Risk Stratification of ICU Patients Using Arterial Blood Pressure Waveforms

by

Mathura J. Sridharan
S.B., Massachusetts Institute of Technology (2012)

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 24, 2013

Massachusetts Institute of Technology

All rights reserved.

Author: Department of Electrical Engineering and Computer Science
May 24, 2013

Certified By:

Collin M. Stultz MD, PhD
Associate Professor of Electrical Engineering and Computer Science
Associate Professor Harvard-MIT Division of Health Sciences and Technology
Institute for Medical Engineering and Sciences (IMES)
Massachusetts Institute of Technology
May 24, 2013

Accepted By:

Dr. Dennis Freeman
Chairman, Masters of Engineering Thesis Committee
May 24, 2013
Risk Stratification of ICU Patients Using Arterial Blood Pressure Waveforms

by

Mathura J. Sridharan
S.B., Massachusetts Institute of Technology (2012)

Submitted to the Department of Electrical Engineering and Computer Science

May 24, 2013

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

ABSTRACT

Identifying patients at high risk for adverse events is very important to the practice of clinical medicine. Non-invasive ECG-based methods of risk stratification such as T wave Alterans, Morphological Variability, and Heart Rate Variability extract prognostic information from the electrocardiograph. However, there is still a wealth of data collected from ICU patients and left unused every year that can augment risk-stratification methods. This thesis extends non-invasive risk stratification to Arterial Blood Pressure (ABP) Waveforms. We derive and analyze classifiers based on the morphological distance time series (derived from beat-to-beat morphology changes in the ABP waveform) including $ASDNN_{md}$, $SDANN_{md}$, $rMSSD_{md}$, the MV$_{ABP}$ score etc. We also derive and analyze classifiers based on the Downstroke Time Series (derived from the decay from peak systole to diastole) including $ASDNN_{Downstroke}$, $SDANN_{Downstroke}$, $rMSSD_{Downstroke}$, etc. While this body of work suggests the classifiers we developed are not effective in risk stratification of ICU patients, we discuss other methods which may extract prognostic information from the ABP waveform more effectively.
Acknowledgements

I would like to express my sincerest thanks to Prof. Collin Stultz who first introduced me to computational biology as my professor in 6.022. It is quite a statement to say that everything I know in cardiology I have learnt the past two and a half years from class and discussion with Prof. Stultz. He has been an integral part of my MIT undergraduate and graduate education as a professor, as a mentor, and as a PI and for that, I am very grateful.

I also would like to Jenna Wiens for giving me my first introduction to Morphological Variability and providing support with the initial stages of coding. I want to give special thanks to Yun Liu who generously imparted his wisdom having been working with risk stratification longer than I have. I also want to extend a thanks to all my lab friends Orly Ullman, Thomas Gurry, Gordon Lu, and Elke Socher for providing me with valuable feedback and criticism during lab meetings, helping shape my ideas, and being fun, supportive lab-mates through this entire year. I also want to thank Ramkumar Venkateswaran for going into lab at random times to see if the code finished running, pulling data for me at moment's notice, and being a great friend for the last four years.

Finally, I want to thank all my friends and family who pulled me up when I was down and gave me the confidence and encouragement to pursue my dreams. Thank you all.
Contents

Abstract 2
Acknowledgements 3
Introduction 9
Background 11
  Cardiovascular Physiology 11
  The Windkessel Model 13
  Cardiac Electrophysiology 18
  Arterial Blood Pressure 20
  Arterial Blood Pressure Waveform and the Intensive Care Unit 22
Risk Stratification Measures 23
  Invasive Risk Stratification Methods 23
  Non-Invasive Non-ECG Risk Stratification 24
  Non-Invasive ECG-Based Risk Stratification 25
Methods of Risk Stratification 27
  Heart Rate Variability of an Electrocardiogram 28
    Instantaneous Heart Rate Time Series 28
    Time Domain Metrics in Heart Rate Variability 30
    Frequency Domain Analysis of Heart Rate Variability 33
Methods 37
  Preprocessing the Arterial Blood Pressure Waveform 41
  Morphological Distance Time Series 41
  Time Domain Analysis of the Morphological Distance Time Series 44
  Frequency Domain Analysis of the Morphological Distance Time Series 49
  Downstroke Time Series 51
  Time Domain Analysis of the Downstroke Time Series 53
Methods of Statistical Analysis 55
  Defining an 'Adverse Event' 55
  Cross Validation 55
  Receiver Operating Characteristic (ROC) Curves 57
List of Figures and Tables

Figure 1: Physiology of the Cardiovascular System .................................................................12

Figure 2: Systemic Circulation ..............................................................................................13

Figure 3: Windkessel Simplification of Peripheral Circulation ..............................................14

Figure 4: Circuit model of peripheral circulation ....................................................................15

Figure 5: Windkessel Model ....................................................................................................16

Figure 6: ECG Tracing .............................................................................................................19

Figure 7: Arterial Blood Pressure Waveform ..........................................................................21

Figure 8: NN Interval ...............................................................................................................29

Figure 9: Instantaneous Heart Rate Time Series ....................................................................29

Figure 10: Segmented Instantaneous Heart Rate Time Series ..............................................31

Figure 11: Bandpass Filter for Noise Removal .......................................................................38

Figure 12: Baseline Wander Removal ....................................................................................39

Figure 13: Artifact Removal ..................................................................................................41

Figure 14: ABP Alignment by Dynamic Time Warping .........................................................42

Figure 15: Morphological Distance Time Series (ABP) .........................................................43

Figure 16: Distribution of Morphological Distance Time Series ...........................................46

Figure 17: Difference in Morphological Distance Time Series .............................................48

Figure 18: Frequency Spectrum of Morphological Distance Time Series ...........................50

Figure 19: Linear Slope of Downstroke ..................................................................................52

Figure 20: Downstroke Time Series ......................................................................................53

Figure 21: Receiver Operating Characteristic .......................................................................57
Figure 22: Statistical Methodology for Time Domain Metrics ...................................................... 60

Figure 23: Statistical Methodology for Frequency Domain Metrics ........................................... 61

Figure 24: Data ........................................................................................................................................ 65

Figure 25: Morphological Distance Time Series Time Domain Results ....................................... 69

Figure 26: Downstroke Time Series Time Domain Results ............................................................... 72

Figure 27: Optimal Diagnostic Frequency Grid .............................................................................. 74

Figure 28: Average beat-to-every-other-beat Distances ................................................................... 78

Figure 29: Δb Time Series .................................................................................................................... 79

Figure 30: Linear Slope of Upstroke ................................................................................................... 83

Figure 31: Upstroke Time Series ....................................................................................................... 83

Figure D.1: ECG Alignment by Dynamic Time Warping .............................................................. 105

Table 1: Electrode Labels and Placement ......................................................................................... 18

Table 2: Morphological Distance Time Series Results (Holdout Set) ............................................. 70
Introduction

Identifying patients at high risk for adverse events is very important to the practice of clinical medicine. A number of computational techniques have been developed to identify high risk patients; however, their performance has been far from perfect. Computational methods have since been studied and have revealed the ability to extract valuable information from biological signals such as electrocardiograms (which contains information on the condition of the patient’s heart). Moreover, computational methods pioneered by researchers show that subtle variations in the beat-to-beat morphology of the electrocardiogram may improve one’s ability to identify high risk patients following acute coronary syndromes. I intend to extend these computational methods beyond the use of just electrocardiographic waveforms to arterial blood pressure waveforms (which contain additional information on the condition of vasculature and peripheral circulation), and develop an algorithm physicians can use to identify patients in the intensive care unit (ICU) who are at high risk of adverse events.

The main hypothesis underlying our work is that Arterial Blood Pressure waveforms contain important prognostic information that can be used to identify patients at high risk for death after admission to the ICU. Therefore, we have developed methods that extract prognostic information from the arterial blood pressure (ABP) waveforms of patients admitted to the ICU. Changes in the beat to beat morphology of the arterial blood pressure waveform can hold prognostic information beyond electrophysiological information gleaned from electrocardiograms such as the changes in the compliance and resistance of the arterial system,
stroke volume, and sympathetic tone. These factors will consequently manifest themselves as changes in the beat to beat morphology of the arterial blood pressure waveform.

In this paper, we will explore methods to process continuous arterial blood pressure waveforms into two different discretized time series: one that captures the beat-to-beat differences in morphology and one that quantifies the decay of each beat from peak systole to diastole. Such changes may encode information about the peripheral vascular resistance and the arterial compliance. We will thus capture subtle changes in the morphology of adjacent beats in the arterial blood pressure waveform that are difficult to detect by visual inspection alone. We wish to determine if these changes enable us to improve our ability to identify high risk patient subgroups.

Unlike prior studies which are exclusively based on patients with acute coronary syndromes, we have expanded our study to include patients in several different ICUs including medical, surgery, and coronary care ICUs. We are therefore implicitly postulating that the metrics derived from the arterial blood pressure waveform are not restrained to only predicting coronary death but all-cause death in physiologically compromised patients.
Background

**Cardiovascular Physiology**

The circulatory system distributes and transports essential substances such as oxygen, nutrients, and hormones while removing byproducts of metabolism from tissues through a heterogeneous fluid, blood. It also participates in regulation of body temperature, maintenance of fluid, adjustment of oxygen and nutrient supply, and other homeostatic mechanisms [3]. The system consists of the heart, which serves as the pump, and the blood vessels which collect and distribute blood to various tissues as well as engage in exchange of materials (at the level of the capillaries or the smallest category of blood vessels approximately 5-10 μm in diameter). This system is collectively known as the cardiovascular system.

The heart, which functions as the pump for the entire circulatory system, is primarily made of cardiac muscle or myocardium and is sandwiched by the epicardium (which is contact with the pericardium or protective sac covering the entire heart) and the endocardium which is in contact with the blood. Electrically stimulated, coordinated contractions of the myocardium pump blood to the pulmonary and systemic circulations.

The heart is composed of four chambers: the left and right atrium and the left and right ventricles (Figure 1). Blood is ejected by ventricular contraction to both the pulmonary and systemic systems. Blood flows from the right atrium to the right ventricle followed by the pulmonary arteries and then the lungs. There, it is oxygenated and returns to the left atrium via the pulmonary veins. Oxygenated blood flows from the left atrium to the left ventricle and is pumped out of the aorta to the rest of the body (and delivered to tissue through a branching
system of blood vessels). The deoxygenated blood returns to the right atrium through the superior and inferior vena cava (Figure 1). Blood flow is unidirectional and achieved through the appropriate arrangement of valves. Further, blood leaving the left ventricle is pumped at a higher pressure.

Figure 1: Diagrammatic representation of the heart, pulmonary circulation, and systemic circulation. Vessels carrying oxygenated blood are represented in red and those carrying deoxygenated blood in blue (taken from [1]).
pressure than blood ejected from the right ventricle to the pulmonary system. The muscle making up the left ventricle is thus, thicker in order to achieve higher pumping pressure.

The blood vessels play a role in the exchange of water, oxygen, carbon dioxide, nutrients, and metabolic waste. The vascular system is like a tree beginning at the aorta and branching into progressively smaller (in diameter) branches until it ends in numerous vessels of 5-10 μm in diameter called capillaries (Figure 2).

**Figure 2:** Systemic Circulation: The aorta branches into progressively smaller blood vessels ending in capillaries where exchange of oxygen, carbon dioxide, and other chemical substances occurs. The blood is brought back to the heart through a similar branching system of veins and venules that bring blood back to the heart (taken from [2]).

As blood exits the heart into the arterial system, it is very fast and slows as it enters the periphery. As the surface area of the branches increase, the blood flow and pressure consequently decrease until blood flow is slowest with lowest mean pressure in the capillaries.
Moreover, it is not just the diameter of the blood vessels that differ amongst the aorta, peripheral arteries, arterioles and capillaries. The walls of the blood vessels become progressively thinner as it branches further into tissue and away from the heart and histological differences that aid in the different functions that the vessels serve in the systemic circulation [3].

**The Windkessel Model**

*...and Implications on the Morphology of the Arterial Blood Pressure Waveform*

Originally proposed by the German physiologist, Otto Frank [4], the Windkessel Model reduces our model of peripheral circulation to a lumped parameter model [5]:

\[
Q \quad R_a \quad Pa \quad \frac{1}{C_a} \quad R_a
\]

**Figure 3:** the Windkessel simplification of the peripheral circulation where \( Q \) is blood flow through the peripheral system, \( R_a \) is resistance of the peripheral system, \( C_a \) is arterial capacitance (or compliance) and we observe \( Pa \), the pressure drop across the peripheral system.

where \( Q \) is the blood flow through the peripheral system, \( R_a \) is the resistance of the peripheral system, \( C_a \) is the arterial capacitance (or compliance) and \( P_a \) is the pressure drop across the peripheral system. We then model this system as if it were driven by the heart, which is simplified as a periodic impulse generator that ejects a stroke volume of \( \Delta V \) at each impulse. The resulting model of the peripheral system is shown in Figure 4[5].
If we assume that there is no volume in the capacitance vessels at time $t = 0$, we can solve this system for some key parameters (Figure 5). At each stroke, the stroke volume will be deposited into the arteries resulting in an increase in the arterial pressure of magnitude $\Delta V / C_a$. The pressure then decays exponentially, as the volume exits the capacitor into the peripheral resistance, with a time constant, $\tau$, where $\tau = R_a C_a$. The model thus shows the mechanism by

$$
\text{Area of impulse} = \Delta V
$$

---

**Figure 4:** (Top) Circuit model of peripheral circulation driven by periodic impulse generator. The impulse generator (Bottom), $Q_0(t)$, is a simplification of the periodic ejection from the heart, is modeled as a series of impulses of area $\Delta V$ corresponding to an ejection of the stroke volume at each cycle.

---

$^1$ While it is assumed, for simplicity, that at $t=0$, the capacitance vessels are empty, this is not physiologically reasonable. Therefore, the Windkessel model is realistically applicable only at steady state.
which the systolic and diastolic pressure is obtained and quantified as a function of the stroke volume, arterial compliance, and peripheral resistance [5].

