Hazard Ratio Results. The WMV(n = 600) metric was trained on the MERLIN Placebo data and tested on the DISPERSE-2 data. The metric did not perform as well as MV on the test data.

**ii. Weighted Morphologic Variability Results (n = 3)**

Three-band weighted morphologic variability, is a subset of WMV in which the power spectrum is divided into n = 3 bands, each with an associated weight. As mentioned, each 5-minute power spectrum is computed from 0.001 Hz to 0.60 Hz in 0.001 increments. In the three-band WMV measure, there are two variable cutoffs, \( \beta_1 \) and \( \beta_2 \), which can take on all possible values between 0.001 Hz and 0.60 Hz in 0.01 Hz increments under the condition that \( \beta_2 > \beta_1 \). Thus, the power spectrum is divided into three bands where \( \text{band}_1 = 0 - \beta_1 \text{ Hz} \), \( \text{band}_2 = \beta_1 - \beta_2 \),
Hz, and \( \beta_3 = \beta_2 - 0.06 \) Hz. As in MV, for each patient, the final energy in each band is taken as the 90th percentile value across all 5-minute intervals. In addition, each band has an associated weight, \( \alpha_1, \alpha_2, \) and \( \alpha_3, \) which range from 0 to 1 in 0.1 increments under the condition that all weights sum to 1. The three-band WMV measure is the weighted sum of energy in each band.

\[
\text{Three-Band WMV} = \alpha_1 \sum_{\nu = 0.03 \text{Hz}}^{\beta_1} \text{Power}_\theta (\nu) + \alpha_2 \sum_{\nu = \beta_1}^{\beta_2} \text{Power}_\theta (\nu) + \alpha_3 \sum_{\nu = \beta_2}^{0.6 \text{Hz}} \text{Power}_\theta (\nu)
\]

(Equation II-9)

The c-statistics were then calculated for three-band WMV under all possible combinations of the cutoffs, \( \beta_1 \) and \( \beta_2, \) and the weights, \( \alpha_1, \alpha_2, \) and \( \alpha_3. \) The optimal combination of cutoffs and weights was defined as that which resulted in the highest c-statistic. The best cutoffs and best weights were found using an exhaustive search method.

As in WMV with \( n = 600, \) WMV with \( n = 3 \) was derived from the MERLIN Placebo data set. An exhaustive search was performed on the training group to find which set of cutoffs and weights maximized the c-statistic. This combination of cutoffs and weights was then applied to the DISPERSE-2 data which served as the test group. The final three-band WMV measure is defined as:

\[
\text{WMV} (n=3) = 0.3 \sum_{\nu = 0.03 \text{Hz}}^{0.03 \text{Hz}} \text{Power}_\theta (\nu) + 0 \sum_{\nu = 0.03 \text{Hz}}^{0.22 \text{Hz}} \text{Power}_\theta (\nu) + 0.7 \sum_{\nu = 0.22 \text{Hz}}^{0.6 \text{Hz}} \text{Power}_\theta (\nu)
\]

(Equation II-10)

It is important to note that the c-statistic of the three-band WMV must perform at least as well as the MV metric in predicting cardiovascular death in post-ACS patients. When \( \beta_1 = 0.30 \text{Hz}, \beta_2 = 0.55 \text{Hz}, \alpha_1 = 0, \alpha_2 = 1, \) and \( \alpha_3 = 0, \) the three-band WMV measure is exactly the same as the MV measure and thus should result in the same c-statistic when computed on the
same dataset as previous studies. Table II-3 shows the c-statistic results for the three-band WMV metric on the training and test data. On both data sets, the c-statistic was larger than 0.7, indicating good predictive performance.

<table>
<thead>
<tr>
<th>Data</th>
<th>( \beta_1 )</th>
<th>( B_2 )</th>
<th>( \alpha_1 )</th>
<th>( \alpha_2 )</th>
<th>( \alpha_3 )</th>
<th>c-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Data (MERLIN Placebo)</td>
<td>0.03</td>
<td>0.22</td>
<td>0.3</td>
<td>0</td>
<td>0.7</td>
<td>0.709</td>
</tr>
<tr>
<td>Test Data (DISPERSE)</td>
<td>0.03</td>
<td>0.22</td>
<td>0.3</td>
<td>0</td>
<td>0.7</td>
<td>0.781</td>
</tr>
</tbody>
</table>

Table II-3. WMV\((n = 3)\) Train vs. Test C-Statistic Results. The c-statistic resulting from the three-band WMV measure was greater than 0.7 on both the training and test data indicating good performance.

In the previous work done with MV, the upper quartile cutoff was determined to be optimal in risk stratifying patients [1818]. Similarly, for the three-band WMV metric, the quartile marker was used to determine the cutoff between high risk and low risk groups. The three-band WMV values of the all patients in the MERLIN Placebo data set were sorted in ascending order and the upper quartile cutoff was determined to be 253.8. Patients were dichotomized into low-risk (WMV\((n=3) < 253.8\)) and high risk (WMV\((n=3) \geq 253.8\)).

The hazard ratios at several points in time are shown in Table II-4 and Table II-5 for the training and test data. The test results in Table II-5 show that the MV metric gives a higher hazard ratio for 30 days following the start of the study while the three-band WMV metric outperforms the MV metric in the 90-day and 1-year hazard ratios. From the results, it seems that the WMV with \( n = 3 \) metric is on par with the MV metric performance but does not result in any major improvements.
Table II-4. *WMV(n = 3) vs. MV Training Results.* The three-band WMV metric was trained on the MERLIN Placebo data. The measure outperforms the MV measure with regards to the 1-year hazard ratio, however the MV metric performs better with respect to the 30-day hazard ratio and 90-day hazard ratios.

