# Risk Stratification By Analysis of Electrocardiographic Morphology Following Acute Coronary Syndromes

by

Philip Pohong Sung

S.B., Massachusetts Institute of Technology (2007)

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

February 2009



Author .....
Department of Electrical Engineering and Computer Science
January 16, 2009

Certified by .....
John V. Guttag
Professor, Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by .....
Arthur C. Smith
Chairman, Department Committee on Graduate Theses

# Risk Stratification By Analysis of Electrocardiographic Morphology Following Acute Coronary Syndromes

by

Philip Pohong Sung

Submitted to the Department of Electrical Engineering and Computer Science
on January 16, 2009, in partial fulfillment of the
requirements for the degree of
Master of Engineering in Electrical Engineering and Computer Science

## Abstract

Patients who have suffered an acute coronary syndrome (ACS) are at elevated risk of future adverse events, including fatal arrhythmias or myocardial infarction. Risk stratification—the identification of high-risk patients—is an important step in determining who is most likely to benefit from aggressive treatments.

We propose a new automated risk stratification technique that uses the long-term electrocardiographic data routinely recorded in the days following an ACS. Data obtained from clinical drug trials indicates that our technique, called MV-DF (morphologic variability diagnostic frequencies), can significantly improve prognostication for ACS patients. Patients with MV-DF values in the highest quartile show a more than five-fold elevated risk of death in the 90 days following a non-ST-elevation ACS.

We also propose techniques to construct models of the dynamics of cardiac behavior. Preliminary results suggest that such techniques may be useful for short-term prediction of fatal arrhythmias.

Our results suggest that long-term ECG-based risk assessment techniques—in particular, methods incorporating information about morphologic variability—are an effective and practical way to select appropriate treatment options for cardiovascular disease patients.

Thesis Supervisor: John V. Guttag
Title: Professor, Electrical Engineering and Computer Science

# Acknowledgments

The Industrial Technology Research Institute of Taiwan (ITRI) and the Center for Integration of Medicine and Innovative Technology (CIMIT) generously funded this work.

It has been a privilege to work with John Guttag, who supervised this thesis with great care. John's suggestions and explanations frequently helped me to gain new insights, but he also ensured that I never lost sight of the essentials. It was his optimism in the potential for this research that encouraged me to excel.

Zeeshan Syed's guidance—on matters clinical, technical, and professional—was crucial to this work. Not only did his work lay the foundation for mine, but almost all of this thesis bears some imprint of his ideas.

Collin Stultz shared his extensive knowledge of physiology and clinical practice. Feedback from my frequent meetings with him invariably helped me to improve my work.

Dorothy Curtis was always ready to help in a jam—of which there were quite a few. Dorothy has also been a wonderful academic advisor throughout my time at MIT.

Irene Fan has been a precious friend through both good and bad times. I have drawn much inspiration from her support.

Finally, I am deeply indebted to my family. Little would have been possible without their encouragement, their faith, their sacrifices, and their love.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

Each year, approximately 1.4 million Americans are hospitalized following an acute coronary syndrome (ACS) [17], an incident in which the blood supply to part of the heart muscle is blocked or severely reduced. Acute coronary syndromes are broadly classified into two major categories: unstable angina, in which heart tissue is not permanently damaged, and myocardial infarction (MI, or heart attack), in which it is.

People who have suffered an ACS are at greatly increased risk for future adverse events. In the GUSTO-IIb trial [4], patients who suffered a non-ST elevation MI (NSTEMI)—one particular subclass of MI—had mortality rates of 5.7% within 30 days and 11.1% within one year; for unstable angina patients, the mortality rates were 2.4% and 7.0%, respectively. In addition, 9.8% of NSTEMI patients and 6.2% of unstable angina patients suffered a myocardial (re)infarction within six months.

There are two reasons for this increased risk. First, the underlying cause of the ACS—cardiovascular disease—may cause future heart attacks. Second, ACS often leads to scarred or damaged cardiac tissue, the presence of which can interfere with the normal electrical conduction patterns of the heart. This may lead to arrhythmias, or abnormal heart rhythms, the most severe of which are fatal.