At equilibrium (denoted by the black dashed lines in Figure 5), the volume deposited

\[ T = \frac{R_a C_a}{1} \]

\[ \frac{1}{C_a} \]

\[ \tau = R_a C_a \]

\[ \Delta V/ C_a \]

\[ 0 \quad T \quad 2T \quad 3T \quad ... \]

**Figure 5:** Pressure waveform modeled by the Windkessel model. The model predicts a pressure buildup (denoted by the black dashed line above) that reaches steady state that varies between the systolic and diastolic pressures (denoted by the red dashed line above). The difference in pressure between the systolic and diastolic pressures is \( \Delta V/ C_a \) and the pressure builds up to the systolic pressure at the beginning of the cardiac cycle and drops back to the diastolic pressure at the end with a decay rate, \( \tau \), where \( \tau \) is directly proportional to both the peripheral resistance and arterial compliance (\( \tau = R_a C_a \)).

by the capacitance per beat is equal to the volume leaving the capacitor and flowing into the peripheral resistance. The relationship between the steady state systolic pressure (\( P_s \)) at the peak pressure and steady state diastolic pressure (\( P_d \)) is:
\[ P_d = P_s e^{-\frac{\tau}{R_aC_a}} \]

Setting the change in pressure, \( \Delta P = P_s (1 - e^{-\frac{\tau}{R_aC_a}}) \), equal to \( \Delta V/ C_a \), we can develop the following equations for steady state systolic and steady state diastolic pressure [5]:

\[
P_s = \frac{1}{(1 - e^{-\frac{\tau}{R_aC_a}})} \cdot \frac{V_a}{C_a}
\]

\[
P_d = \frac{e^{-\frac{\tau}{R_aC_a}}}{(1 - e^{-\frac{\tau}{R_aC_a}})} \cdot \frac{V_a}{C_a}
\]

In this study, these steady state approximations are of interest as it changes in the morphology of the arterial blood pressure waveform is indicative of underlying changes in the stroke volume, arterial capacitance, and peripheral resistance. Further, it has been experimentally determined that the exponential decay during diastole is a realistic representation of the pressure in the human aorta. Thus, changes in the shape and duration of the decay is directly related to changes in the peripheral resistance and arterial compliance. An increase (or decrease) of magnitude \( x \) in either \( R_a \) or \( C_a \) is directly proportional to an increase (or decrease) of magnitude \( x \) in the time constant \( \tau \), which controls the duration (and consequently, shape) of the decay of the ABP waveform.
**Cardiac Electrophysiology**

Clinicians use an electrocardiogram (ECG or EKG), a non-invasive method, to visualize the electrical activity of the heart. The standard method by which to measure the electrical activity of the heart is to place 10 electrodes across the thorax to achieve a 12 lead recording (summarized in Table 1).

<table>
<thead>
<tr>
<th>Electrode Label</th>
<th>Electrode Placement</th>
</tr>
</thead>
<tbody>
<tr>
<td>RA</td>
<td>Right Arm</td>
</tr>
<tr>
<td>LA</td>
<td>Left Arm</td>
</tr>
<tr>
<td>RL</td>
<td>Right Leg, lateral calf muscle</td>
</tr>
<tr>
<td>LL</td>
<td>Left Leg, lateral calf muscle</td>
</tr>
<tr>
<td>V₁</td>
<td>Between ribs 4 and 5, to the right of the sternum</td>
</tr>
<tr>
<td>V₂</td>
<td>Between ribs 4 and 5, to the left of the sternum</td>
</tr>
<tr>
<td>V₃</td>
<td>Between leads V₂ and V₄</td>
</tr>
<tr>
<td>V₄</td>
<td>Between ribs 5 and 6, mid-clavicular line.</td>
</tr>
<tr>
<td>V₅</td>
<td>Horizontally even with V₄, in the left anterior axillary line.</td>
</tr>
<tr>
<td>V₆</td>
<td>Horizontally even with V₄ and V₅ in the midaxillary line.</td>
</tr>
</tbody>
</table>

Table 1: Electrode labels (as used in USA) and physical placement of electrode

These electrodes measure small electrical changes on the skin caused by the heart depolarizing for each heartbeat. The resultant tracing for each lead is a net vector projection of the heart's electrical activity onto each lead.
The SA node is responsible for initiating the action potential under normal conditions and can be seen in the lead II ECG tracings as the P wave that precedes the large QRS complex (Figure 6). The delay at the AV node is seen in a large part of the PR segment where there is a brief delay before the P wave and the QRS complex. As the electrical impulse spreads through the ventricular myocardium, the QRS complex itself is produced and can be seen as the large complex (Figure 6). Finally, repolarization of the ventricles is responsible for the T wave as seen in Figure 6.

**Figure 6:** ECG tracing of a normal beat as seen in a Lead II tracing. The individual segments corresponding to physiologically significant segments of the tracing are labeled as well. (Courtesy Anthony Atkielski, Public Domain)
Arterial Blood Pressure

Arterial blood pressure (ABP) is a primary vital sign that is defined as the pressure exerted on the walls of blood vessels. ABP varies over the heartbeat with a maximum systolic blood pressure, which occurs towards the end of cardiac cycle during ventricular contraction, followed by the minimum diastolic pressure, which occurs during the beginning of the cardiac cycle when the ventricles are still filled with blood. Mean arterial blood pressure drops as the blood flows into the peripheral vasculature until there is continuous flow across capillary beds (in a normal circulatory system) with low blood pressure.

The major physical factors that determine arterial blood pressure waveform is the arterial resistance, compliance, and the stroke volume [3]. Therefore, the arterial blood pressure waveform is a valuable signal that contains information on the mechanical condition of the heart (ability to pump blood effectively as measured by the cardiac output), the condition of the vasculature (as determined by the peripheral resistance and arterial compliance), functionality of mechanical parts of the heart (defects in the valves can be observed in the shape of the ABP waveform), and even nervous system regulation of the heart rate and vascular tone.
Figure 7: Normal arterial blood pressure waveform. Peak pressure occurs during the systolic phase of the heartbeat. The dicrotic notch corresponds to the closing of the aortic valve. The small peak following the dicrotic notch corresponds to contraction of the atrium.

Typically, Invasive methods are used to procure a continuous arterial blood pressure waveform tracing (Figure 7). The blood pressure waveform is obtained by placing a cannula in an artery, typically the femoral or radial arteries. The continuous waveform displays three key characteristics: the systolic peak, diastolic minimum, and dicrotic notch (Figure 7). The systolic peak occurs when the ventricles contract, thereby ejecting blood into the aorta. The diastolic minimum occurs prior to ventricular contraction. The dicrotic notch is caused by the sudden closing of the aortic valve following ejection of the blood from the ventricles. This interrupts the drop from the systolic peak and causes a momentary upward increase in the ABP waveform before the drop from the systolic peak resumes.
Arterial Blood Pressure Waveform and the Intensive Care Unit

The arterial blood pressure waveform is a key vital that is routinely recorded in the intensive care unit (ICU). Patients in the intensive care unit are generally high risk patients with need for constant monitoring and thus monitoring the ABP with an eye for subtle changes can provide invaluable information above just information gathered from the electrocardiogram on the cardiovascular system.

When a patient's blood pressure drops rapidly, there is usually an activation of the sympathetic nervous system that is evidenced by the increasing of the heart rate (both observed in the ECG and ABP) and the clamping down of blood vessels, resulting in changes to the arterial blood pressure waveform. Thus the ECG and ABP together give complementary information that may be of prognostic value.

Arterial blood pressure waveforms can also bring to light information that other vital signs cannot. During sepsis, secretions of vasoactive chemicals cause the blood vessels to dilate. Because this is not an electrophysiological change, it does not affect the ECG tracings. However, these changes cause the beat to beat morphology of the arterial blood pressure waveform to change thus allowing clinicians to obtain prognostic information on changes in the vasculature that may be indicative of a more dangerous underlying problem.
Risk Stratification Measures

Identifying patients at high risk of an adverse event following admission to the ICU is important for the appropriate allocation of resources in the ICU. Currently there are few methods to systematically determine which patients are at high risk for another adverse event. When a patient is admitted to the ICU, it is standard practice to regularly monitor the patient's ECG and ABP waveform as the state of the cardiovascular system is integral to predicting the onset of an adverse event. Therefore, we rely on these measures of the electrophysiological and mechanical health of the heart and vasculature to develop a globally applicable method for risk stratification. Existing methods of classifying high risk patients are used on a very specific subset of patients: ones who have suffered an acute coronary syndrome (ACS). These risk stratification metrics can be broadly classified into invasive and noninvasive procedures. The methods can further be classified into ECG and non-ECG based metrics.

Invasive Risk Stratification Methods

Cardiac catheterization is a common invasive procedure used for both diagnostic as well as interventional purposes. Sometimes called coronary angiography, as catheterization of the coronaries is most common, cardiac catheterization allows physicians to visualize blockages in the arteries. During coronary catheterization, the physician inserts a catheter (small, tube-like device) into the large arteries of the body until the tip reaches the opening of one of the coronary vessels. Then a radio-opaque contrast agent, is injected allowing a 3-5 second visibility via fluoroscopy. The primary diagnostic goal of this procedure is to determine vascular occlusions. Sometimes, this procedure is also used for interventional purposes in a procedure known as angioplasty.
Often, biochemical markers found in blood are also used to identify patients at high risk of adverse events following ACS. Standard markers for identification of patients at high-risk for death following acute coronary syndromes include biochemical markers such as cardiac specific troponins, C-reactive protein, and brain neurotic peptide. However, these markers are limited in their predictive power and do not identify all patients at high risk of death following acute cardiac syndromes. Further, using these biochemical markers in routine clinical practice is at best, unclear.

**Non-Invasive Non-ECG Risk Stratification**

The TIMI (Thrombolysis in Myocardial Infarction) risk score is one metric that is used to assess the risk of a patient having subsequent ischemic events and is often used as a prognostic tool by physicians for informed decision making. The score itself is determined by seven independent features which receive a score of 1 if the patient has that particular feature thus allowing the TIMI score to range from 0 to 7. The features are age greater than or equal to 65 years, presence of at least three risk factors for heart disease, ST segment deviation present in ECG, the chance of prior coronary stenosis being greater than 50%, at least two anginal episodes in the past 24 hours, elevated serum cardiac markers, and the use of aspirin in the past 7 days. A score of 0-2 is classified as low risk, 3-4 as intermediate risk, and 5-7 as high risk of subsequent ischemic events or death [6]. A major downfall of this metric is its inability to make finer risk stratifications that are clinically relevant.

Non-invasive, non-ECG metrics also include non-invasive imaging techniques such as cardiac magnetic resonance imaging (imaging through use of powerful magnetic fields), cardiac
ultrasound (imaging through echocardiography), and cardiac computed tomography (imaging through the use of x-rays). Similarly, cardiac MRI and cardiac computed tomography can be extended beyond simply imaging the heart organ itself, to measuring and monitoring total and regional myocardial blood flow [3].

Non-Invasive, ECG-Based Risk Stratification

Many non-invasive, ECG-based models, often combined with other risk stratification metrics discussed above, are used to improve the identification of a high risk patient. These computational measures can be computed from measurable waveforms such as data from a patient’s electrocardiogram (ECG). One such marker is Heart Rate Variability (HRV) which measures beat to beat changes in the heart rate using various computational metrics on the NN intervals (or beat to beat intervals) [7]. Similarly, Morphological Variability (MV), a measure developed by Zeeshan Syed et al [8], is determined by the high-frequency variability in the beat-to-beat morphological distance derived from a patient's ECG. Patients with high-frequency variability in a population are determined to be high-risk patients for death within 90 days following an ACS. Further, MV provides complementary information to existing non-ECG based risk metrics including age, the TIMI risk score, Left Ventricle Ejection Fraction, etc., this computational biomarker has been clearly associated with determining high risk patients.
Methods of Risk Stratification

As discussed earlier, the Windkessel model establishes a direct relationship between the vascular resistance/compliance and the shape of the ABP waveform. Further, we discussed the relationship between the compliance of large arteries, the elastic recoil of the arterioles and the shape of the waveform. Events such as hypotension lead to activation of the sympathetic nervous system and deactivation of the parasympathetic nervous system. This results in quantifiable changes to the arterial blood pressure waveform. Therefore, we will extract the change in beat-to-beat morphology and define metrics based on the resulting series that may help us risk stratify ICU patients.

As part of our study on prognostic information contained in the ABP waveform, we will be doing both time domain analysis as well as frequency domain analysis on changes in the beat-to-beat morphology of the ABP waveform.

A study of changes in beat-to-beat morphology of the electrocardiogram has already been conducted, and frequency domain analysis was used to develop a metric that can be used to stratify patients at high risk of death in 90 days following an acute coronary syndrome. This method, known as morphological variability (MV), will be modified and adapted to stratifying patients at risk for death in 90 days in the ICU from the arterial blood pressure waveform. Thus, we will be closely following methods used in the prior publication entitled “Relation of Death Within 90 Days of Non-ST-Elevation Acute Coronary Syndromes to Variability in Electrocardiographic Morphology” [9] including deriving the series of beat-to-beat changes in
the ABP waveform and performing frequency domain analysis on this time series. A detailed methodology of MV is given in Appendix D.

We will also be performing time domain analysis on the series of beat to beat changes in the ABP waveform. We will be adopting and modifying some metrics more commonly used in Heart Rate Variability (HRV) studies and adapting them for use in time domain analysis. While HRV relies on changes in the RR interval, we will show that these metrics can successfully be adapted to our analysis of the morphological variability of ABP waveforms with relevant physiological justification on the relevance of these derived metrics.

Heart Rate Variability of an Electrocardiogram

Heart Rate Variability is a risk metric developed in the late 60's and 70's following experimental evidence connecting increased sympathetic or reduced vagal activity to lethal arrhythmias [7]. Heart rate variability quantifies and analyzes the oscillation in the interval between consecutive heart beats with information gleaned from an electrocardiogram. Further, it measures the oscillations between consecutive instantaneous heart rates. The term 'Heart Rate Variability' is accepted to describe variations of both RR intervals and instantaneous heart rates.