<table>
<thead>
<tr>
<th>Training Results</th>
<th>30-Day Hazard Ratio</th>
<th>90-Day Hazard Ratio</th>
<th>1-Year Hazard Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>WMV (n = 3)</em></td>
<td>3.92</td>
<td>5.03</td>
<td>3.67</td>
</tr>
<tr>
<td><em>MV</em></td>
<td>4.84</td>
<td>5.16</td>
<td>3.29</td>
</tr>
</tbody>
</table>

Table II-5. *WMV(n = 3) vs. MV Test Results.* The three-band WMV measure was tested on the DISPERSE-2 data. The measure outperforms the MV measure with regards to 90-day and 1-year hazard ratios, however the MV metric performs better with respect to the 30-day hazard ratio.

<table>
<thead>
<tr>
<th>Test Results</th>
<th>30-Day Hazard Ratio</th>
<th>90-Day Hazard Ratio</th>
<th>1-Year Hazard Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>WMV(n = 3)</em></td>
<td>10.00</td>
<td>9.42</td>
<td>9.42</td>
</tr>
<tr>
<td><em>MV</em></td>
<td>12.30</td>
<td>8.46</td>
<td>8.46</td>
</tr>
</tbody>
</table>
Chapter III: 
MACHINE LEARNING APPLICATIONS TO MORPHOLOGIC VARIABILITY

In this chapter, we tackle the binary classification problem of dichotomizing post-ACS patients into either a high-risk or low risk group through the use of support vector machines with morphology-based features.

A. Supervised Learning

Supervised learning infers input/output relationships from training data. Generally, the input/output pairings in the training data reflect an underlying functional relationship mapping of inputs to outputs and this functional relationship is the decision function for a supervised learning problem. The decision function, also known as the classifier, classifies inputs into certain output groups. When the learning problem has two output classifications, as is the one described in this thesis, it is referred to as a binary classification problem. It is important that the classifier not only perform well on that training data, but also generalize reasonably well to unseen data.

While supervised learning can be a powerful method of classification, there are a number of issues to consider, the first is the dimensionality of the feature vector. Typically, inputs are represented by a vector of features that are descriptive of the input. When the dimensionality of the feature vectors is large, the learning problem can become much more difficult. For example, if our feature vector is of length 1,000, we are trying to find a classification function that can separate out examples in 1,000-dimensional space. In this case, it is very difficult to find an accurate classification function which correctly maps the inputs to outputs in the training data.
Another important issue is the amount of training data needed to correctly infer a classification function. If the true classification function does not have a high degree of complexity, then it can be learned from a small amount of data. However, if the true classification function is highly complex, meaning it involves complex interactions among many different input features, a large amount of training data is needed to learn the function [26].

1. **Support Vector Machines**

Support vector machines (SVM’s) constitute a set of supervised learning methods used for classification. The learning machine is given a set of inputs, or training examples, with associated labels that are the output values. The input examples are in the form of feature vectors which are vectors containing a set of attributes of that particular input. Following this a classification function which optimally separates examples of different labels is chosen. Support vector machines incorporate several different types of classification functions or kernels, including linear kernels as well as non-linear functions such as radial basis kernels [27].

2. **Support Vector Machines with Linear Kernels**

The linear SVM is a method which represents examples as points in space mapped so that the examples that belong to separate categories are divided by a clear gap that is as wide as possible. As shown in Figure III-1, the positively labeled blue examples and the negatively labeled red examples are separated by a linear boundary. In addition, the two lines that border the decision boundary create a geometric margin which is the largest possible separation between the positive and negative examples. The goal of the linear SVM is to find the largest possible geometric margin that separates differently labeled examples.
The linear SVM is an optimization problem that is solved by directly maximizing the geometric margin. Many times, it is difficult to separate the labeled training examples using a linear SVM and thus slack is introduced into the optimization problem which allows for examples to fall within the margin. The linear support vector machine relaxed quadratic programming problem that needs to be solved is:

\[
[\theta^*, \theta_0^*] = \min_{\theta, \theta_0, \xi} \left( \frac{1}{2} \|\theta\|^2 + C \sum_{i=1}^{n} \xi_i \right) \quad \text{subject to} \\
y_i (\theta \cdot x_i + \theta_0) \geq 1 - \xi_i \quad \text{where } i = 1, ..., n \\
\xi_i \geq 0 \quad \text{where } i = 1, ..., n
\]

(Equation III-1)
The underscored variables in the optimization problem above indicate that they are vectors. The parameter $\theta$ is a vector that is normal to the decision boundary while the $\theta_0$ term is the offset parameter which makes the decision boundary more flexible as it no longer needs to pass through the origin. The C term in the problem represents the penalty that is given to examples that violate the margin constraint and the $\xi_i$ terms are the slack variables and correspond to the distance of the particular example from the margin to the decision boundary. Thus, if $\xi_i = 0$ the example lies on the margin, if $0 < \xi_i \leq 1$ the example lies within the margin but is correctly classified, and if $\xi_i > 1$ the example lies outside the margin and is misclassified. While the parameter vector, $\theta$, defines the location of the decision boundary, the magnitude of the parameter vector $\|\theta\|$ does not affect the location of the decision boundary and therefore it also does not affect the margin. We are allowed to use this extra degree of freedom to specify the margin in terms of this magnitude. The geometric margin is defined as $\gamma_g = \frac{1}{\|\theta\|}$, and thus, minimizing the $\frac{1}{2} \|\theta\|^2$ term is equivalent to maximizing the geometric margin. (Figure III-2).
Figure III-2. Linear SVM with Slack. The linear decision boundary is defined by an optimal parameter vector, \( \theta^* \), that is normal to the boundary and the optimal offset, \( \theta_0^* \), and these are found from the optimization problem defined in Equation III-1. The linear decision boundary labels the examples as either positive or negative depending on which side of the boundary they fall. Since slack is allowed in the linear SVM, examples are allowed to violate the margin constraint. These examples are the circled examples in the figure above. Image courtesy of Jaakkola [28].