People have different risk levels for cardiovascular disease due to varying risk factors, both hereditary and lifestyle-related. In addition, acute coronary syndromes vary widely in their severity. ACS patients are therefore quite heterogeneous in their

risk for death and adverse cardiac events [4, 11, 53]. One key goal for clinicians is post-ACS risk stratification—assessment of both a patient's risk for a particular adverse event and the expected benefit to that patient of a particular intervention.

Accurate risk assessment is important for identifying patient-appropriate treatment options. The most invasive treatments—which are also usually expensive and risky—are often of the most benefit to high-risk patients. Conversely, patients deemed low-risk may receive quite conservative therapies. For ACS patients, treatments may include anti-arrhythmic drugs, any of a number of surgical or catheterization procedures, or the implantation of a device such as an ICD (implantable cardioverter defibrillator, a device that can detect arrhythmias and administer an electrical shock to restore normal cardiac rhythm). Risk stratification may also be used to decide when additional tests (e.g. imaging procedures) can justify their cost and risk.

Risk prediction for acute arrhythmic events on shorter timescales—minutes or hours—would also be valuable. Such information could be used to allocate resources and attention among patients in a hospital.

Two risk stratification techniques that are commonly employed today are the TIMI risk score and echocardiography. The TIMI risk score [3, 31, 32] incorporates clinical information available at the time of patient presentation, such as patient history and electrocardiographic (ECG) features, to yield an assessment of risk. One component of the TIMI risk score is based on the analysis of short-term ECG recordings—recordings of the heart's electrical activity—which can indicate the extent of an ACS. An echocardiogram [38] is an ultrasound image of the heart that may be used to diagnose certain mechanical abnormalities.

We advocate the adoption of another class of risk stratification techniques, those based on long-term ECG recordings. ECG data is routinely acquired during hospital stays for monitoring purposes and is often collected for long periods of time. It provides a wealth of diagnostic information and captures subtle features of cardiac behavior. Various ECG-based risk stratification techniques have been proposed, including heart rate variability [30], deceleration capacity [7], and T-wave alternans [44]. It has been shown that risk stratification techniques based on long-term ECG

collected in the hours after admission can refine an initial prognosis made at the time of presentation. However, despite their value and easy availability, the use of ECG-based risk stratification is still uncommon.

In this thesis, we present an improved technique for post-ACS risk stratification and risk prediction based on the MV-LF/HF technique proposed by Syed *et al.* [50, 51], which gives a measure of *morphologic variability* (MV). MV-based techniques estimate risk by quantifying heterogeneity in the shapes (morphologies) of heartbeats on an ECG.

At a high level, the process of taking an input ECG signal and computing the MV-LF/HF measure from it can be considered in four steps, as shown in Figure 1-1. ECG data for a patient is segmented into heartbeats and noisy parts of the signal are excluded from further computation (1). A measure of *morphologic distance* (or *morphologic difference*; MD) is computed for each pair of adjacent heartbeats, yielding a sequence of successive morphologic distances—the *MD time series* (2). The MD time series is then summarized by a single number giving a measure of risk in steps (3) and (4).

The primary contributions of this thesis are the following:

- An evaluation of various methods for computing a morphologic difference between two time-series signals. We demonstrate how three techniques originally designed for comparing shapes, sequences, and distributions in a general metric space may be adapted to compare time-series data; we compare them to the technique employed in [50] for comparing heartbeats—a variant of dynamic time-warping (DTW) [35]. All of the techniques are compared in discriminative power and robustness against noise. We then evaluate how well the techniques identify morphologic differences in ECG data that are associated with adverse outcomes. We show that DTW has certain desirable properties and best identifies those particular morphologic differences that are associated with high risk of death.