Instantaneous Heart Rate Time Series

We will discuss some of the common metrics used to quantify the oscillations between both RR intervals and instantaneous heart rates (IHR). We must first derive the instantaneous
Figure 8: An NN interval is the interval between adjacent QRS complexes resulting from sinus node depolarizations. (ECG tracing courtesy of Anthony Atkielski, Public Domain)

heart rate time series from the electrocardiogram. In order to derive this series, we take the continuous electrocardiogram and identify the NN intervals (normal-normal intervals: intervals between adjacent QRS complexes that result from sinus node depolarizations) (Figure 8).

Figure 9: Depiction of a continuous ECG and corresponding instantaneous heart rate time series. Time between adjacent NN intervals is extrapolated to the number of beats per minute and plotted.
We find the time length of each NN interval and extrapolate that to the number of beats per
minute implied by the NN interval. We can then construct the entire time series of
instantaneous heart rates. This process is shown in Figure 9.

Another time series of interest is the differences in the IHR time series. This series is
constructed by finding the difference between adjacent points in the IHR time series and can be
interpreted either as the difference in instantaneous heart rates or the cycle length [7].

**Time Domain metrics in Heart Rate Variability**

Of the time-domain based statistical measures that can be derived from these series the
most obvious metric is the average heart rate which is simply the mean of the IHR time series.
The average heart rate is of interest in many heart studies linking elevated heart rates with
coronary atherosclerosis and cardiovascular (and all-cause, non-cardiovascular) mortality,
[10][11].

We can also derive more sophisticated, time domain based statistical metrics for each
IHR time series and differences in IHR time series that can be split into two groups: metrics
derived from the analysis of the total ECG recording and metrics derived from smaller segments
of the recording period. The metrics derived from smaller segments of the recording period can
be used to analyze the RR intervals changes during periods of varying activity (rest, exercise,
etc.). On a basic level, moreover, it can be used to summarize changes in the heart rate longer
than a single cardiac cycles such as sympathetic and parasympathetic responses which can
occur on the order minutes.
In addition to the average heart rate (the mean of the IHR time series), another metric derived from the analysis of the total ECG recordings is the standard deviation of the entire 24-hour recording which is referred to as $SDNN$ (standard deviation of the NN intervals). Here, it is important to note that the standardization of the length of the ECG recording is necessary to ensure cross-comparison of the SDNN intervals because the standard deviation is so closely dependent on the sample length [7].

**Figure 10:** Segmentation of the instantaneous heart rate time series and development of the *Means of 5-minute Windows in IHR time series*. The mean of each five minute window is shown with a dotted green line and the five minute segments are denoted by the red line dividing the time series into smaller time series.
The same metrics (mean and standard deviation) on the entire 24-hour time series can be computed on the difference in IHR time series. We further interpret the mean and standard deviation to be the mean and standard deviation of the first difference in IHR (so that a mean of 0 indicates on average, no difference in IHR across the entire time series and the standard deviation indicates the spread subtle changes of IHR across 24 hours). We also include, in our analysis of the entire 24-hour long ECG recording, the square root of the mean squared differences of successive NN intervals (rMSSD).

As mentioned earlier, the second set of metrics is based on segmenting the 24-hour ECG recording into smaller time segments and analyzing the summary statistics resulting from the segmented time series. We first segment the IHR time series into mini time series of 5 minutes in length. Then we find the mean of each mini time series for each five minutes which will result in a time series of the means. We then find the standard deviation of the time series which is called the standard deviation of the average of NN interval or, in short, SDANN. Similarly, we develop a series of standard deviations by finding the standard deviation of the IHR mini time series for each five minute window. We then find the average of the time series which is called the average of the standard deviations of NN interval or, in short, ASDNN. The process of deriving the segmented time series of averages for the five minute windows is shown in Figure 10.

Other commonly used metrics in HRV analysis include the NN50 and pNN50. NN50 is a count of the number of adjacent NN intervals that differ by more than 50ms in length over the 24 hour period. pNN50 finds the percentage of adjacent NN intervals that differ by more than
50ms in length over the 24 hour period. We will not be using either of these metrics in conjunction with our study of changes in beat morphology.

**Frequency Domain Analysis of Heart Rate Variability**

Another method of analyzing the instantaneous heart rate time series and the difference in IHR time series is to use perform analysis in the frequency domain rather than the time domain. In order to capture the information in the frequency domain, we turn to the power spectral density (PSD) which is the power of the signal as a function of frequency. There are two generally accepted methods of computing the PSD of the IHR and difference in IHR time series: parametric methods (which most often employ the Fast Fourier Transform algorithm) and non-parametric methods. While both give an almost equivalent result, both offer their distinct advantages and disadvantages. Non-parametric methods offer simpler algorithms and higher processing speed. Parametric methods offer smoother spectral components, easy post-processing of the spectrum, and robustness of estimation to sample size [7]. We can also derive more sophisticated, frequency domain based statistical metrics for each IHR time series and differences in IHR time series that can be split into two groups: metrics derived from the analysis of the total ECG recording and metrics derived from smaller segments of the recording period. Prior to doing so however, we must discuss statistical significance of various frequency spectra, their physiological significance, and how the information they provide varies depending on the length of the original time signal.

Short term ECG recordings (defined as a recording between 2-5 minutes in length) are split into three separate spectral regions: very low frequency (VLF) from less than .04Hz, low
frequency (LF) .04Hz to .15Hz, and high frequency (HF) .15Hz to .4Hz. The physiological significance of the VLF spectrum in such a short time window has no physiological significance and is generally uninformative [7]. The use of these short term ECG recordings is significant when we develop frequency domain based metrics by segmenting the IHR and difference in IHR time series into short time segments (a similar method to developing SDANN and ASDNN). However, the LF and HF spectral ranges represent the two branches of the autonomic nervous system (as determined experimentally).

Long term ECG recordings (defined as anything greater than 5 minutes in length) are split into four separate spectral regions: ultra-low frequency (ULF) which is less than .003Hz, very low frequency (VLF) from .003Hz to .04Hz, low frequency (LF) .04Hz to .15Hz, and high frequency (HF) .15Hz to .4Hz. For shorter term recording, the VLF has little to no physiological significance. Similarly, for long term recordings, the ULF and VLF provide very little physiological interpretation.

The LF and HF components have very specific and significant physiological interpretations. Clinical and experimental observations from autonomic maneuvers such as vagotomy, electrical vagal stimulation, and muscarinic receptor blockade have shown that efferent vagal activity is a major contributor to the HF component of the IHR time series [7]. Interpretation of the LF component has two separate schools of thought: one that believes it is purely influenced by sympathetic modulation and one that believes it includes elements of both sympathetic and parasympathetic modulation [7].
Some common metrics derived from the PSD of the entire 24 hour ECG recording include the raw measurement of the PSD in the ULF, VLF, LF, and HF spectral ranges. Further, we find the total power of the entire 24 hour recording which is interpreted as the variance of the NN intervals. More interesting, however is the segmentation of the 24 hour recording into recordings 5 minutes in length. Within each segment, we generally compute the energy in the VLF, LF, and HF frequencies as well as the total energy over the five minute period which is the variance of the NN intervals in the five minute segment. We also find LF\_norm (which is defined as the LF power in normalized units \( \frac{LF}{Total\ Power-VLF} \) * 100) and HF\_norm (which is defined as the HF power in normalized units \( \frac{HF}{Total\ Power-VLF} \) * 100. Finally, we define the ratio of the LF PSD to HF PSD which gives us an understanding of changes to the ratio of sympathetic activity (assuming the LF contains mostly elements of sympathetic modulation) to parasympathetic activity in the autonomous nervous system.
Methods

Studies such as Heart Rate Variability and Morphological Variability of the Electrocardiogram have shown that prognostic information on risk of a repeat adverse event can be derived from the basic ECG. We seek to derive additional information, which can be used in conjunction with the aforementioned computational metrics derived from the ECG, from the arterial blood pressure (ABP) waveform. The primary focus of our methods will be to extract information from the beat-to-beat differences in morphology in the ABP signal. Because this realm of study has not been explored, in this section we will present the methods we developed that have shown to be effective in processing the ABP waveform and deriving prognostic metrics based on the ABP waveform.

Preprocessing the Arterial Blood Pressure Waveform

In this section we will discuss methods we developed to preprocess an arterial blood pressure (ABP) waveform prior to the derivation of time and frequency domain metrics that identify high risk patients in the population. To be consistent with similar studies done with electrocardiograms, we only use ABP recording of at least 24 hours in length. We then perform an automated method of preprocessing the ABP signal in order to remove noise, ectopic beats, and artifacts thereby ensuring that the signal is of an appropriate quality from which we can glean prognostic information.

We first normalize the signal, remove baseline wander and high frequency noise, and remove artifacts in the raw signal. The baseline wander and higher frequency noise is first
eliminated using a simple bandpass filter (Figure 11) that only allows frequencies from 1.25 Hz to 33.75 Hz.

![Filter Impulse Response](image)

![Filter Frequency Response](image)

**Figure 11:** (Left) Impulse response of bandpass filter used in removing baseline wander and high frequency noise. (Right) Frequency response of bandpass filter used in removing baseline wander and high frequency noise.

Second, the signal is filtered using wavelet denoising, a method by which we use the mathematical construct, wavelet, to further reduce noise in the waveform. We then employ parabolic fitting to ensure that the blood pressure peaks are distinct and clearly indicated as well as to ensure that all missed peaks are detected as well. After detecting the peaks, we further remove any ectopic beats (or irregular beats in an otherwise normal rhythm) that typically represent non-physiological artifacts. Finally, the signal is de-trended (so that the signal is an average mean 0).
Figure 12: Arterial blood pressure signal pictured above has significant baseline wander. The corrected ABP waveform is shown below (clean signal). The baseline wander is removed, the signal has been detrended (mean 0) and normalized.
Figure 13: Demonstration of artifact removal. Above we see the raw signal with mechanical artifacts denoted with red arrows. Below we see the de-trended signal with the mechanical artifacts removed.
Morphological Distance Time Series

Following signal pre-processing, we are left with a de-trended, noise-reduced continuous recording of NN intervals for the ABP waveform. Our primary hypothesis is that the changes in shape of the ABP waveform from beat to beat can provide us with prognostic information as the upstroke provides information on contractility, and the decay provides information on peripheral resistance and arterial compliance. Changes in any of these factors (peripheral resistance, arterial compliance, and stroke volume) will have an effect on the shape of the beat in the ABP waveform and these changes can be gleaned by examining the differences in the shape of the beats.

In order to derive this prognostic information based on the change in morphology, or shape, of the arterial blood pressure waveform, we must quantify the beat-to-beat differences by defining and using beat-to-beat distances as a measure of changes to the beats morphology. We choose to use dynamic time warping (DTW), the same method used in deriving the morphological distance between adjacent beats in an ECG. This is because time length of each beat in the ABP waveform varies and significant features must first be aligned prior to computing the difference in order for the distance rendered to be relevant to changes in morphology.

Dynamic time warping (shown in Figure 14) aligns the significant features of the arterial blood pressure waveform (the upstroke corresponding to systole, peak, dicrotic notch, and downstroke corresponding to diastole). It is evident that using a simple Euclidean difference will result in a distance that does not encompass the true differences in morphology because it
does not find the distance between relevant features of the time-varying beats (Figure 14). We will now develop the algorithm

![ABP beat alignment by dynamic time warping](image)

**Figure 14:** ABP beat alignment by dynamic time warping. Beats on left is unaligned and represents a Euclidean alignment. Beat on right is aligned using DTW.

Given two beats, \( x_1 \) and \( x_2 \) of length \( l_1 \) and \( l_2 \), DTW will first produce an \([l_1 \times l_2]\) distance matrix, \( D \). In this matrix, each entry, \((i,j)\) represents the distance between \( x_1 \) and \( x_2 \). A particular path through this matrix is defined as:

\[
\varphi(k) = (\varphi_1(k), \varphi_2(k)) \quad 1 \leq k \leq K
\]

where \( \varphi_1 \) and \( \varphi_2 \) are row and column indices and \( K \) is the alignment length [8]. We then search for the minimum cost path using dynamic programming which corresponds to the optimal alignment of the beats. This optimal path, \( C(x_1, x_2) \) is defined as:

\[
C(x_1, x_2) = \min_{\varphi} C_{\varphi}(x_1, x_2)
\]

Where,

\[
\min_{\varphi} C_{\varphi}(x_1, x_2) = \sum_{k=1}^{K} d(x_1[\varphi_1(k)], x_2[\varphi_2(k)])
\]
Where

\[ d(p_1, p_2) = (p_1 - p_2)^2 \]

In the special case where two beats are exactly alike morphologically, we expect the morphological distance between the two beats to be 0. The result of applying this to an entire time series of beats is a time series of morphological distances. This derivation is also shown below in Figure 15.

Figure 15: (Above) Continuous arterial blood pressure waveform is discretized into the morphological distance time series (below). \( D_{x,y} \) represents the DTW distance between beat \( x \) (\( b_x \)) and beat \( y \) (\( b_y \)).
**Time Domain Analysis of the Morphological Distance Time Series**

We now analyze the morphological distance time series (MD time series) of the arterial blood pressure (ABP) waveform in the time domain. The motivation for this time domain study stems from heart rate variability studies, and we will be applying metrics derived in the original HRV studies to a different waveform: arterial blood pressure. First, we split our analysis into two separate methods: analysis of the entire 24 hour time series and analysis of smaller time segments (5 minutes in length) of the time series. Analysis of the entire 24 hour time series provides information on the distribution of the time series across the entire recording. Analysis of smaller segments, on the other hand, provides information on the variation of the time series over time periods longer than the time scale of a single beat.