When the penalty is very large, there are few violations and when the penalty is small, there are many violations of the margin constraint. Thus, overall the function aims to minimize the costs of violating constraints together with the geometric margin. The optimization problem specifies the trade-off between the size of the margin and margin violations. Thus, the final classification function is:

\[
 f(x; \theta, \theta_0) = \text{sign}(\theta \cdot x + \theta_0) \\
 = \begin{cases} 
 +1, & \text{if } (\theta \cdot x + \theta_0) > 0 \\
 -1, & \text{if } (\theta \cdot x + \theta_0) \leq 0 
\end{cases}
\]

(Equation III-2)
The final classification depends on which side of the decision boundary an example falls, as indicated by the \textit{sign} function. The function returns +1 if the example falls on the side of the decision boundary to which the parameter vector, $\theta$, points, and -1 if the examples falls on the other side of the decision boundary. The decision boundary essentially divides the space into a positive and negative half and this classification function assigns an example to one of those halves depending on which side of the decision boundary it falls [27].

The linear SVM with slack is beneficial because it is more robust against noisy examples as it draws the separation boundary as far as possible from the differently labeled training examples. In addition, it is advantageous because it is the simplest SVM model and requires the least amount of parameter variation thereby helping to prevent over-fitting of a model to the training data. Data correctly classified outside of the margin are classified with great confidence while the data within the margin are classified with little confidence or may be misclassified. Introducing slack allows violations of the margin constraint and sometimes even misclassifications. Thus, the penalty serves to prevent too many misclassifications [27].

3. \textit{Support Vector Machines with Non-Linear Kernels}

It is possible that examples are not linearly separable and thus a support vector machine with a non-linear kernel may be used (Figure III-2). The non-linear kernel serves to map the input examples to a higher dimensional feature space. The SVM classifiers remain linear in the parameters but perform non-linear operations in the original input space. Now, classifiers are of the form:
In Equation III-3, the \( \phi(x) \) function is the function that maps the input features to higher dimensional space.

One example of a non-linear kernel is the radial basis function (RBF) kernel, which introduces non-linear separation boundaries as it creates circles in the linear space to separate the data. Similar to the linear SVM, there is a penalty factor which relaxes the SVM's margin constraint and allows for violations of the margin. In addition to this parameter, there is a standard deviation parameter which specifies how quickly the kernel vanishes as the points move further away from each other.
The optimization problem needed to be solved for the SVM with an RBF kernel is the following:

$$\begin{align*}
\left[\theta^*, \theta_0^*\right] &= \min_{\theta, \theta_0, \xi_i} \left( \frac{1}{2} \|\theta\|^2 + C \sum_{i=1}^{n} \xi_i \right) \quad \text{subject to} \\
y_i (\theta \cdot x_i + \theta_0) &\geq 1 - \xi_i \quad \text{where } i = 1, \ldots, n \\
\xi_i &\geq 0 \quad \text{where } i = 1, \ldots, n
\end{align*}$$

(Equation III-4)

This optimization problem is typically solved in the dual form in which Lagrange multipliers are included and the constraints are encoded. The final classification function is:
\[ f(x; \theta, \theta_0) = \text{sign}(\theta \cdot \phi(x) + \theta_0) \Rightarrow f(x; \alpha) = \text{sign} \left( \sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + \theta_0 \right) \text{ where} \]

\[ K(x, x') = \exp \left( \frac{-1}{2\sigma^2} \|x - x'\|^2 \right) \]

(Equation III-5)

As with linear SVM's, the \textit{sign} function indicates which side of the decision boundary the example falls. The function returns a +1 if the example is on the side of the decision boundary to which the parameter vector, \( \theta \), points, and -1 if the examples is on the other side of the decision boundary. Ultimately, the max-margin linear separator is found in the feature space, resulting in a non-linear separation in the linear space as shown in Figure III-3. The \( K(x, x') \) term is the radial basis function (RBF) kernel which essentially forms Gaussians in the feature space with standard deviation \( \sigma \). The radial basis function is beneficial in that it allows for better classification of non-separable data, however it is more likely to over-fit the training data due to the addition of the standard deviation parameter [30].

\textbf{B. Linear SVMs with Morphology-Based Features}

A maximum margin linear classifier was used to separate patients into high risk and low risk groups where the feature vectors are obtained from the MD power spectra. As mentioned previously, there are 600 single frequency samples between 0.001-0.6Hz. However, a feature vector of length 600 is of relatively large dimensionality, thereby making the corresponding optimization problem difficult. Hence, we reduced the dimensionality of the feature space by dividing the power spectrum into \( n \) distinct equally sized bands that span the 600 frequencies. The power corresponding to each distinct band is then summed and these sums are used as the features. Thus, the feature space is reduced to the size 600/\( n \). In order to mitigate the effects of
outliers in the training data, we introduce slack into our optimization problem as described above. The values of the slack and penalty parameters where selected by optimizing the f-score which is a statistical measure of a classifiers accuracy and is defined as the harmonic mean of precision and recall:

\[
F_{\text{Score}} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}
\]  
(Equation III-6)

The precision is the fraction of correctly identified high risk patients out of all patients classified as high risk and the recall is the true positive rate or the fraction of correctly identified high risk patients out of all patients that are actually high risk. The f-score ranges from 0 to 1: an f-score of 0 means the classifier is terrible and no patients are correctly identified as high risk (making the precision and recall equal to 0) and an f-score of 1 means the classifier is great and all patients are correctly classified [31].