- The identification and evaluation of a new *diagnostic frequencies* measure, called

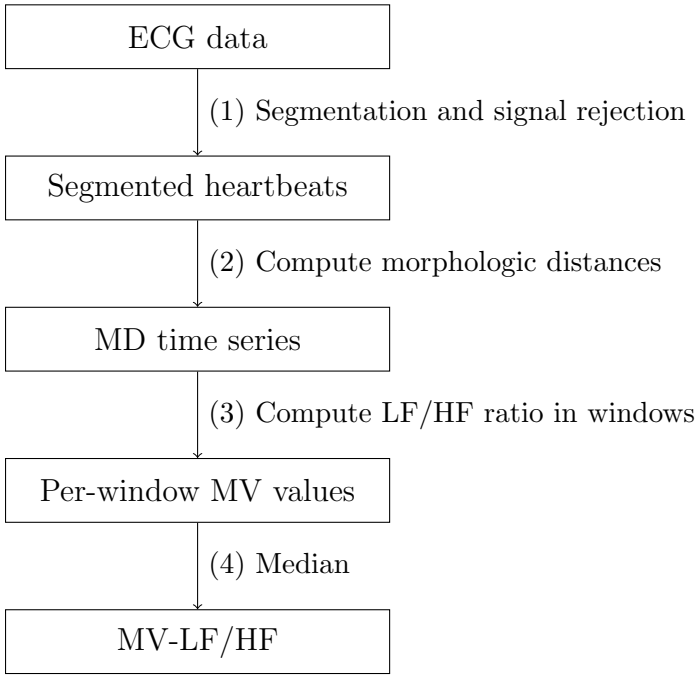



Figure 1-1: Steps in the computation of MV-LF/HF.

MV-DF, that summarizes the MD time series with a measure of risk. The MV-DF measure isolates morphologic change on specific timescales. We developed and trained the technique using data from 764 patients of the DISPERSE-2 (TIMI33) clinical drug trial [13]. We then validated the MV-DF technique on data from 2,302 patients in the placebo group of the MERLIN (TIMI36) trial [33]. In both datasets, for each patient, we had access to continuous ECG recordings at least 24 hours long that started no more than 48 hours following a non-ST-elevation ACS (NSTEACS).

We show that high MV-DF values are significantly associated with increased risk. The quartile of patients with the highest MV-DF shows a more than five-fold increase in risk of death in the 90 days following NSTEACS. Patients with high MV remain at significantly elevated risk even when controlling for other risk variables. MV-DF identifies patients at elevated risk of death in the 90 days following NSTEACS with a c-statistic of 0.77, as compared to 0.72 for the MV-LF/HF measure proposed in [50]. Our results indicate that MV-DF is the

best Holter-based risk stratification technique known today for predicting risk following NSTEACS.

- The development of techniques for modeling the dynamics of morphologic change using hidden Markov models (HMMs). These techniques operate on the MD time series and other sequences that can be derived from ECG data. We argue that such models might yield important information for short-term risk prediction. Some preliminary results suggest that these models may be used to predict acute arrhythmic events, although further study is needed to validate this claim.

The remainder of this thesis is organized as follows. Chapter 2 provides background on cardiac function and acute coronary syndromes. Chapter 3 gives an overview of existing post-ACS risk stratification methods, with a focus on ECG-based techniques. Chapter 4 evaluates various techniques for computing the morphologic difference between two signals. In Chapter 5, we develop the MV-DF measure for summarizing measurements of morphologic heterogeneity to yield an evaluation of patient risk. Chapter 6 proposes methods for modeling cardiac dynamics and extracting features from those models for risk stratification and prediction of imminent death. Chapter 7 concludes with a summary and a discussion of future work.

# Chapter 2

# Cardiac function and disease

This chapter gives an overview of basic cardiac function and electrophysiology, as well as coronary artery disease and its consequences. The interested reader is advised to consult Lilly [28] for a more comprehensive treatment of these topics.

## 2.1 Cardiac function

The heart consists of four muscular chambers: the right and left ventricles, and the right and left atria, which deliver blood to their respective ventricles. The right atrium and ventricle collect deoxygenated blood from the body and deliver it to the lungs where carbon dioxide is exchanged for oxygen. The left atrium and ventricle collect oxygenated blood from the lungs and deliver it to the rest of the body. The aorta, the artery that carries blood out of the left ventricle, branches into smaller arteries which lead to different parts of the body. The first arteries to split off are the left and right coronary arteries, which supply the heart's own muscles with oxygenated blood.