First we examine the average morphological distance across the entire 24 hour time series or $\mu_{md}$. Similarly, we also examine the spread of the morphological distances across the entire 24 hour time series which we define as $\sigma_{md}$. In HRV, the mean value of the entire, 24-hour, instantaneous heart rate time series provided the average heart rate of the patient over 24 hours. The average morphological distance, $\mu_{md}$ does not have a standard physiological definition as the instantaneous heart rate does. It quantifies the average distance in shape between each beat for the particular patient over 24 hours. It is also hypothesized that a patient with a large average change in stroke volume, peripheral resistance, and arterial compliance (factors that affect the morphology of the ABP waveform) will have a significant deviation in mean morphological distance from the average morphological distance of 0 (which characterizes no change, on average, in beat-to-beat morphology). This is probably indicative
of the existence of sympathetic and parasympathetic stimulation over 24 hours as well as provides an indirect quantification of the average stimulation across the 24 hours.

Further, we hypothesize that the standard deviation of the 24 hour morphological distance time series will be able to identify when the level of changes to the stroke volume, arterial compliance, and peripheral resistance vary widely across the 24 hour time series. This is possibly due to varying levels of sympathetic and parasympathetic stimulation across the 24 hour time series. We also find and interpret the root mean squared differences in adjacent morphological distances over the 24 hour period \( rMSSD_{md} \).

We now address statistics derived from splitting the morphological distance time series into smaller, mini-series, computing statistics for the small time series, and developing a summary statistic that will define the variations of each mini time series over the 24 hours. For consistency with previous studies, we pick time segments of 5 minutes in length. We first cut the morphological distance time series into non-overlapping windows 5 minutes in length. We develop two metrics by segmenting this time series:

1. \( ASDNN_{md} \) which is the average of the standard deviation of morphological distances for windows 5 minutes in length. **

2. \( SDANN_{md} \) which is the standard deviation of the average of morphological distances windows 5 minutes in length. **

**Because the morphological distance time series is calculated from normal beats, \( NN \) is appended as a suffix to the metrics as it was done in the original HRV studies.
$ASDNN_{md}$ is calculated by finding the standard deviation of every 5 minute non-overlapping window in the 24 hour MD time series then taking the average of all the standard deviations for

\[ \text{Morphological Distance Time Series} \]

\[ \text{Frequency of Occurrence} \]

**Figure 16:** (Top) Morphological Distance Time Series. (Bottom) Derived histogram of morphological distances over 24 hour period. The mean and standard deviation of the distribution of morphological distances is shown on the histogram.
each five minute window. $SDANN_{mu}$ is calculated by finding the mean of every 5 minute non-overlapping window in the 24 hour MD time series then taking the standard deviation of all the means for each five minute window.

By segmenting the signal and re-summarizing the 24 hour recording from statistics of each five minute interval, we are effectively eliminating the noise that can obfuscate prognostic information that arises from combining an early portion of the signal with later portions of the signal which may be under different conditions. This is especially important when using a 24 hour recording where physiological conditions can vary a lot over 24 hours and therefore a pure average or standard deviation of the entire signal may include too much unrelated information to be useful. By measuring the spread of the averages in each window and the average of the spread in each window, SDANN and ASDNN respectively avoid summarizing the signal by considering averaging the entire signal or finding the standard deviation of the entire signal. SDANN and ASDNN are thus used to group information over short term segments (in this case, five minute windows) and therefore allow us to more reasonably compare segments early in the recording to segments late in the recording.
In analogy to the time domain analysis developed in Heart Rate Variability, we develop a time series of differences between adjacent points in the morphological distance time series (Figure 17). For instance, the first MD time series point $D_{1,2}$ is defined as the morphological distance between beats 1 and 2 and $D_{2,3}$ is defined as the morphological distance between beats 2 and 3 (where $D_{x,x+1}$ is the morphological distance between beats $x$ and $x+1$). The first point in the Difference in MD Time Series is the difference between $D_{2,3}$ and $D_{1,2}$, the second between $D_{3,4}$ and $D_{2,3}$ until for the final beat $n$, the corresponding point is defined as $D_{n-2,n-1}$ and $D_{n-1,n}$.

![Diagram of morphological distance and difference in morphological distance time series](image)

**Figure 17:** Derivation of the Difference in MD time series (below) from the MD time series (above).
The purpose of differencing the MD time series is to smooth out the spurious high frequency components of the MD time series that may obfuscate prognostic information. On the other hand, smoothing the signal using the differences also causes loss of information and may not necessarily perform better than the MD time series itself if prognostic information is contained in the higher frequency components.

Risk metrics we derive from the Difference in MD time series can also be split into two categories: those determined from the entire 24 hour recording, and those derived from segments of the 24 hour recording. The two metrics derived from the entire 24 hour recording are the $\mu_{md}^{dif}$ and $\sigma_{md}^{dif}$. We then segment the Difference in MD time series into non-overlapping windows of five minutes in length and compute the average of the Difference in MD time series in each window. The standard deviation of these averages is $SDANN_{md}^{dif}$. We also compute the standard deviation of the Difference in MD Time Series in each window. The average of these standard deviations is the $ASDNN_{md}^{dif}$.

**Frequency Domain Analysis of the Morphological Distance Time Series**

We now analyze the morphological distance time series (MD time series) of the arterial blood pressure (ABP) waveform in the time domain. Here we will be adopting and modifying methods used in Morphological Variability of ECG analysis as well as frequency domain metrics derived from Heart Rate Variability. We begin by once again segmenting the morphological distance time series into non-overlapping 5 minute windows.
We then derive the frequency spectrum of each five minute time series (Figure 18) using a Lomb-Scargle periodogram which is a semi-parametric spectral method of fitting sinusoids to discrete data using least-squares estimation.

![Morphological Distance Time Series](image)

**Figure 18**: Spectral energy as a function of frequency (above) of the morphological distance time series (above) from 5-10 minutes. The total energy within the experimentally determined diagnostic frequency range is considered the relevant prognostic energy value for that five minute window.

We find the total energy within a particular frequency band called the diagnostic frequency. This diagnostic frequency is experimentally determined and in the original Morphological Variability on ECG studies, was determined to be $0.3 \rightarrow 0.55$ Hz for
electrocardiograms of patients following acute coronary syndrome [9]. The optimal diagnostic frequency for ABP waveforms in the ICU population to be .35→.5Hz (experimental methods leading to this result will be discussed in the Results section).

Thus we convert each morphological distance time series into a series of energy values for each non-overlapping five minute window. The 90th percentile of all the energy values over the 24 hour recording is considered the MVABP score for the patient.

**Downstroke Time Series**

Another time series of interest, is one that captures only the joint changes in the peripheral resistance \( (R_a) \) and arterial compliance \( (C_a) \) of the signal from beat to beat. These changes are indicative of sympathetic and parasympathetic activity regulating peripheral vasculature and consequently, the arterial blood pressure. Because the down-slope from each systole to the next following diastole of the ABP waveform decays exponentially with a time constant proportional to \( R_a \cdot C_a \), the slope from each systole to the next diastole directly captures the rate of decay following systole and indirectly captures information on \( R_a \cdot C_a \). When we make a time series of down-slopes of systole to following diastole, we can identify variations in the rate of decay, the change in pressure over the change in time, over the entire 24 hour ABP recording. This time series, we call the Downstroke time series.

To develop the Downstroke time series, we first follow the same signal pre-processing methods used prior to developing the morphological distance time series (see Methods: Preprocessing the Arterial Blood Pressure Waveform). Following signal pre-processing, we are left with a de-trended, noise-reduced continuous recording of NN intervals for the ABP
waveform. Then, we identify each peak corresponding to the systole which occurs at \((t_s, P_s)\) and the following trough corresponding to the diastole which occurs at \((t_d, P_d)\) where \(t_d > t_s\). We then find the slope, \(m\), of the linear interpolation of \((t_s, P_s)\) and \((t_d, P_d)\):

\[
m = \frac{P_s - P_d}{t_s - t_d}
\]

**Figure 19:** Linear interpolation of the peak of systole and the trough corresponding to the following diastole shown for a single beat in the ABP waveform.
Figure 20: Derivation of the Downstroke time series which converts continuous ABP recording into the discretized time series of the linear slopes of decay between peak systole and following diastole.

We also develop the time series of differences between adjacent points in the Downstroke time series. This will gives us a time series of the differences in systole to diastole decay between adjacent beats. We name this series the Difference in Downstroke time series.

**Time Domain Analysis of the Downstroke Time Series**

We will use the same time domain metrics used to analyze the morphological distance time series but with different physiological interpretations. We first find the mean, $\mu_{\text{Downstroke}}$, and standard deviation, $\sigma_{\text{Downstroke}}$, of the Downstroke time series which gives us insight into the average slope of decay from systole to diastole as well as a measure of the deviation of the
decays from the average over 24 hours. It is postulated that a large variation is evidence of increased sympathetic and parasympathetic activity that actively regulate peripheral resistance and arterial compliance resulting in variation in the rates of decay from systole to diastole.

We also derive the Difference in Downstroke time series by finding the difference between each successive point in the Downstroke time series. Slightly more interesting is the mean and average of the Difference in Downstroke Time Series ($\mu_{\text{Downstroke}}^{\text{dif}}$ and $\sigma_{\text{Downstroke}}^{\text{dif}}$ respectively), because the Difference in Downstroke Time Series informs us on beat-to-beat change in decay rate and the mean and standard deviation of this series inform us about how the decay rates change over the 24 hour period. For an ideal person with perfectly identical decay from each beat to its following beat, the $\mu_{\text{Downstroke}}^{\text{dif}}$ will be 0. Further, the $\sigma_{\text{Downstroke}}^{\text{dif}}$ metric helps us understand if the changes themselves are consistent over time ($\sigma_{\text{Downstroke}}^{\text{dif}}$ is small) or inconsistent with large variation in the beat to beat changes themselves ($\sigma_{\text{Downstroke}}^{\text{dif}}$ is large). Large variations in the beat-to-beat changes is indicative of varying levels of sympathetic and parasympathetic stimulations whereas a small value in $\sigma_{\text{Downstroke}}^{\text{dif}}$ is more indicative of a constant level of sympathetic and parasympathetic stimulations.

We also develop the metrics developed from segmenting the time series into smaller windows, then finding a summary of the distributions of the time series across all windows in a 24 hour period. We use the same procedure as we used to analyze the MD time series and start by segmenting the Downstroke time series into non-overlapping windows five minute in length. We then compute the average the Downstroke time series in each window. The standard deviation of these averages is $SDANN_{\text{Downstroke}}$. We also compute the standard deviation of
the Downstroke time series in each window. The average of these standard deviations is the ASDNN\textsubscript{Downstroke}. We apply the same method to the Difference in Downstroke time series to obtain SDANN\textsubscript{Downstroke} and ASDNN\textsubscript{Downstroke} \textsuperscript{dif}. Finally, we compute the square root of the mean squared differences ($r$MSSD\textsubscript{Downstroke}) of adjacent points of the Downstroke time series.

**Methods of Statistical Analysis**

**Defining an 'Adverse Event'**

Because we are interested in risk metrics to determine patients at risk for an adverse event following admission to the ICU, we must strictly define an 'adverse event'. We define an adverse event to be the occurrence of death within 90 days of ICU admission. We label the patient records (with known follow-up death status) as 'dead at 90th day following admission' or 'not dead at 90th day following admission.'

**Cross-Validation**

Cross-validation is a technique used to assess how the results of a statistical analysis generalize when applied to an independent dataset. This model validation technique is often used when additional, independent data (most likely a separate dataset with no overlapping data points) is unavailable and we run the risk a of error generated by testing a hypothesis on a dataset using parameters implied by the same dataset). Cross-validation involves partitioning the dataset into two mutually exclusive datasets called the Training and Testing sets. Conventionally, the training set is larger than the testing set.
After partitioning the dataset, we apply the model to the training set and vary the adjustable parameters. We then find the optimal combination of adjustable parameters that produce the most significant statistical results. Then we use the fine-tuned model derived from the training set and apply it to the testing set. The resulting performance in the testing set is considered to be an approximation of the model's performance on an independent dataset.

k-fold cross validation involves partitioning the dataset into $k$ parts. Then, $k-1$ subsets will be retained as training sets and one subset will be held out as the testing set. Then, the procedure above is repeated $k$ times to estimate the distribution of the performance of the model on independent datasets.

In our analysis, we will be using 2-fold cross validation (also known as holdout validation because the testing set is held out in the first portion of the analysis). We randomly partition 67% of our viable patient records and assign them to the training set and the other 33% are assigned to the testing or holdout set. For the time domain metrics, we do not have adjustable parameters. Instead, we test the performance of the 9 metrics on the testing set, choose the most predictive metrics from those 9 metrics, and apply them to the testing set to determining their performance on an independent dataset. For the frequency domain metric, MV$_{ABP}$ scores, we have one adjustable parameter: the spectral range that contains the most prognostic information. We test many different spectral ranges and pick the optimal range which is the diagnostic frequency (more detailed approach in Results: Frequency Domain Metrics). We then apply this diagnostic frequency to compute the MV$_{ABP}$ scores for the patients in the testing set and analyze its effectiveness as predictor of a repeat adverse event.
Receiver Operating Characteristic (ROC) Curves

Because our risk metrics function as binary classifiers where a patient is defined as 'high risk' if they fall above some cutoff $Y$ and 'low risk' otherwise, we use a Receiver Operating Curve (ROC) to determine the performance of the metric. The ROC curve is derived by varying the discrimination threshold $Y$ and plotting the true positive rate (also called the sensitivity) against the false positive rate (also 1-specificity). We then integrate under this curve and the resulting area under the curve (AUC) is a measure of the effectiveness of the binary classifier. The AUC is also known as the c-statistic and is equal to the probability that the classifier will rank a randomly chose 'high risk' patient as 'high risk' and thus a higher AUC means a more effective

![ROC Curve Example](http://www.medcalc.org/manual/roc-curves.php)

**Figure 21:** Example ROC curve with the no-discrimination line depicted as a dotted line. Image courtesy of: [http://www.medcalc.org/manual/roc-curves.php](http://www.medcalc.org/manual/roc-curves.php)
binary classifier. The line where true positive rate = false positive rate (shown in Figure 21 as a dotted line) is called the no discrimination line, where the AUC = 0.5 which indicates that the binary classifier is completely un-predictive. When the AUC is less than 0.5, this means the inverse of the classifier (i.e. 'high risk' if they fall below some cutoff Y and 'low risk' otherwise) and the \( \text{AUC}_{\text{inverse classifier}} = 1 - \text{AUC}_{\text{classifier}} \).