1. **Normalization of the Power Spectrum**

Before applying machine learning techniques to the power spectra data, it is important to normalize across frequencies so that each frequency contributes equally in determining the ultimate classifier. Since the power variation in the lower frequencies is much larger than the variation in the high frequencies, the classifier will tend to be dominated by these lower frequencies. When normalizing across frequencies, the lowest power value across all patients is set to -1 and the highest power value is set to +1, while all other values are scaled between those two bounds.
2. **Validation**

To develop and test the linear SVM classifiers, we used data from the MERLIN Placebo group which consists of post-ACS patients. In order to validate the results, the MERLIN Placebo group was randomly divided such that approximately 50% of the patients were in the test group and 50% of the patients were in the training group. In order to ensure balanced test and train sets, the deaths were randomly divided such that there were roughly equal number of patients who died in both the test and train groups. This random division of the MERLIN Placebo data into test/train sets was done 10 times. The linear SVM was developed from the training set and tested on the test set for each of these 10 trials to check for consistent results.

3. **Evaluation of Linear SVM’s with Morphology-Based Features**

When dividing the power spectrum frequencies into 60 bands, the f-score and 90-day hazard ratio results show that on average across the 10 test/train trials the linear SVM performs much better on the training data relative to the test data. This suggests that the linear SVM has over-fit to the training data (Figure III-4). The downside here is that since the linear classifiers developed from the training sets don’t perform very well on unseen data the classifiers are not generalizable. In addition, Figure III-5 shows the average f-score and 90-day hazard ratio values of MV applied to the test groups versus the average f-score and 90-day hazard ratio values of the linear SVM classifiers applied to the same test groups. On average, the linear SVM classifier does not perform as well as the morphologic variability metric based on f-score. Similar results were found when performing the same analysis after dividing the power spectrum 30 bands and even 10 bands as shown in the Appendix (A-1-A-4).
Figure III-5. Linear SVM Test vs. Train Results (60 Bands). The linear SVM classifiers developed from the training sets are over-fitting on the training data and performing poorly on the test data.
Figure III-6. *MV vs. Linear SVM Results (60 Bands)*. Based on both f-score and 90-day hazard ratios, the linear SVM applied to the power spectrum does not perform as well as the morphologic variability metric.
In addition to the linear SVM, SVM’s with RBF kernels were also developed. These SVM’s with RBF kernels did not perform as well as the linear SVM and over-fit to the training data even more so than the linear SVM. The reason for the decreased performance when using more complex kernels is most likely due to an increase in the number of parameters that need to be varied. In order to find the most optimal combination of parameters with the RBF kernel, not only does the penalty and cost factor need to be varied, but the standard deviation must be swept as well. This makes the classifier much more sensitive to the data.

There are several reasons why our implementation of SVM’s performed poorly with respect to classifying patients post-ACS using power spectra data. It may be that the entire power spectrum does not provide enough discriminatory power to differentiate high risk from low risk patients. There may be no underlying functional relationship between the power spectrum inputs and whether or not a post-ACS patient experiences cardiovascular death in the future. In addition, the classification problem is difficult to solve because the data is so imbalanced and there are many more non-deaths than deaths in the datasets.
Chapter IV: INCREASING THE CLINICAL UTILITY OF MORPHOLOGIC VARIABILITY

While morphologic variability has proven to be a successful metric for risk stratification of ACS patients for cardiovascular death, one of its major downsides is that it requires the collection of 24 hours of electrocardiographic data before risk assessment can be made. This chapter aims to see if we can increase the clinical utility of MV by decreasing the amount of data needed to identify post-ACS patients who are at high risk for cardiovascular death in the future.

A. Morphologic Variability Performance over Time

First, it is important to analyze the performance of MV over time. More specifically, is MV calculated from the first 10 hours after a patient is admitted to the hospital after an ACS more predictive than MV calculated from the last 10 hours in the day? The results from this analysis revealed that the 10-hour period at the start of a patients recording was more predictive than the last 10 hours based on the c-statistic. Figure IV-1 shows the c-statistics resulting from risk-stratification using the MV metric on different 10-hour intervals. The green points represent the start times of each 10-hour interval. The 10-hour intervals were slid by 5 minutes for each data point so a large change in the MV value is not expected. However, the plot shows that there is a relatively steady decrease in c-statistic as we look at later and later 10-hour intervals occurring in the first 24 hours of data.
C-Statistic values for MV calculated on 10-hour intervals of Data Across the First 24 Hours

Figure IV-1. MV Performance over the Time. The plot above shows the predictive results of calculating MV from 10-hour intervals of data. The green points represent the start of a particular 10-hour intervals and there is a difference of 5 minutes between adjacent 10-hour interval data points. The results show that the c-statistic decreases from the first 10-hour interval to the last 10-hour interval in the first 24 hours.

As a result, we conclude that data collected directly following an ACS incident is more predictive than data collected from later periods. After patients are admitted to the hospital following an acute coronary syndrome, they are given medications to quell their symptoms such as medicine to bring their heart rate back down to a normal speed. These medications cause slight changes to the physiologic signal as time goes on. Thus, data closer to the beginning of
the recording is the most informative because the information is not confounded with effects from these various medications.