The heart circulates blood via the coordinated action of the various parts of the *myocardium*, or heart muscle tissue. In each heartbeat, the sequence of contraction of the atria followed by contraction of the ventricles pumps blood through the circulatory system. This process is coordinated by the orderly propagation of electrical impulses which cause *depolarization*, a temporary reversal of the cell membrane volt-

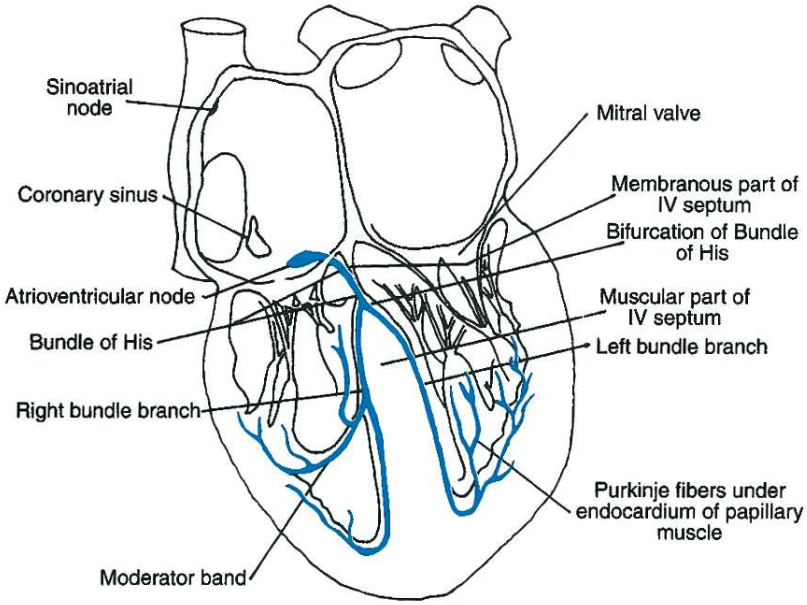



Figure 2-1: Main components of the cardiac conduction system. From Lilly [28].

age, throughout the myocardium. Depolarization of cells in the myocardium causes those cells to contract.

The conduction system of the heart is pictured in Figure 2-1. A wave of depolarization begins in the sinoatrial (SA) node, which contains *pacemaker cells* that spontaneously produce electrical impulses. From there, depolarization spreads throughout the atria, causing them to contract. The wave then reaches the atrioventricular (AV) node. This is the only connection between the conduction systems of the atria and the ventricles, which are elsewhere separated by insulating fibrous tissue. The AV node consists of specialized tissue that conducts slowly, so it delays electrical impulses that pass through it for a short time (about 0.1 sec). This delay is important for efficient circulation because it allows the atria to completely empty their blood into the ventricles before the ventricles begin to contract. Finally, the wave of depolarization spreads throughout the ventricles by way of the Bundle of His and the left and right bundle branches, causing the ventricles to contract.

### 2.1.1 Electrophysiology

The membrane of a myocardial cell contains *ion channels*, specialized proteins that span the cell membrane and regulate the movement of specific ions across the membrane [28]. Different types of ion channels are selective for different kinds of ions, allowing only ions of a specific type to pass. In addition, the conformation of ion channels changes with the membrane voltage difference to allow (or block) the diffusion of ions. Ion channels act as voltage-regulated passive gates for ions: the flow of ions through ion channels is determined by the concentration gradient and by the electrical potential difference (voltage) across the membrane. Cell membranes also contain active *ion pumps*, which consume energy in the form of adenosine triphosphate (ATP) to pump ions across a membrane against their natural gradient.

In a cardiac cell at rest, the ion channels and ion pumps together maintain a *resting potential* of $-90$ mV inside the cell by selectively moving $Na^+$ and $Ca^{++}$ ions out of the cell and $K^+$ ions into the cell. If the membrane voltage goes above approximately $-70$ mV, an *action potential* begins. Some sodium ion channels open, allowing $Na^+$ ions to enter the cell, raising the potential inside, causing more sodium ion channels to open, and so on, creating a positive feedback loop. The cell quickly (within milliseconds) becomes depolarized and reaches a peak voltage of slightly more than 0 mV. This voltage is high enough to raise the membrane voltage in a nearby area of the cell or a neighboring cell, causing the action potential to propagate.