For example, one metric we computed is \( \text{ASDNN}_{md} \) which is the average of the standard deviation of NN intervals (in 5 minute windows) for each 24 hour recording. We pick a cutoff \( Y_0 \) such that if the patient's \( \text{ASDNN}_{md} \) score is greater than \( Y_0 \), then he is considered 'high risk' and the patient's \( \text{ASDNN}_{md} \) score is less than \( Y_0 \), then he is considered 'low risk'. We compute the entire vector of predictions for each patient based on the cutoff \( Y_0 \).

We then compute the true positive rate (the number of patients correctly predicted to be high risk divided by the true number of patients that actually died in 90 days following admission to the ICU) and the false positive rate (the number of patients incorrectly predicted to be high risk divided by the true number of patients alive at the 90th day following admission to the ICU). We then vary the cutoff and repeat this process for every possible \( Y \) and plot the resultant true positive rate vs. false positive rate for each \( Y \). The area under this curve (AUC) then measures the probability that the metric predicts a 'high risk' patient as 'high risk'.

**Bootstrapping**

Bootstrapping is a technique used to assess the sampling distribution of a statistic using an approximating distribution. This is an important technique in assessing the distribution of a
statistic when the entire population cannot be sampled or known. In this case, we cannot possibly use the world's population of ICU patients and are limited to the population represented by the MIMIC II database from which we use patient records in the ICU. Therefore, we must use a method by which to measure the variance in the statistic used to analyze our classifiers' performance on the dataset. In this case, we are using the AUC as our statistic to analyze the classifier's performance in our training set and will use the bootstrapping technique to create a sampling distribution for the AUCs in order to extrapolate their performance to the overall population of ICU patients.

The general implementation of bootstrapping is by constructing a new dataset (of the size of the original dataset or smaller) by randomly sampling with replacement from the original dataset and compute the statistic used to analyze the new dataset. Then the process is repeated N times (where N is a large number, typically 1,000 or 10,000) in order to obtain a distribution of statistics from each one of these N population subsets.

Similarly, for a metric (we use ASDNN_{md} as an example again) we will first sample with replacement from the patients in the training set to construct a new training set. Then we classify each patient as either high risk or low risk over every possible cutoff and compute the resulting ROC curve and AUC from this set of patients. We repeat this process 100 times to get 100 AUCs from each derived training set. The resulting distribution of AUCs is used to analyze where the true mean for the world population of ICU patients truly lies and with what certainty we can say the AUC we determine for a metric is representative of the true AUC for the entire population of ICU patients.
Diagram of Methods Used in Statistical Analysis

Time Domain Metrics Statistical Methodology

Figure 22: Statistical methodology shown for time domain metrics. The dataset is partitioned into the screening set and testing set. From the screening set we derive 100 populations of the same size as the training set using random sampling with replacement. We then compute the AUC for each population and derive an estimated sampling distribution for AUCs. This process is repeated for all time domain metrics (classifiers) and estimated sampling distributions are found for each metric. We screen for the metrics for which the 95% confidence interval does not contain 0.5 and apply only those to the testing set.
Frequency Domain Metrics Statistical Methodology

Figure 23: Statistical methodology shown for frequency domain metrics. The dataset is partitioned into the training set and testing set. We compute the MV\textsubscript{ABP} scores for every possible diagnostic frequency range. We then compute the AUC for each set of MV\textsubscript{ABP} scores and the diagnostic frequency which produced the MV\textsubscript{ABP} scores with the highest AUC is considered the optimal diagnostic frequency. MV\textsubscript{ABP} scores for patients in the testing set are then computed using the optimal diagnostic frequency and the resulting is computed from using these scores as a classifier.
Data

The data used in our studies come from the MIMIC II waveform and clinical databases from the PhysioNet [12], a compilation of physiological signals (Physiobank) as well as related-open source software (Physio Toolkit). The MIMIC II waveform database specifically contains a collection of physiological signals collected from 2001-2008 from a variety of ICUs (in a single tertiary teaching hospital) including medical, surgery, coronary care, and neonatal ICU population [12]. However, we will be restricting our study to data collected from the medical, surgery, and coronary care units and will be disregarding neonatal waveforms.

Electrocardiographic signals included in patient records from the MIMIC II database are AVF, AVL, AVR, I, II, III, MCL, MCL1, V (unspecified precordial lead), V1, and V2 [13]. Blood pressure waveforms included in patient records from the MIMIC II database are arterial blood pressure recorded from one or both radial arteries (ABP and ART), pulmonary artery pressure (PAP), amongst other measures of blood pressure [13].

Many of the patient records from the waveform database are matched with the corresponding clinical records in the MIMIC II Clinical database. These records include [14]:

- General Information (admission & discharge dates, death dates, patient demographics etc.)
- Physiological Information (hourly vital sign metrics, SAPS, SOFA, ventilator settings)
- Medications Administered During ICU Admission
- Lab Tests (Hematology, Imaging, Chemistry, etc.)
Fluid Balance (inputs including solutions and blood, outputs including urine and estimated blood loss)

Notes (Cardiac Catheterization, Radiology, Discharge Summary, etc.)

Admission, discharge, and death records are de-identified for privacy using an arbitrary offset added to the year of admission, discharge, and death. Multiple admissions to the ICU are also noted in the patient clinical records [14].

We restricted our data to the ABP waveforms that are 24 hours or more in length from the MIMIC II Waveform Database with 125 Hz sampling rate. We matched clinical records, specifically the admission and death date, to the records in the waveform database with viable ABP waveforms in order to determine if the patient is alive at the 90th day following admission to ICU and acquisition of the arterial blood pressure waveform. In some of the extensions of our study (discussed in the Results section), we also used the clinical information on drugs administered during the ICU stay.

We began with a total of 6443 patient records of which only 1822 records had ABP records. After removing the patient records with ABP waveforms less than 24 hours in length, we removed the patient records with no follow-up information on death date of the patient. Because we have defined a repeat adverse event to be death within 90 days of ICU admission, we are interested in stratifying the population into those dead at the 90th day following ICU admission and those alive at the 90th day following admission. Of the remaining 1234 patients, 305 were dead at the 90th day following admission to the ICU and 929 were alive at the 90th
day following admission to the ICU (a 7.9% rate of death within 90 days). This reduction process is detailed in Figure 24 below.

Figure 24: Out of a total 6443 records, 1234 contain viable arterial blood pressure recordings with information on the death date of the patient. In the 1234 viable patient recording, 7.9% were dead at the 90th day following admission to the ICU.

As detailed in the Results section, 90 days is the arbitrary cutoff used as the time horizon for a recurring adverse event resulting in death. Therefore, the viable patient population is split into patients dead at 90 days following admission and patients alive at 90 days following admission to the ICU.
Results and Discussion

As mentioned in the Methods section, we define an adverse event to be the occurrence of death within 90 days of ICU admission. We first gleaned prognostic information from the morphological distance time series which measures subtle changes in beat to beat morphology across time and developed various metrics that can be used to predict a repeat adverse event. We use these various risk metrics in both the time domain ($SDANN_{md}$, $ASDNN_{md}$, etc.) as well as the frequency domain ($MV_{ABP}$ score, LF/HF energy ratio), to determine how well they are able to predict whether or not the patient is dead at the 90th day following admission to the ICU.

We further derived time domain metrics from the Downstroke time series (see Methods: Downstroke Time Series) which we assess in its ability to predict a repeat adverse event, death within 90 days following admission to the ICU.

Time Domain Metrics

Morphological Distance Time Series

From the MD time series and the Difference in MD time series, we derived 9 metrics:

\[ \mu_{md}, \sigma_{md}, ASDNN_{md}, SDANN_{md}, rMSSD_{md}, \mu_{md}^{diff}, \sigma_{md}^{diff}, ASDNN_{md}^{diff}, SDANN_{md}^{diff} \]

We develop Receiver Operating Curves (see Methods: Methods of Statistical Analysis) for each metric, $m$, by defining every possible cutoff $Y$ and defining patients with $m \geq Y$ to be at high risk for death in 90 days and $m < Y$ to be low risk for death in 90 days. We plot the ratio of false positives to true positives at every possible cutoff $Y$ by comparing against the true labels for death at day 90. We define the area under this curve (AUC) to be the probability that this metric will rank a randomly chosen patient at risk for a repeat adverse event higher than a randomly chosen
patient not at risk for a repeat adverse event. Finally, we will use the bootstrapping method (see Methods: Methods of Statistical Analysis) to estimate the sampling distribution for the AUCs. Metrics that do not contain 0.5 in the 95% confidence interval obtained from bootstrapping on the training set are tested in the holdout set as they may contain prognostic information that identify patients at high risk for an adverse event.

Of the total 1234 patient records, we randomly select (without replacement) a training set of size 827 (67% of total population) holding the ratio of patients dead at the 90th day following admission to patients alive at the 90th day following admission constant at 7.9%. Therefore, in the training set, we have 762 patients alive at the 90th day following admission and 65 patients dead at the 90th day following admission.

The remaining 33% of the 1234 patients is the holdout set in which we will test only the metrics that do not contain 0.5 in the 95% confidence interval. The ratio of patients dead at the 90th day following admission to patients alive at the 90th day following admission is maintained at the original ratio in the total population (7.9%) as well. Therefore, in the holdout set, we have 375 patients alive at the 90th day following admission and 32 patients dead at the 90th day following admission.
We will select 827 patients with replacement from the training set and find the AUC by developing the ROC curve for this population. We repeat this process 100 times and develop a distribution of AUCs for each population selected from the training set. The results are shown in Figure 25. Additional details can be found in Appendix B.

Figure 25: Estimates of AUC sampling distributions metrics derived from the training set for the 9 time domain metrics derived from the MD time series. The metrics are labeled as follows: (A) is $\mu_{md}$, (B) is $\sigma_{md}$, (C) is $\mu_{md}^{diff}$, (D) is $\sigma_{md}^{diff}$, (E) is $ASDNN_{md}$, (F) is $SDANN_{md}$, (G) is $ASDNN_{md}^{diff}$, (H) is $SDANN_{md}^{diff}$, (I) is $rMSSD_{md}$. Edges of the box are the 25th percentile and 75th percentile cutoffs, the red line is the median, and the green dotted line is plotted as reference at AUC=0.5.
The 95% confidence intervals for all 9 metrics (see Appendix B) do not include 0.5.

Therefore, we tested all nine metrics on the entire holdout set and got the following results:

<table>
<thead>
<tr>
<th>Metrics</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUC</td>
<td>0.5279</td>
<td>0.5194</td>
<td>0.4955</td>
<td>0.5191</td>
<td>0.5153</td>
<td>0.5009</td>
<td>0.5197</td>
<td>0.4583</td>
<td>0.5191</td>
</tr>
</tbody>
</table>

**Table 2**: AUC for time-domain metrics A-I tested on the holdout set. The metrics are labeled as follows: (A) is $\mu_{md}$, (B) is $\sigma_{md}$, (C) is $\mu_{md}^{dif}$, (D) is $\sigma_{md}^{dif}$, (E) is $ASDNN_{md}$, (F) is $SDANN_{md}$, (G) is $ASDNN_{md}^{dif}$, (H) is $SDANN_{md}^{dif}$, (I) is $rMSSD_{md}$.

Because we do not see a significant deviation from an AUC of 0.5 (with the accepted benchmark of around 0.7 being significantly predictive or usable clinically and the converse 0.3 as being inverse predictive - i.e. when the patients score from the metric is lower than $\gamma$, it is predictive of high risk for a repeat adverse event), it appears that the morphological distance time series for the ABP waveform may not capture prognostic information as expected.

The lack of significant power to predict high risk patients may be due to two main reasons:

1) The population is not restricted to patients at predisposition for cardiovascular problems.

Therefore indicators of future adverse events may not be as strong in cardiovascular waveforms (such as the ECG and ABP). It is possible also that the magnitude of these indicators is small enough that the signal to noise ratio (SNR) is low in the resulting MD time series.

We can approach these problems by restricting the population to only patients with predisposition to cardiovascular problems, developing better processing techniques that amplify the subtle prognostic information that may be contained in the morphology of the beats, or developing other time domain metrics that better extract prognostic information from the MD time series that may have a low SNR ratio. We will explore the first method, restricting
the population to patients predisposed to cardiovascular problems in the next section and the other methods in the Future Work Section.

2) Sympathetic and parasympathetic changes occur in the order of minutes whereas the MD time series explores beat-to-beat changes which occur on the order of ~1 second. The difference between time scales may mean that we are not looking at difference in morphology of the right beats and rather should compare beat x with the morphology of a beat that follows several minutes later. Some approaches towards achieving this goal are described in the Future Work section. Further, an attempt to perform this analysis on a reduced dataset of only patients predisposed to Cardiovascular Problems is detailed in Appendix A.