B. Morphologic Variability Performance using less Data

1. 1st Hour Analysis

Currently MV provides an analysis of the risk level after 24 hours of data has been collected. However, prior experiments suggest that adequate risk stratification can be achieved with just 10 hours of Holter data. Nevertheless, the clinical utility of MV would be increased if the length of time required for accurate risk assessment could be further reduced. This section aims to create a modified MV metric that can be applied to 1 hour of data in the hopes of getting at least as good risk stratification as MV applied to 10 hours of data. MV is a function of the frequency range and the percentage threshold (90th percentile) which dictates which 5-minute interval is used in the computation. The final MV value for each patient is taken as the 90th percentile value sum across the frequency band from 0.3Hz to 0.55Hz over all 5-minute intervals or the 90th percentile value of the MV's for each patient as seen in Equation II-4. This particular frequency range is important in that prior results that involved changing the frequency range did not significantly alter the results. Thus, we are left with two aspects of MV that can be modified to create this new MV-1hr measure: the percentage threshold and the 5-minute intervals used.

a. Varied Thresholds

A patient’s MV is defined as the 90th percentile of the distribution of energies (within the pre-specified MV frequency band) across all 5-minute windows. One concern is that since there are only 12 (60min/5min) 5-minute windows that span the first hour of data, the 90th percentile
of this small set may not be optimal. To see what percentage threshold is optimal for this new
MV-1hr we calculated MV-1hr using each of the 12 possible thresholds.

To test the MV-1hr measure using varied thresholds, we used data from the MERLIN
Placebo group. There were several patients who did not have enough usable data in the first
hour. Because MV is calculated from such a large amount of data, if a patient has some unusable
data in the first 24 hours, it has a negligible effect on the MV measure. On the other hand, MV-
1hr is calculated using only a single hour of data and if a patient has some unusable data in the
first hour, it would have a much larger effect on the final MV-1hr measure. Thus, patients who
did not have 1 full hour of data after enrollment into the study were discarded. After these
patients were removed, 1589 MERLIN Placebo patients remained of which 69 experienced
cardiovascular deaths. The remaining patients were randomly divided such that approximately
50% of the patients were in the test group and 50% of the patients were in the training group and
roughly equal numbers of patients who died were in each group. This random division of the
MERLIN Placebo data into test/train sets was done 10 times.

Thresholds were varied from 1 to 12 corresponding to each 5-minute interval in the first
hour (0-5min, 5-10min… 55-60min). When varying the thresholds from 1 to 12, it was found
that using the 11th value (which corresponds to 50-55min) in the 0.3Hz-0.55Hz band resulted in
the highest c-statistic in a majority of the 10 trials. The c-statistic results from the 10 test/train
trials when using threshold 11 compared to the MV baseline measure (which is calculated on the
test groups) are shown in Figure IV-2. The plot shows similar results between the test and train
groups which means that the MV-1hr measure is not over-fitting the data. However, the MV-1hr
measure does not do as well as the MV measure applied to 10 hours of data. A paired t-test was
performed to check whether the difference between the c-statistics calculated from MV-1hr and
MV were significant. The t-test resulted in a p-value of 0.00093 which implies that those calculated from MV-1hr are significantly less than those calculated from MV.

Figure IV-2. MV-1hr using Threshold 11 vs. MV C-Statistic Results. The blue and green bars in the plot summarize the c-statistics when applying MV-1hr using threshold 11 to the 10 train and 10 test groups. The red bar summarizes the c-statistics when applying MV to 10 hours of data to each of the 10 test groups. The results show that MV performs significantly better than MV-1hr (p-value = 0.00093).

In addition to calculating MV-1hr by varying the thresholds from 1 to 12, we also tried combinations of averages to see if these resulted in higher c-statistics. Since the first hour data only contains 12 possible percentage thresholds to choose from, it is likely that there is lots of variation between each of these thresholds. In order to lessen this variation, we take combinations of averages. Results showed that averaging the 11th and 12th threshold values between 0.3Hz-0.55Hz resulted in the highest c-statistic. However, MV-1hr with averaged thresholds did not perform as well as MV-1hr using threshold 11. In addition, MV still performed significantly better than MV-1hr using average thresholds (p-value = 0.00056).
Figure IV-3. **MV-1hr Averaging Threshold 11 & 12 vs. MV C-Statistic Results.** The blue and green bars in the plot summarize the c-statistics when taking the MV-1hr value as the average of the 11th and 12th thresholds between 0.3Hz-0.55Hz in the 10 train and 10 test groups. The red bar summarizes the c-statistics when applying MV to 10 hours of data to each of the 10 test groups. The results show that MV performs significantly better than MV-1hr (p-value = 0.00056).

<table>
<thead>
<tr>
<th></th>
<th>Average c-statistic across 10 Test Trials</th>
<th>Standard Deviation of C-Statistics across 10 Test Trials</th>
<th>P-value (t-test between test results &amp; MV baseline)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MV-1hr (Threshold 11)</strong></td>
<td>0.660</td>
<td>0.032</td>
<td>0.00093</td>
</tr>
<tr>
<td><strong>MV-1hr (Average Threshold 11 &amp; 12)</strong></td>
<td>0.658</td>
<td>0.035</td>
<td>0.00056</td>
</tr>
<tr>
<td><strong>MV 10 Hour Baseline</strong></td>
<td>0.711</td>
<td>0.044</td>
<td>---</td>
</tr>
</tbody>
</table>

Table IV-1. **MV-1hr with Varying Thresholds vs. MV.** This table shows the averages and standard deviations of c-statistic across the 10 test trials for the MV-1hr metric and MV calculated from 10 hours of data. Both p-values are under the 0.05 threshold which signifies that the MV metric significantly outperforms the MV-1hr metric across the 10 trials in terms of c-statistic.