At the peak voltage, the sodium channels close and remain inactivated until the cell has returned to resting potential (as described below). In healthy myocardial tissue, this refractory period prevents recently depolarized cells from depolarizing again, regardless of the membrane voltage. This ensures that the wave of depolarization propagates forward and never backward.

The cell now begins the process of *repolarization* in order to prepare for the next action potential. When the membrane voltage becomes high enough, the potassium and calcium channels open, allowing $K^+$ and $Ca^{++}$ ions to flow out of and into the cell, respectively. Calcium ions entering the cell during this phase activate a pathway

that induces the physical contraction of cardiac muscle cells. Finally, the original concentrations of each ion, and the resting potential, are restored by ion pumps in order to prepare the cell for another action potential.

## 2.2 Electrocardiogram

An electrocardiogram (ECG) is a recording of the electrical activity of the heart. ECG data is routinely recorded for hospitalized patients, as it is useful for both monitoring them and diagnosing conditions such as ACS or arrhythmias. ECG can be acquired inexpensively and with minimal invasiveness; a Holter monitor (a portable ECG device worn on a patient) can record data for 24 hours or more. As such, ECG data is useful for analysis of rare and noisy phenomena. Depending on the setting and on the reason for the recording, varying numbers of electrodes may be used in order to capture fewer or more channels of data. Typical ECG monitors record between 3 and 12 channels. MV, as well as some other ECG-based risk measures, can be computed from just a single channel of Holter data.

A cardiac muscle cell at rest maintains a negative voltage with respect to the outside of the cell. While at rest, the surface of the cell is uniformly charged with a positive voltage, but during depolarization, this voltage decreases and may even become negative. Consequently, when depolarization is propagating through a cell, there exists a potential difference on the membrane between the part of the cell that has been depolarized and the part of the cell at resting potential. After the cell is completely depolarized, its membrane is uniformly charged again (although now negatively instead of positively).

These changes in potential, summed over many cells, can be measured by electrodes placed on the skin. For any pair of electrodes, a voltage is recorded whenever the direction of depolarization (or repolarization) is aligned with the line connecting the two leads, with the sign indicating the direction of depolarization. Multiple electrodes along different axes are used so that the average direction of depolarization, as a three-dimensional vector, can be reconstructed from the ECG tracings.

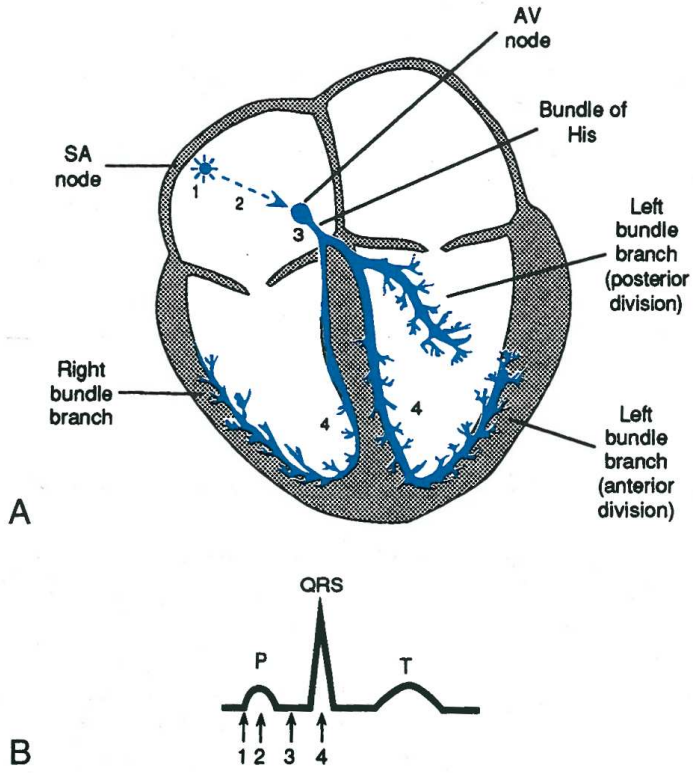



Figure 2-2: Cardiac conduction pathway, with corresponding waveforms on the ECG recording. The electrical impulse begins at the SA node (1). The wave of depolarization traverses the atria (2). Conduction is slowed at the AV node (3). The wave of depolarization traverses the ventricles (4). From Lilly [28].