**Downstroke Time Series**

From the *Downstroke* time series and the Difference in *Downstroke* time series, we derived 9 metrics: $\mu_{\text{Downstroke}}$, $\sigma_{\text{Downstroke}}$, $A_{\text{SDNN}_{\text{Downstroke}}}$, $SD_{\text{Downstroke}}$, $R_{\text{MSSD}_{\text{Downstroke}}}$, $\mu_{\text{Downstroke}}^{\text{dif}}$, $\sigma_{\text{Downstroke}}^{\text{dif}}$, $A_{\text{SDNN}_{\text{Downstroke}}^{\text{dif}}}$, $SD_{\text{Downstroke}}^{\text{dif}}$. We also tested the power of these time domain metrics that are obtained from the *Downstroke* time series using the same methodologies used to test time domain metrics derived from the morphological distance time series. We once again use the 1234 patient records with 7.9% dead at the 90th day following admission to the ICU. We then derive our 67% training set of size 827 (while maintaining the 7.9% of patients dead at 90 days in the population) with 65 deaths within 90 days following admission and 762 not dead at the 90th day following admission. The results are plotted below in Figure 26.
Figure 26: Estimates of AUC sampling distributions metrics derived from the training set for the 9 time domain metrics derived from the Downstroke time series. The metrics are labeled as follows: (A) is $\mu_{\text{Downstroke}}$, (B) is $\sigma_{\text{Downstroke}}$, (C) is $\mu_{\text{Downstroke}}^\text{diff}$, (D) is $\sigma_{\text{Downstroke}}^\text{diff}$, (E) is $\text{ASDNN}_{\text{Downstroke}}$, (F) is $\text{SDANN}_{\text{Downstroke}}$, (G) is $\text{ASDNN}_{\text{Downstroke}}^\text{diff}$, (H) is $\text{SDANN}_{\text{Downstroke}}^\text{diff}$, (I) is $r\text{MSSD}_{\text{Downstroke}}$. Edges of the box are the 25th percentile and 75th percentile cutoffs, the red line is the median, and the green dotted line is plotted as reference at AUC=0.5.

We see that the 95% confident interval for $\text{ASDNN}_{\text{Downstroke}}$ and $\text{SDANN}_{\text{Downstroke}}$ do not contain 0.5 so we test the other 7 metrics on the holdout set and find the following AUCs for $\mu_{\text{Downstroke}}$, $\sigma_{\text{Downstroke}}$, $\mu_{\text{Downstroke}}^\text{diff}$, $\sigma_{\text{Downstroke}}^\text{diff}$, $\text{ASDNN}_{\text{Downstroke}}^\text{diff}$, $\text{SDANN}_{\text{Downstroke}}^\text{diff}$, and $r\text{MSSD}_{\text{Downstroke}}$ respectively: 0.5171, 0.4712, 0.5075, 0.4868, 0.4845, 0.4491, and 0.4866. Additional details and results are provided in Appendix B.
**Frequency Domain Metrics**

Our first step towards analyzing the Morphological Distance Time Series in the frequency domain is to determine the appropriate diagnostic frequency, spectral range, in the frequency domain whose energy contains the most prognostic information. Ideally, we would train on one dataset by determining the optimal diagnostic frequency that produces $MV_{ABP}$ scores that are most predictive of death in 90 days. Then, once again ideally, we would test this diagnostic frequency by applying it to another dataset to determine its efficacy in producing predictive $MV_{ABP}$ scores.

In order to circumvent this situation, we once again divide our entire dataset into a training set and a holdout or test set. We begin with 1234 viable patients (patients with computable $MV_{ABP}$ scores and death records) from which we randomly, without replacement, draw 827 (67%) of patients to be in our training set. The other 407 patients (33%) of patients are then held out to be the test set.

In order to determine the optimal diagnostic frequency, we compute the vector of $MV_{ABP}$ scores for every possible diagnostic frequency range. Each range is identified by the start frequency and the end frequency. We stepped through each possible where start frequencies range from $0\rightarrow 0.55$ Hz by steps of $0.05$ Hz and end frequencies range from $0.05 \rightarrow 0.6$ Hz by
steps of 0.05 Hz. This corresponds to 66 possible frequency ranges (only 66 are viable because \(^{12}_2\) have unique start and end values).

We develop an ROC curve for each vector and compute the resulting area under the curve (AUC). This provides us with 66 values for each possible diagnostic frequency range (shown in Figure 27).

![AUC values from ROC curves obtained MV_{\text{ABP}} Scores computed from every possible diagnostic frequency range where start frequencies range from 0\rightarrow0.55 Hz by steps of 0.05 Hz and end frequencies range from 0.05 \rightarrow 0.6 Hz by steps of 0.05 Hz. Optimal diagnostic frequency ranges are 0.3 \rightarrow 0.4Hz and 0.3 \rightarrow 0.55 Hz.](image)

Figure 27: AUC values from ROC curves obtained MV_{\text{ABP}} Scores computed from every possible diagnostic frequency range where start frequencies range from 0\rightarrow0.55 Hz by steps of 0.05 Hz and end frequencies range from 0.05 \rightarrow 0.6 Hz by steps of 0.05 Hz. Optimal diagnostic frequency ranges are 0.3 \rightarrow 0.4Hz and 0.3 \rightarrow 0.55 Hz.

\(^3\) For a heart rate of \~60 beats per minute, it will be sufficient to extend the end frequency to a maximum 0.5Hz but we extended to 0.6Hz in order to be consistent with the original studies published on developing the optimal diagnostic frequency.
We can see from image Figure 27 that the optimal frequency ranges are from $0.3 \text{Hz} \rightarrow 0.4 \text{Hz}$ and $0.3 \text{Hz} \rightarrow 0.55 \text{Hz}$. Interestingly, the $0.3 \text{Hz} \rightarrow 0.55 \text{Hz}$ is the same optimal frequency range determined for ECG waveforms in the population of patients with Acute Coronary Syndrome [9]. We then test the $\text{MV}_{\text{ABP}}$ scores computed with the diagnostic frequency range of $0.3 \text{Hz}$ to $0.4 \text{Hz}$ on the 427 patients in the holdout set. After computing the ROC curve for these scores, we get an AUC of 0.5001. We also test the $\text{MV}_{\text{ABP}}$ scores computed with the diagnostic frequency range of $0.3 \text{Hz}$ to $0.55 \text{Hz}$ on the 427 patients in the holdout set. After computing the ROC curve for these scores, we get an AUC of 0.4921. Both results indicate that the $\text{MV}_{\text{abp}}$ scores may not be predictive. This may be either because of lack of prognostic information from the MD time series itself, the time length of 5 minutes to segment the MD time series and compute energies may be too short or too long, and there may be more no information to be gleaned from the frequency domain at all for the ABP waveform.
Future Work

**Modifying the Morphological Distance Time Series**

The morphological distance (MD) time series derived from the ABP waveform does not actually capture the important morphological changes that occur on the order of minutes. Instead, the MD time series focuses on changes that occur on a beat-to-beat basis. Therefore, it would be beneficial to explore series that find the distances from a beat to a beat a few minutes away. However, we are blindsided in that we do not know the optimal time lag that will produce the most predictive MD time series. We will discuss a few approaches to get around this problem.

1.) Average beat-to-every-other-beat in five minute windows. We begin by dividing the 24-hr. preprocessed (see **Methods: Preprocessing the Arterial Blood Pressure Waveform**) ABP waveform into non-overlapping 5 minute windows (288 total windows). This results in approximately 300 beats per window. We now compute $\binom{300}{2} = 44850$ distances from each beat in the 5 minute window to every beat that follows (Figure 28).

We then compute the average of all these distances within the five minute window. We repeat this procedure for every five minute window in the 24 hour recording and obtain the 5 Minute Average MD Time Series. We can now derive time domain metrics that summarize the entire 24 hr. recording such as the mean, standard deviation, and square root of the mean squared differences of the entire 5 Minute Average MD Time Series.
Figure 28: The first five minutes of an example ABP waveform. Blue arrows represent distance from the first beat to every other beat that follows it in the five minute window and green arrows represent distance from the first beat to every other beat. We compute these distances for every beat $x$ to every beat that follows beat $x$ in the five minute window.

It is unclear how effective metrics like SDANN or ASDNN will be because the time series already averages distances within each five minute window. Further, SDANN (standard deviation of the average NN intervals - see Methods: Time Domain Metrics) and ASDNN (average of the standard deviation of NN intervals - see Methods: Time Domain Metrics) will have to be modified to encompass time lengths longer than five minutes (such as 60 minutes) because the 5 Minute Average MD Time Series is sparse and contains one data point for every five minutes. This will in turn change what information the SDANN and ASDNN provides us.

2.) We can find the optimal time or 'beat' lag, $\Delta b$, from a beat $x$ to the beat $x+\Delta b$ and create a time series of the distance from every beat to the beat that is an optimal $\Delta b$ away.

For computational as well as physiological purposes, $\Delta b$ should be limited to a maximum of $\sim 600$ (which is about 10 minutes from each beat) because there is no reason to examine
beats that are farther away than the typical time scale of a sympathetic or parasympathetic modulation.

Figure 29: Examples of time series computed using various $\Delta b$ lags. (A) shows the derivation of the canonical MD time series where $\Delta b = 1$. (B) shows derivation of an MD time series computed using a $\Delta b = 2$. (C) shows derivation of an MD time series computed using a $\Delta b = 3$.

In order to find the optimal $\Delta b$, we probe the $\Delta b$ by varying $\Delta b$ from 1 to 600 with either steps of 1, 2, or 5 (based on computational feasibility and time constraints). After computing MD time series from the preprocessed ABP waveforms with these various $\Delta b$ lags, we will be
able to compute the same time domain metrics computed for the canonical MD time series
including \( \mu_{md}, \sigma_{md}, rMSSD_{md}, ASDNN_{md} \) and \( SDANN_{md} \).

To validate this method, we first choose a single metric (let's use \( \mu_{md} \) as an example).
We compute the \( \mu_{md} \) of each patient for an MD time series with a particular \( \Delta b \). We use that as a classifier and compute the resulting AUC. We repeat this method for every possible \( \Delta b \)
time series. The optimal \( \Delta b \), or \( \Delta b^* \), time series will provide the highest AUC. Following our cross-validation methodology, we can then compute the \( \Delta b^* \) MD time series for each patient in the testing set, compute \( \mu_{md} \) for each patient and compute the AUC of the resulting classifier. This process is then repeated for every time domain metric in order to identify the optimal 'beat lag' that provides the best prognostic information.

This same procedure can be applied to Difference in \( \Delta b \) MD Time Series to determine if the difference in adjacent points of the \( \Delta b \) MD Time Series can provide a significant classifier as well.

**Frequency Domain Analysis of the Downstroke Time Series**

We derived the Downstroke Time Series from the ABP waveform in the section Methods: Downstroke Time Series and analyzed this time series in the time domain. Another approach to analyzing the time series, which was not completed due to the lack of time, is analysis in the frequency domain. Because the Downstroke contains very different prognostic information from the MD time series, we need to once again find the optimal diagnostic frequency from which we can derive a \( MV_{Downstroke} \) Score for each patient that can be used as a classifier.
We define a spectral range called the diagnostic frequency. We begin by segmenting the time series into not overlapping windows of 5 minutes in length. We compute the spectral energy of the five minute mini time series using lomb scargle parametric estimation and sum the total energy in the diagnostic frequency. We compute this total energy for every five minute window over a 24 hour period for the patient and consider the 90th percentile to be the MV\textsubscript{Downstroke} score for the patient. To test this score as a classifier, we once again develop an ROC curve and find the AUC where we consider an MV\textsubscript{Downstroke} score above a specified cutoff, Y to be predictive of high risk for a future adverse event. We then vary the diagnostic frequency range and compute every possible AUC. The optimal diagnostic frequency range is the one that results in the highest AUC. We can apply this diagnostic frequency to an independent set of patients (most likely the holdout set) to determine the predictive power of the MV\textsubscript{Downstroke} score in an independent dataset.

*Deriving the 'Upstroke' Time Series*

We have studied the decay of the ABP waveform from peak systole to following diastole which contains information on the peripheral resistance and arterial compliance, but another portion of interest in an ABP waveform is the shape of the upstroke from diastole to following peak systole. This portion of the curve contains mainly prognostic information on the contractility.

To develop the 'Upstroke' time series, we first follow the same signal pre-processing methods used prior to developing the morphological distance time series (see Methods: Preprocessing the Arterial Blood Pressure Waveform). Following signal pre-processing, we are
left with a de-trended, noise-reduced continuous recording of NN intervals for the ABP waveform. Then, we identify each trough corresponding to the diastole which occurs at \((t_d, P_d)\) and the following peak corresponding to the systole which occurs at \((t_s, P_s)\) where \(t_s > t_d\). We then find the slope, \(m\), of the linear interpolation of at \((t_d, P_d)\) and the \((t_s, P_s)\).

\[
m = \frac{P_d - P_s}{t_d - t_s}
\]

The series of upstroke slopes for each beat in the ABP recording creates the 'Upstroke' Time Series. This time series can be analyzed in both the time domain as well as the frequency domain (for which relevant metrics such as the mean, standard deviation, ASDNN, SDANN, and MV scores) have already been discussed several times in this paper. Analogous metrics can be derived and analyzed from the 'Upstroke' Time Series as well.
Figure 30: Linear interpolation of the peak of diastole and the peak corresponding to the following systole shown for a single beat in the ABP waveform.

Figure 31: Derivation of the 'Upstroke' Time Series which converts continuous ABP recording into the discretized time series of the linear slopes of the upstroke from each diastole and to the following peak systole.
Of more interest would be the Differences in the Upstroke Time Series which is derived by taking the difference between adjacent points in the Upstroke Time Series because it quantifies the changes in the upstroke from beat to beat. This consequently provides prognostic information of changes in the ratio of the stroke volume to the arterial compliance. We may also benefit from doing an analysis of changes in the upstroke from each beat to the beat Δb beats away. This may be especially effective in discovering the effects of parasympathetic to sympathetic modulation that act on the order of minutes as well as more long-term effects to the stroke volume from muscle weakness and deterioration of the myocardium.

**Adjustable Parameters**

Throughout this thesis, we have kept some parameters constant as though they have already been shown to be optimal for the processing and analysis of the ABP waveform. Namely, when doing frequency domain analysis, we segment the MD time series into smaller series of 5 minutes in length and we take the 90th percentile MV_{ABP} score to be the score for the patient. I do not think a window size smaller than five minutes can produce an accurate spectral range due to the sparseness of data points in the smaller window, there is nothing that stops us from expanding the window to larger lengths like 8 minutes, 10 minutes, 20 minutes, or even a half hour. Further, there is no conclusive evidence that shows that the 90th percentile is the optimal energy for the entire 24 hour time series to be the MV_{ABP} score. Further, we may even benefit from setting the MV_{ABP} score to be an interpolation of all the energies from each
five minute window. There is much work needed to fine-tune these parameters to their optimal values in order to create a more predictive \( MV_{ABP} \) score.