When looking at the 90-day hazard ratios when varying thresholds and also when varying average thresholds, we find that there is extreme variations in both the train and test groups. As
shown in Figure IV-4a, the standard deviation of the 90-day hazard ratios when using threshold 11 is extremely large for the test groups, indicating that there is lots of variation when using this metric on a single hour of data. The same can be seen when averaging thresholds 11 and 12 (Figure IV-4b). Even though both the test and train 90-day hazard ratio averages are larger than that for MV, there is so much variation, making the difference insignificant.
Figure IV-4. *MV-1hr vs. MV 90-day Hazard Ratio Results.* As seen in figures (a) and (b) above, even though the average 90-Day hazard ratios for the train and test groups when using threshold 11 and the average of threshold 11 & 12 are larger than that resulting from MV, the extreme variation in train and test results cause this difference to be insignificant.

**b. Sliding 5-Minute Windows**

Rather than change the threshold, another way to alter the MV-1hr metric to work with a smaller amount of data is to increase the number of possible thresholds in the first hour of data.
Instead of using adjacent 5-minute windows, we created sliding 5-minute intervals in the MD-time series and then converted these into the frequency domain. The 5-minute windows were slid in 1-minute increments thus creating 56 5-minute intervals in the first 1 hour of data. After the additional windows are created, the MV-1hr metric (with the 90\% threshold) was applied to the data. The main downside to creating more data points in this manner is that the 5-minute windows are overlapping and thus, we are re-using some of the same information in the extra data points. This MV-1hr metric was applied to the MERLIN Placebo data in the same manner described in section above. These results also showed that the MV-1hr metric does not perform as well as MV as shown in Figure IV-5. A paired t-test was applied to the c-statistic results from the MV baseline versus the MV-1hr test groups and it was found that the difference between the MV and MV-1hr c-statistics was significant (p-value = 0.0015).

Figure IV-5. MV-1hr with Sliding 5-min Windows vs. MV C-Statistic Results. The blue and green bars in the plot summarize the c-statistics when applying MV-1hr using sliding 5-minute windows and the 90\textsuperscript{th} percentile on the 10 train and 10 test groups. The red bar summarizes the c-statistics when applying MV to 10 hours of data to each of the 10 test groups. The results show that MV performs significantly better than MV-1hr (p-value = 0.0015).
Average c-statistic across 10 Test Trials | Standard Deviation of C-Statistics across 10 Test Trials | P-value (t-test between test results & MV baseline)
---|---|---
MV-1hr (Sliding 5-min Windows) | 0.661 | 0.038 | 0.00093
MV 10 Hour Baseline | 0.711 | 0.044 | ---

Table IV-2. MV-1hr with Sliding 5-min Windows vs. MV. This table shows the averages and standard deviations of c-statistic across the 10 test trials for the MV-1hr metric and MV calculated from 10 hours of data. The p-value is under the 0.05 threshold which signifies that the MV metric significantly outperforms the MV-1hr metric across the 10 trials in terms of c-statistic.

When looking at the 90-day hazard ratios for after using sliding 5-minute windows in our MV-1hr measure, we find that there is less extreme variations in both the train and test groups than there was when varying thresholds. As shown in Figure IV-6, the averages of the 90-day hazard ratios for both the test and train groups are less than that for MV applied to 10 hours of data. This difference is significant (p-value = 0.00030) indicating that once again, MV applied to a large amount of data outperforms MV-1hr applied to a single hour of data.
MV-1 hr Sliding 5-minute Windows
90-Day Hazard Ratio Results

avg = 2.66
stdev = 0.84

avg = 1.41
stdev = 0.53

avg = 1.17
stdev = 0.35

Train
Test
MV 10hr Baseline

Figure IV-6. MV-1hr with Sliding 5-min Windows vs. MV 90-day Hazard Ratio Results. The plot above shows that MV applied to 10 hours of data outperforms MV-1hr using sliding 5-minute windows in terms of 90-day hazard ratios (p-value = 0.00030).

c. Sliding 5-Minute Windows and Varying Thresholds

Results from MV-1hr when varying thresholds and when creating sliding 5-minute intervals were not as good as those from the MV baseline. The final step was to combine both of these modifications in the hopes of achieving a higher c-statistic. We created sliding 5-minute windows in 1-minute increments and then we varied the thresholds from 1 to 56. Results showed that using threshold 45 gave the highest c-statistic. While MV-1hr with sliding 5-min windows and a threshold of 45 did result in a c-statistic higher than varying the thresholds and creating the 5-minute windows separately, it still falls short of MV applied to 10 hours of data. As shown in Figure IV-7, the MV baseline resulted in a significantly higher c-statistics than this MV-1hr measure (p-value = 0.00083).
Figure IV-7. *MV-1hr with Sliding 5-min Windows and Threshold 45 vs. MV C-Statistic Results.* The blue and green bars in the plot summarize the c-statistics when applying MV-1hr using sliding 5-minute windows and 45th threshold from 0.3Hz-0.55Hz on the 10 train and 10 test groups. The red bar summarizes the c-statistics when applying MV to 10 hours of data to each of the 10 test groups. The results show that MV performs significantly better than MV-1hr (p-value = 0.00083).