Further, for the derivations of ASDNN and SDANN, we segment the time series, again into five minute windows. The five minute window is not magic and is merely a reasonable cutoff that has been arbitrarily decided. Our analysis may require windows to the length of 10 or even 20 minutes so before effects from sympathetic and parasympathetic modulation become apparent. This way, ASDNN and SDANN may possibly highlight these influences without being mired by extraneous fluctuations present in shorter time spans of the MD time series.

Therefore, there is much to be fine-tuned even in methods that we executed and analyzed in the MIMIC II dataset. There may be prognostic information in the MD time series and the Downstroke time series that can be extracted by finding optimal parameters to build time domain as well as frequency domain classifiers. In conjunction to the various time series explored above, this research would benefit from the fine tuning of adjustable parameters that make more optimal classifiers in both the time and frequency domain.
Conclusion

We began with the hypothesis that the arterial blood pressure waveform provides prognostic information that can help identify patients at high risk for a future adverse event. We specifically restricted our study to patients who are physiologically compromised but not necessarily in the ICU because of a coronary problem. Therefore, we focused our entire study on patients in the MIMIC II database (see Data) who are admitted to the ICU for various reasons (and share the common characteristic of being physiologically compromised). This implicitly postulates cardiac waveforms provide signals of physiological deterioration into a future adverse event regardless of the original cause of the initial adverse event. We rigorously defined an adverse event to be death within 90 days following admission to the ICU.

We developed new methods to process the ABP waveform and analyzed the processed waveform in two different time series derived from a 24 hour ABP recording: the morphological distance time series and the Downstroke time series. The morphological distance time series is used to quantify changes in beat-to-beat morphology and converts the continuous waveform into a series of distances between adjacent beats. From this time series, we developed 3 summary statistics of the entire 24 hour recording (μ_{md}, σ_{md}, and rMSSD_{md}) and 2 statistics derived by segmenting the waveform into 5 minute non-overlapping windows then finding the average and standard deviation of the standard deviations and averages respectively from each 5 minute window (ASDNN_{md} and SDANN_{md}). We found that all these metrics were not effective in identifying patients at risk for death within 90 days following admission to the ICU.
We also derived the Difference in MD time series which removed high frequency fluctuations in the MD time series and smoothed the waveform further (at the risk of loss of prognostic information at high frequencies). Metrics derived from this series include: $\mu_{md}^{dif}$, $\sigma_{md}^{dif}$, $SDANN_{md}^{dif}$, and $ASDNN_{md}^{dif}$ where $\mu_{md}^{dif}$ and $\sigma_{md}^{dif}$ are summary statistics of the entire 24 hours of the Difference in MD time series and $SDANN_{md}^{dif}$ and $ASDNN_{md}^{dif}$ are computed by segmenting the series into 5 minute windows then summarizing the average and standard deviations of the time series in these five minute windows. Once again, we were unable to find metrics that show a significant ability to predict patients at high risk for a future adverse event.

We also examined the MD time series in the frequency spectrum by computing Morphological Variability (of the ABP Waveform) Scores ($MV_{ABP}$ scores) for the patients. We first experimentally determined an optimal frequency band, or the diagnostic frequency, to serve as our spectral frequency in the derivation of the $MV_{ABP}$ scores. We determine this band to be 0.3Hz to 0.55 Hz (which is the same optimal diagnostic frequency derived in other studies on Morphological Variability of the ECG Waveform). Even with this optimal diagnostic frequency, the $MV_{ABP}$ scores were not significant predictors of death within 90 days following admission to the ICU.

We postulate that the lack of predictive power from these 9 metrics derived from the MD time series is for two large reasons:

1) The ABP waveform does not contain significant prognostic information to predict a future adverse event for ICU patients.
2) The morphological distance time series computes beat-to-beat distances which occur at an average rate of about 1 distance per second when changes to the arterial blood pressure waveform is on the order of minutes. This is because the main components of the ABP waveform are dictated by sympathetic and parasympathetic modulation (which affect stroke volume, arterial compliance, and peripheral resistance) which occur on the order of several minutes.

To deal with this, we have developed alternate methods by which to derive and analyze the MD time series in the Future Work section. These include computing beat-to-every-other-beat distances in small windows of about 2-5 minutes in length and taking the average distance of each window thereby including information in the MD time series of fluctuations in morphology happening at greater than the span of the time from one beat to the next. Further, we discuss finding the optimal beat-to-beat paired distance (i.e. such as a time series computed by determining the distances of each beat with the fifth beat that follows) within each 2-5 minute window and computing those distances over the entire signal instead of computing the distance of a beat to its adjacent beat.

3) Patients in the ICU don’t necessarily exhibit strong indicators of future adverse events in their cardiac waveforms unless the future adverse event is triggered by a cardiovascular problem or they are predisposed to cardiovascular problems.

We also tried to quantify changes to the joint effects of peripheral resistance and arterial compliance by computing the rate of decay from peak systole to following diastole for every beat in a 24 hours recording thus deriving the Downstroke time series. We examined the
same time domain metrics used to analyze the MD time series but applied to the Downstroke time series. These include $\mu_{Downstroke}$, $\sigma_{Downstroke}$, and $rMSSD_{Downstroke}$ which statistically summarize the entire 24-hour time series and $ASDNN_{Downstroke}$ and $SDANN_{Downstroke}$. which are derived by segmenting the waveform into 5 minute non-overlapping windows then finding the average and standard deviation of the standard deviations and averages respectively from each 5 minute window respectively.

Of more interest was the Difference in Downstroke time series which provides us with a beat-to-beat change in decay rates over 24 hours with the postulate that changes to the decay are evidence of underlying changes to the cardiovascular system including sympathetic and parasympathetic modulation. The metrics we derive from this include $\mu_{dP/dt}^{dif}$, $\sigma_{dP/dt}^{dif}$, $ASDNN_{dP/dt}^{dif}$, and $SDANN_{dP/dt}^{dif}$. The $\mu_{dP/dt}^{dif}$ and $\sigma_{dP/dt}^{dif}$ especially provide us with information on the average change in rate of decay as well the variation in this average change over 24 hours.

However, we once again find that the metrics appear unable to significantly predict patients at risk for a future adverse event. This may be because the time scale of the changes is not significant in capturing sympathetic and parasympathetic modulations, the ABP waveform itself isn't predictive, or the ABP waveform is not predictive in ICU patients but may be predictive in patients predisposed to cardiovascular problems or an adverse event related to a cardiovascular malfunction. Due to lack of time, frequency domain analysis was not performed on the Downstroke time series but details on carrying out frequency domain analysis on this time series are provided in the Future Work section.
We also detail in Future Work an alternate, but complementary time series that derives the upstroke (from diastole to following peak systole) and contains mainly prognostic information on the stroke volume and arterial compliance. Due to lack of time, this was not completed but is detailed in the Future Work section.

As a conclusion, we developed methods by which to process the continuous ABP waveform and derive discrete time series from which risk metrics, or classifiers of patients at high risk for a future adverse event can be computed. The time domain and frequency domain analysis of the MD time series did not yield metrics with sufficient power to predict adverse events.

That being said, the arterial blood pressure waveform is vastly unstudied as a computational biomarker and this is one of the first forays into the study of the ABP waveform in risk stratification. Therefore, these experiments are merely gateways to the experiments describes in the Future Work section and beyond even that. The arterial blood pressure waveform has a vast trove of information on the mechanical nature of the heart and its vasculature that the ECG cannot always express. There is definitely prognostic information that can be exploited when the ABP is processed and analyzed in the optimal manner to produce a strong classifier of high risk patients for future adverse events.
Appendix A

Morphological Distance Time Series: Patients Predisposed to Cardiovascular Problems

As mentioned in the previous section, we now turn our attention to only patients predisposed to cardiovascular problems in the ICU population. There are two approaches we use to determining which patients are predisposed to cardiovascular problems:

1) Limit the population to patients who have been administered a beta blocker or a anti-hypertensive drug during their ICU stay.\(^4\)

2) Limit the population to patients who are classified as having 'supraventricular tachycardia' or an average heart rate greater than 100 beats per minute.\(^5\)

We first approached the problem by looking at patients who are administered a beta blocker or anti-hypertensive drug. Beta blockers are a class of drugs that target and block beta receptors found in the cells of tissues that are part of the sympathetic nervous response system such as the heart muscles, smooth muscles, arteries, kidneys, etc. They are often administered to prevent a second myocardial infarction following the first heart attack and are also administered to control hypertension and cardiac arrhythmias (but not restricted to only these uses) [15]. The list of beta blockers considered in our study is detailed in Appendix C. Anti-hypertensive drugs are used to control hypertension and are indicative of the patient having

\(^4\) One problem with this approach is that the drug may affect the ABP waveform sufficiently to remove prognostic information that can be gleaned from the waveform. Further, there are many reasons other than just cardiovascular reasons that may require the patient to be administered these drugs.

\(^5\) There are many reasons why the heart rate is elevated such as bleeding, sepsis, etc. Therefore, the average heart rate being greater than 100 bpm may not really narrow the patients to only those with cardiovascular disease.
prior history of cardiovascular problems as well. The list of anti-hypertensive drugs in our study is in Appendix C.

After reducing our patient population to only those administered either a beta blocker or anti-hypertensive drug (or both), we found that there were a total of 137 such patients of which 128 patients were alive at the 90th day following admission and 9 patients were dead at the 90th day following. Because the 67% training set derived from this reduced population has only 7 deaths and the AUC is extremely sensitive when the number of a particular label (in this case, dead at the 90th day) is low, we considered the sample size too small to make conclusions from the results of this reduction. The results of this study are detailed in Appendix B. We would further benefit from having a larger dataset in which to test the postulate that the MD time series derived from the ABP waveform is predictive in patients predisposed to cardiovascular problems.

We also restricted the original 1234 patients to only those with supraventricular tachycardia or a heart rate greater than 100 bpm. Prior literature on patients with high heart rates, specifically those with heart rates greater than 100 bpm, suggests that there is significant correlation between the high heart rates and all-cause, cardiovascular and non-cardiovascular death, increased risk of atherosclerosis, and sudden cardiac death [10][11]. This is especially interesting as deaths in the ICU may be due to both cardiovascular as well as non-cardiovascular reasons and help group patients with a prior cardiac indicator of high risk for death. We believe the patients with supraventricular tachycardia are especially predisposed to a repeat adverse event and thus reduce the population to only patients with heart rates greater than 100 bpm (the threshold for supraventricular tachycardia). However, we once again run
into the problem of sparse data with 96 total patients with heart rates greater than 100 bpm (with only 8 deaths by the 90th day following admission and 88 alive at the 90th day following admission). We once again conclude that the results from this reduced population are not sufficient to draw conclusions. The results are detailed in Appendix B.
Appendix B - Additional Results

Morphological Distance Time Series

Key:

A: $\mu_{md}$
B: $\sigma_{md}$
C: $\mu_{md}^{dif}$
D: $\sigma_{md}^{dif}$
E: ASDNN$_{md}$
F: SDANN$_{md}$
G: ASDNN$_{md}^{dif}$
H: SDANN$_{md}^{dif}$
I: rMSSD$_{md}$

Full Data Set Results:

Training Set
867 Total Patients
65 Dead at 90th day following admission
762 Alive at 90th day following admission

Testing Set
407 Total Patients
32 Dead at 90th day following admission
375 Alive at 90th day following admission

Full Dataset

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Means</td>
<td>0.5304</td>
<td>0.5398</td>
<td>0.4606</td>
<td>0.5428</td>
<td>0.5418</td>
<td>0.5334</td>
<td>0.5431</td>
<td>0.5145</td>
<td>0.5451</td>
</tr>
<tr>
<td>CI Lower Bound</td>
<td>0.5234</td>
<td>0.5326</td>
<td>0.4527</td>
<td>0.5355</td>
<td>0.5331</td>
<td>0.5257</td>
<td>0.5349</td>
<td>0.5066</td>
<td>0.538</td>
</tr>
<tr>
<td>CI Higher Bound</td>
<td>0.5373</td>
<td>0.547</td>
<td>0.4685</td>
<td>0.5501</td>
<td>0.5506</td>
<td>0.5411</td>
<td>0.5514</td>
<td>0.5224</td>
<td>0.5521</td>
</tr>
</tbody>
</table>

Testing AUC
0.5279 0.5194 0.4955 0.5191 0.5153 0.5009 0.5197 0.4583 0.5191
Patients Administered Beta Blockers or anti-Hypertensive Drugs:

Training Set
93 Total Patients
7 Dead at 90th day following admission
86 Alive at 90th day following admission

Testing Set
44 Total Patients
2 Dead at 90th day following admission
42 Alive at 90th day following admission

<table>
<thead>
<tr>
<th>Beta Blockers or anti-Hypertensive Drugs Administered</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Means</td>
<td>0.5807</td>
<td>0.5225</td>
<td>0.3883</td>
<td>0.5278</td>
<td>0.4908</td>
<td>0.5722</td>
<td>0.5087</td>
<td>0.4449</td>
<td>0.5175</td>
</tr>
<tr>
<td>CI Lower Bound</td>
<td>0.5635</td>
<td>0.5045</td>
<td>0.371</td>
<td>0.511</td>
<td>0.4771</td>
<td>0.5534</td>
<td>0.4946</td>
<td>0.4162</td>
<td>0.5008</td>
</tr>
<tr>
<td>CI Higher Bound</td>
<td>0.598</td>
<td>0.5405</td>
<td>0.4057</td>
<td>0.5447</td>
<td>0.5045</td>
<td>0.591</td>
<td>0.5227</td>
<td>0.4735</td>
<td>0.5342</td>
</tr>
<tr>
<td>Testing AUC</td>
<td>0.5238</td>
<td>0.5</td>
<td>0.6786</td>
<td>0.5</td>
<td>0.5238</td>
<td>0.631</td>
<td>0.5595</td>
<td>0.6548</td>
<td>0.5</td>
</tr>
</tbody>
</table>

AUC distributions from Bootstrapping: Training Set (67%) of Patients Administered Beta Blockers or Hypertension Drugs
Patients with Heart Rate Greater than 100 bpm

Training Set
65 Total Patients
6 Dead at 90th day following admission
59 Alive at 90th day following admission

Testing Set
31 Total Patients
2 Dead at 90th day following admission
29 Alive at 90th day following admission

Supraventricular Tachycardia (Heart Rate >100 bpm)

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Means</td>
<td>0.4708</td>
<td>0.5043</td>
<td>0.3964</td>
<td>0.5246</td>
<td>0.4043</td>
<td>0.2952</td>
<td>0.3984</td>
<td>0.1325</td>
<td>0.5236</td>
</tr>
<tr>
<td>CI Lower Bound</td>
<td>0.4479</td>
<td>0.485</td>
<td>0.3797</td>
<td>0.5006</td>
<td>0.3822</td>
<td>0.2741</td>
<td>0.3746</td>
<td>0.1202</td>
<td>0.4984</td>
</tr>
<tr>
<td>CI Higher Bound</td>
<td>0.4937</td>
<td>0.5235</td>
<td>0.413</td>
<td>0.5486</td>
<td>0.4263</td>
<td>0.3164</td>
<td>0.4223</td>
<td>0.1448</td>
<td>0.5489</td>
</tr>
</tbody>
</table>

Testing AUC
0.4655 0.5 0.2586 0.5172 0.3966 0.431 0.431 0.2241 0.5172

AUC distributions from Bootstrapping; Training Set (67%) of Patients with Heart Rate Greater than 100 bpm
**Downstroke Time Series**

Key:

- **A**: \( \mu_{\text{Downstroke}} \)
- **B**: \( \sigma_{\text{Downstroke}} \)
- **C**: \( \mu_{\text{Downstroke}}^\text{diff} \)
- **D**: \( \sigma_{\text{Downstroke}}^\text{diff} \)
- **E**: \( ASDNN_{\text{Downstroke}} \)
- **F**: \( SDANN_{\text{Downstroke}} \)
- **G**: \( ASDNN_{\text{Downstroke}}^\text{diff} \)
- **H**: \( SDANN_{\text{Downstroke}}^\text{diff} \)
- **I**: \( rMSSD_{\text{Downstroke}} \)

**Full Data Set Results:**

- **Training Set**
  - 867 Total Patients
  - 65 Dead at 90th day following admission
  - 762 Alive at 90th day following admission

- **Testing Set**
  - 407 Total Patients
  - 32 Dead at 90th day following admission
  - 375 Alive at 90th day following admission

<table>
<thead>
<tr>
<th>Downstroke Full Dataset</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Means</td>
<td>0.4571</td>
<td>0.4242</td>
<td>0.5705</td>
<td>0.457</td>
<td>0.4989</td>
<td>0.4969</td>
<td>0.4922</td>
<td>0.5389</td>
<td>0.4601</td>
</tr>
<tr>
<td>CI Lower Bound</td>
<td>0.4507</td>
<td>0.4173</td>
<td>0.5638</td>
<td>0.4504</td>
<td>0.4919</td>
<td>0.4889</td>
<td>0.4844</td>
<td>0.5312</td>
<td>0.4538</td>
</tr>
<tr>
<td>CI Higher Bound</td>
<td>0.4634</td>
<td>0.4311</td>
<td>0.5773</td>
<td>0.4636</td>
<td>0.506</td>
<td>0.5049</td>
<td>0.4999</td>
<td>0.5465</td>
<td>0.4663</td>
</tr>
<tr>
<td>Testing AUC</td>
<td>0.5271</td>
<td>0.4712</td>
<td>0.5075</td>
<td>0.4868</td>
<td>0.4842</td>
<td>0.5143</td>
<td>0.4845</td>
<td>0.4491</td>
<td>0.4866</td>
</tr>
</tbody>
</table>
Appendix C - Drugs

Beta Blockers for High Blood Pressure queried in MIMIC II Clinical Database

<table>
<thead>
<tr>
<th>Generic Name</th>
<th>Brand Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>acebutolol</td>
<td>Sectral</td>
</tr>
<tr>
<td>atenolol</td>
<td>Tenormin</td>
</tr>
<tr>
<td>beta blockers</td>
<td>--</td>
</tr>
<tr>
<td>carvedilol</td>
<td>Coreg</td>
</tr>
<tr>
<td>labetalol</td>
<td>Trandate</td>
</tr>
<tr>
<td>metoprolol</td>
<td>Lopressor, Toprol</td>
</tr>
<tr>
<td>nadolol</td>
<td>Corgard</td>
</tr>
<tr>
<td>penbutolol</td>
<td>Levatol</td>
</tr>
<tr>
<td>pindolol</td>
<td>--</td>
</tr>
<tr>
<td>Propranolol</td>
<td>Inderal</td>
</tr>
</tbody>
</table>

Courtesy WebMD

**Two drugs found in MIMIC II Clinical Database: metaprolol and esmolol**

anti-Hypertensive Drugs queried in MIMIC II Clinical Database

**One anti-hypertensive drug found in MIMIC II Clinical Database: diltiazem**

<table>
<thead>
<tr>
<th>DRUG NAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>lisinopril Oral</td>
</tr>
<tr>
<td>Ziac Oral</td>
</tr>
<tr>
<td>Mavik Oral</td>
</tr>
<tr>
<td>acebutolol Oral</td>
</tr>
<tr>
<td>amlodipine-valsartan-hctz Oral</td>
</tr>
<tr>
<td>atenolol Oral</td>
</tr>
<tr>
<td>Coreg CR Oral</td>
</tr>
<tr>
<td>Microzide Oral</td>
</tr>
<tr>
<td>reserpine Oral</td>
</tr>
<tr>
<td>olmesartan-amlodipine-hctz Oral</td>
</tr>
<tr>
<td>Diovan Oral</td>
</tr>
<tr>
<td>nifedipine Oral</td>
</tr>
<tr>
<td>Teveten Oral</td>
</tr>
<tr>
<td>Calan Oral</td>
</tr>
<tr>
<td>azilsartan medoxomil Oral</td>
</tr>
<tr>
<td>Bystolic Oral</td>
</tr>
<tr>
<td>Zestril Oral</td>
</tr>
<tr>
<td>metoprolol ta-hydrochlorothiaz Oral</td>
</tr>
<tr>
<td>amiloride Oral</td>
</tr>
<tr>
<td>metoprolol su-hydrochlorothiaz Oral</td>
</tr>
<tr>
<td>hydrochlorothiazide Oral</td>
</tr>
<tr>
<td>Cardizem CD Oral</td>
</tr>
<tr>
<td>valsartan-hydrochlorothiazide Oral</td>
</tr>
<tr>
<td>aliskiren Oral</td>
</tr>
<tr>
<td>Sodium Edecrin IV</td>
</tr>
<tr>
<td>Norvasc Oral</td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>Toprol XL Oral</td>
</tr>
<tr>
<td>metoprolol succinate Oral</td>
</tr>
<tr>
<td>metoprolol tartrate Oral</td>
</tr>
<tr>
<td>Lisinopril-hydrochlorothiazide Oral</td>
</tr>
<tr>
<td>Cozaar Oral</td>
</tr>
<tr>
<td>Clonidine Oral</td>
</tr>
<tr>
<td>Propranolol Oral</td>
</tr>
<tr>
<td>Azor Oral</td>
</tr>
<tr>
<td>Losartan Oral</td>
</tr>
<tr>
<td>Coreg Oral</td>
</tr>
<tr>
<td>Benicar HCT Oral</td>
</tr>
<tr>
<td>carvedilol Oral</td>
</tr>
<tr>
<td>Exforge Oral</td>
</tr>
<tr>
<td>Avapro Oral</td>
</tr>
<tr>
<td>Lotrel Oral</td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td>Hyzaar Oral</td>
</tr>
<tr>
<td>furosemide Oral</td>
</tr>
<tr>
<td>Altace Oral</td>
</tr>
<tr>
<td>Micardis Oral</td>
</tr>
<tr>
<td>Avalide Oral</td>
</tr>
<tr>
<td>Micardis HCT Oral</td>
</tr>
<tr>
<td>Lopressor Oral</td>
</tr>
<tr>
<td>enalapril maleate Oral</td>
</tr>
<tr>
<td>ramipril Oral</td>
</tr>
<tr>
<td>diltiazem HCl Oral</td>
</tr>
<tr>
<td>terazosin Oral</td>
</tr>
<tr>
<td>amlodipine-benazepril Oral</td>
</tr>
<tr>
<td>labetalol Oral</td>
</tr>
<tr>
<td>triamterene-hydrochlorothiazide Oral</td>
</tr>
<tr>
<td>Atacand Oral</td>
</tr>
<tr>
<td>losartan-chlorthalidone</td>
</tr>
</tbody>
</table>

102
<table>
<thead>
<tr>
<th>Drug Name</th>
<th>Formulation</th>
<th>Drug Name</th>
<th>Formulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>hydrochlorothiazide Oral</td>
<td>Oral</td>
<td>benazepril Oral</td>
<td>Oral</td>
</tr>
<tr>
<td>triamterene Oral</td>
<td>Oral</td>
<td>Zebeta Oral</td>
<td>Oral</td>
</tr>
<tr>
<td>doxazosin Oral</td>
<td>Oral</td>
<td>eplerenone Oral</td>
<td>Levatol Oral</td>
</tr>
<tr>
<td>hydralazine Oral</td>
<td>Oral</td>
<td>Nifediac CC Oral</td>
<td>betaxolol Oral</td>
</tr>
<tr>
<td>Tribenzor Oral</td>
<td>Oral</td>
<td>bumetanide Oral</td>
<td>torsemide IV</td>
</tr>
<tr>
<td>Maxzide Oral</td>
<td>Oral</td>
<td>irbesartan Oral</td>
<td>candesartan-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>telmisartan-amlodipine Oral</td>
<td>hydrochlorothiazid Oral</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dyrenium Oral</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>telmisartan Oral</td>
<td>Teveten HCT Oral</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Calan SR Oral</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>torsemide IV</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>amiloride-hydrochlorothiazide</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oral</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oral</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oral</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oral</td>
<td></td>
</tr>
</tbody>
</table>

*Courtesy WebMD*
Appendix D

*Morphological Variability of an Electrocardiogram*

Preprocessing the ECG signal

In this section we will discuss the method developed in *Relation of Death Within 90 Days of Non-ST-Elevation Acute Coronary Syndromes to Variability in Electrocardiographic Morphology* [9] to derive the Morphological Variability (MV). In the DISPERSE2 study on morphological variability, 862 continuous ECG recording of at least 24 hours in length were used. An automated method of preprocessing the signal was performed in order to remove noise and reject signals that failed to meet appropriate signal quality requirements. To tackle noise removal, we begin by estimating the baseline wander using median filtering and then subtracting the baseline wander from the signal. Then the signal is filtered using wavelet denoising with a soft threshold. Finally, the signal is normalized by the average amplitude of the R peaks. Ectopic beats were also removed thereby reducing the signal to only the NN intervals (beats arising from normal SA node depolarizations).

Certain signals were also rejected based on the quality of the signal as well as the physiological feasibility of the signal. The signal quality index (SQI) is also calculated for each ECG and the ECG is rejected if the signal has a low SQI. Following the preprocessing, the beat to beat distance between adjacent beats is calculated to determine the morphological distance time series.
Morphological Distance Time Series

The simplest way to quantify the difference between two adjacent beats is to subtract the samples of one beat from the other one. In order to do this, we must segment each beat from the start of the P-wave to the start of the next P-wave and subtract samples from one beat to the next. However, because heart rates vary between patients and even within a single patient's waveform, adjacent beats are slightly different lengths, blindly subtracting the beats may not be matching and subtracting one beat with the analogous physiological portion of the other beat. For instance, subtracting the QRS complex of one beat from the P-wave of the other beat does not give us an accurate distance between methods. Therefore, we must first align the two signals before deriving a distance metric. We do this using a dynamic time warping.
algorithm which finds an optimal alignment between two beats and sets the morphological distance between two beats to the cost of the optimal path alignment.

This dynamic time warp will align the corresponding parts of each beat as shown in Figure D.1. Given two beats, $x_1$ and $x_2$ of length $l_1$ and $l_2$, DTW will first produce an $[l_1 \times l_2]$ distance matrix, $D$. In this matrix, each entry, $(i,j)$ represents the distance between $x_1$ and $x_2$. A particular path through this matrix is defined as:

$$\varphi(k) = (\varphi_1(k), \varphi_2(k)) \quad 1 \leq k \leq K$$

where $\varphi_1$ and $\varphi_2$ are row and column indices and $K$ is the alignment length [8]. We then search for the minimum cost path using dynamic programming which corresponds to the optimal alignment of the beats. This optimal path, $C(x_1, x_2)$ is defined as:

$$C(x_1, x_2) = \min_{\varphi} C_{\varphi}(x_1, x_2)$$

Where,

$$\min_{\varphi} C_{\varphi}(x_1, x_2) = \sum_{k=1}^{K} d(x_1[\varphi_1(k)], x_2[\varphi_2(k)])$$

The final morphological distance between the two beats is the cost of the optimal alignment [9]. The result of applying this to an entire time series of beats is a time series of morphological distances.

**Morphological Variability: Analysis in the Frequency Domain**

Following development of the morphological distances in time, we convert every five minute window to the energy contained in the morphological distance time series in that five minute window. We sum the frequencies of energies between .3→.55Hz, known as the
diagnostic frequency; this sum is the morphological variability of the five minute window [9]. We then find the 90th percentile of all five minute window energy sums and that is the final, morphological variability for the individual patient's 24-hour ECG. In the original study, the patient population was dichotomized at the top quartile with the highest quartile corresponding to high risk patients. This cutoff was determined to be 52.5 in the DISPERSE2 Population.
References


109