<table>
<thead>
<tr>
<th>Metric</th>
<th>Average c-statistic across 10 Test Trials</th>
<th>Standard Deviation of C-Statistics across 10 Test Trials</th>
<th>P-value (t-test between test results &amp; MV baseline)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MV-1hr (Threshold 11)</strong></td>
<td>0.669</td>
<td>0.037</td>
<td>0.00083</td>
</tr>
<tr>
<td><strong>MV 10 Hour Baseline</strong></td>
<td>0.711</td>
<td>0.044</td>
<td>---</td>
</tr>
</tbody>
</table>

Table IV-3. *MV-1hr with Sliding 5-min Windows and Varied Thresholds vs. MV.* This table shows the averages and standard deviations of c-statistic across the 10 test trials for the MV-1hr metric and MV calculated from 10 hours of data. The p-value is under the 0.05 threshold which signifies that the MV metric significantly outperforms the MV-1hr metric across the 10 trials in terms of c-statistic.

When looking at the 90-day hazard ratios after using sliding 5-minute windows and threshold 45 we find that there is increased variations in both the train and test groups and on
average, the 90-day hazard ratios do not perform as well as those for MV applied to 10 hours of data. As shown in Figure IV-8, the averages of the 90-day hazard ratios for both the test and train groups are less than that for MV applied to 10 hours of data. This difference is significant (p-value = 0.00084) indicating that once again, MV applied to a large amount of data outperforms MV-1hr applied to a single hour of data.

**Figure IV-8. MV-1hr with Sliding 5-min Windows and Threshold 45 vs. MV 90-day Hazard Ratio Results.** The plot above shows that MV applied to 10 hours of data outperforms MV-1hr using sliding 5-minute windows and threshold 45 in terms of 90-day hazard ratios (p-value = 0.00084).
Chapter V: SUMMARY AND CONCLUSIONS

Patients who have experienced an acute coronary syndrome remain at risk for future adverse cardiovascular events, thus making risk stratification metrics extremely important. The availability and low cost of electrocardiograms make electrocardiographic risk stratification metrics extremely useful in identifying high risk patients. Morphologic variability (MV) is one such metric for risk stratification post-ACS. This thesis performed a comprehensive assessment of morphologic variability and aimed to modify the metric in order to increase its predictive capabilities and clinical utility.

A. Morphologic Variability

Morphologic variability (MV) is a powerful risk stratification metric that measures changes in the morphology of successive heart beats. Because damaged myocardial tissue does not conduct an electric signal the same way that undamaged tissue does, beat-to-beat changes in the waveform may indicate some underlying problem with the conduction path through the heart. In order to calculate MV, a morphologic distance (MD) time series is formed by performing the dynamic time warping (DTW) algorithm on successive beats. The power spectral density (PSD) is then formed from the MD time series and essentially converts the data into the frequency domain. The MV is quantified by taking the sum of the powers in the PSD across a specific frequency band. The MV metric was found to be highly correlated with the prediction of death in post-ACS patients based on hazard ratio results (Table II-1) however, because it only takes into account one specific band of frequencies, it misses out on relevant information that may exist in the other frequencies.
B. Weighted Morphologic Variability

Weighted morphologic variability (WMV) is an ECG-based risk stratification technique that is a variation on morphologic variability (MV). While MV is quantified by the sum of powers over one particular frequency band, the WMV metric is quantified by the weighted sum of powers over \( n \) frequency bands. As in the MV calculation, WMV uses the dynamic time warping algorithm to quantify beat-to-beat changes in the morphology of the ECG which leads to the morphologic distance (MD) time series. The MD time series then converted to the frequency domain into the power spectrum. WMV with \( n = 3 \) also known as three-band WMV was analyzed in this thesis. The three-band WMV measure is the weighted sum of powers over three frequency bands within the power spectral density of the MD time series.

In order to identify optimal cutoffs for each of the three bands as well as their associated weights, and exhaustive search method was used to find which combinations of weights and cutoffs resulted in the highest c-statistic. The c-statistic describes the probability of correctly identifying the patient with a disease given two patients, where one has the disease and the other does not. When evaluated on the DISPERSE-2 data set, three-band WMV had a c-statistic of 0.781 which was slightly better than the MV metric which had a c-statistic of 0.771. In order to evaluate cardiovascular performance, a quartile cutoff was used to dichotomize between patients at high risk and low risk. A hazard ratio, which captured death rates of high risk individuals compared to low risk individuals, was then calculated. The three-band WMV metric outperformed MV in the 90-day and 1-year hazard ratios but not the 30-day hazard ratio. From the results, it seems that the three-band metric does not result in any major improvements in
performance over the MV metric. This shows that the frequency band used to quantify the MV metric is quite important.

The goal in testing the three-band WMV metric was to see if we could achieve better risk stratification when segmenting the power spectrum into multiple bands. While the MV metric is defined as the sum of powers over a single band within the power spectrum, the modified version of MV presented in this thesis had three bands in order to create a metric that took advantage of information in other frequencies but at the same time, did not over-fit on the training data from which the metric was developed. The results of the three-band WMV analysis show that the metric does, in fact, outperform the existing MV metric.

There is evidence that the weighted frequency bands in the three-band WMV metric correspond to conduction path changes of the myocardium. Certain aberrant heart rhythms are found to appear as patterns in a fixed period of time. For example, in ventricular bigeminy and trigeminy, premature ventricular contractions (PVC) occur every second or third beat, respectively. It is possible that a group of myocardial cells spontaneously generate a wave of depolarization and then take longer than normal myocardial cells to repolarize. Thus, this group of cells can only depolarize once every $N$ heartbeats. At the times when this region of cells is electrically unexcitable, the wave of depolarization is deflected around this region, resulting in periodic conduction path changes. The morphologic difference captures differences between morphologies that result from the normal conduction path versus those that are associated with the abnormal conduction paths. This periodic change in the conduction path causes the MD time series to be periodic with period $N$ and causes the power spectrum to contain information about these morphology changes at frequencies $\text{heart rate} / N [6]$. 
The three-band WMV measure a weight of 0.7 on the frequency band 0.3-0.6Hz and a weight of 0.3 on the frequency band 0-0.03Hz. Information about anomalous beat that occurs once every three beats, such as ventricular trigeminy, would present itself in the frequency band 0.33Hz-0.56Hz (where normal heart rates range from 60-100bpm). This frequency range falls directly into the more heavily weighted upper band of the three-band WMV measure. In addition, because the three-band WMV measure places some weight on extremely low frequencies, it is likely that there is some abnormal phenomenon occurring in periods $N$ where $N$ is quite large, causing $\text{heart rate} / N$ to be arbitrarily close to 0.

Overall, three-band WMV is a strong metric that can be used to risk stratify post-ACS patients for cardiovascular death. Our data suggests that it is an efficient prognostication tool which is inexpensive and non-invasive. The metric easily allows clinicians to identify which patients may be at high risk and then provide them with the appropriate treatments to decrease their risk for the future adverse cardiovascular event. This simple, yet effective metric may help improve the care patients who are admitted to the hospital with an acute coronary syndrome.

C. Linear Support Vector Machines with Morphology-Based Features

This thesis also analyzed the ability of linear support vector machines to accurately separate post-ACS patients into high risk and low risk groups. As in MV, dynamic time warping algorithm was used to create the morphologic distance (MD) time series which captures beat-to-beat morphology changes and this was then converted to the frequency domain to create the power spectrum. Then a maximum-margin linear classifier was created using the MERLIN Placebo data set. The feature space for each patient was the 600 single frequency samples between 0.001Hz and 0.6Hz divided summed into different sized bands and each patient had an
associated label indicating that whether or not he/she died. The max-margin classifier aimed to find a line and associated margin that separated positively labeled patients who died from negatively labeled patients who did not die.

Support vector machines are generally thought to be superior to many other classification techniques. Generally, however, if the features of a certain dataset are strong predictors, any of the various classification methods perform well. In this analysis, it is likely that the feature space was not particularly strong thus causing the linear SVM to perform poorly in classification. Even after dividing the 600 frequency power spectrum into bands, thereby limiting the dimensionality of the feature space, it was difficult to find an appropriate linear classifier that separates the patients. In addition, because the data sets we worked with were so imbalanced (much fewer deaths than non-deaths), it was difficult to derive a strong classifier. Overall, linear support vector machines applied to the power spectrum performed worse than MV because the classifier tends to greatly over-fit the training data from which it was derived.

D. Increasing the Clinical Utility of Morphologic Variability

The final goal of this thesis was to see whether we could increase the clinical utility of MV. Currently MV needs at least 10 hours of electrocardiographic data to adequately risk stratify post-ACS patients. We aimed to see if we could modify the MV metric such that it could be applied to only a single hour of ECG data and still achieve as good if not better predictive ability as MV. We modified the MV metric in two ways to create an MV-1hr metric that could be applied to one hour of data: varying the thresholds and creating sliding 5-minute windows. MV is quantified as the 90th percentile sum of over the power spectrum across all 5-minute intervals. However, when working with a single hour there are only 12 5-minute windows, thus
the 90\textsuperscript{th} percentile sum may not be the most optimal threshold. In addition, we created sliding 5-minute windows by increments of 1 min, to create more data points in the first hour with the goal of better quantifying MV-1hr.

The results from this analysis show that based on the c-statistic, the MV measure applied to a large amount of data significantly outperforms the MV-1hr measure applied to only a single hour of ECG data based on c-statistic. In addition, there seems to be a considerable amount of variability in 90-day hazard ratios when applying the MV-1hr metric to a single hour of patient data. Thus, the more data that is included in risk assessment, the more robust and accurate predictions the metric gives.

MV is a powerful predictive metric that can be used in clinical settings to risk stratify patients for cardiovascular death. Because MV is an electrocardiogram-based risk metric, it can easily be implemented using data from an ECG-recording device and is easy to interpret as well as cost-effective. This thesis demonstrates that the MV metric outperforms more complex variations of MV, and works best when using large amounts of data. This simple metric should help improve the care for millions of patients who are admitted to the hospital every year following an acute coronary syndrome.
BIBLIOGRAPHY


doi: 10.1109/TBME.2004.827359


Figure A-1. Linear SVM (30 Band Power Spectrum) Test vs. Train F-Score and 90-Day Hazard Ratio Results. The linear SVM classifiers developed from the training sets are over-fitting on the training data and performing poorly on the test data.
Figure A-2. *MV vs. Linear SVM (30 Band Power Spectrum) F-Score and 90-day Hazard Ratio Results.* Based on both f-score and 90-day hazard ratio, the linear SVM applied to the power spectrum does not perform as well as the morphologic variability metric.
Figure A-3. Linear SVM (10 Band Power Spectrum) Test vs. Train F-Score and 90-Day Hazard Ratio Results. The linear SVM classifiers developed from the training sets are over-fitting on the training data and performing poorly on the test data.
Figure A-4. MV vs. Linear SVM (10 Band Power Spectrum) F-Score and 90-day Hazard Ratio Results. Based on both f-score and 90-day hazard ratio, the linear SVM applied to the power spectrum does not perform as well as the morphologic variability metric.