Three major segments can be identified in a normal ECG, corresponding to different parts of the action potential. Schematics of the cardiac conduction pathway and a typical ECG recording are shown in Figure 2-2. The P wave is associated with depolarization of the atria. The QRS complex is associated with depolarization of the ventricles. The T wave is associated with repolarization of the ventricles. The QRS complex is larger than the P wave because the ventricles are much larger than the atria. The QRS complex coincides with repolarization of the atria, which is therefore usually not seen on the ECG. The T wave has a larger width and smaller amplitude than the QRS complex because repolarization takes longer than depolarization.

## 2.3 Atherosclerosis

Atherosclerosis is a disease of the arteries characterized by the accumulation of fat, cholesterol, cells, and other substances within the arterial wall. Such deposits may form lesions in the arteries known as plaques. When this occurs in a coronary artery, this condition is known as coronary artery disease. Coronary artery disease is dangerous because it interferes with the biochemical pathways that maintain healthy vessel walls; this can lead to complications such as an ACS. Coronary artery disease can also reduce the cross-section (and therefore effective capacity) of the artery.

In healthy arteries, the endothelium—the thin layer of cells closest to the bloodstream—performs a number of functions to preserve the integrity of the vessel wall and promote normal blood health. The endothelium forms a tight barrier so that large molecules and cells in the bloodstream do not enter the tissue of the vessel wall. Endothelial cells normally inhibit *thrombosis*—blood clotting—by synthesizing certain antithrombotic substances. They also inhibit the immune system inflammatory response unless an injury or infection has actually occurred. When endothelial cells become damaged, these functions may be compromised. A damaged endothelial layer is more permeable and is more likely to release chemicals that inappropriately induce clotting or inflammation.

One critical factor in atherosclerosis is the presence of low-density lipoproteins (LDLs), biochemical assemblies that bind to cholesterol—which is not water-soluble by itself—so that it can be transported in the bloodstream. Cholesterol carried by LDL is commonly referred to as "bad cholesterol" for its role in contributing to cardiovascular disease. When the endothelium is damaged, LDL can diffuse across the endothelium into the *intima*, the layer of tissue just outside the endothelium. In the intima, LDL participates in a pathway that may lead to the formation of atherosclerotic plaques. LDLs may become oxidized or otherwise modified and cause the endothelium to activate the inflammatory response. This causes leukocytes (white blood cells) to enter the intima from the bloodstream and ingest the LDL. When the leukocytes become loaded with LDL, they are known as *foam cells*, and they

manifest as fatty streaks in the arteries. These fatty streaks are the first visible effects of atherosclerosis, although they are asymptomatic and do not cause significant narrowing of the arteries.

Subsequently, foam cells emit chemical signals that induce the migration of smooth muscle cells into the intima from the surrounding muscle layer. Once they are established, smooth muscle cells reproduce within the intima and synthesize proteins that produce a *fibrous cap* around the foam cell layer. The combination of these processes causes the intima to expand in a lesion, shrinking the artery's effective diameter.

Atherosclerosis is associated with a number of complications. The accumulated deposits can reduce the blood vessel's cross-sectional area enough to impair circulation; this may cause *stable angina*, recurrent chest pain that occurs with physical activity or stress but goes away with rest. The fibrous cap of the plaque can rupture, exposing the foam cell core to the blood. Foam cells release chemicals that induce thrombosis, causing a blood clot to form over the site of rupture which may partially or completely occlude the artery. In a coronary artery, this may lead to an acute coronary syndrome, which can cause myocardial cell death and may be life-threatening in the absence of prompt medical attention. The blood clot may also break off from the original site and travel to another part of the body (this is called an embolism). This can cause a stroke if it blocks an artery that supplies the brain.

### 2.3.1 Treatment

The presence and extent of atherosclerosis may be diagnosed by angiography, a procedure for imaging blood vessels. In this procedure, an x-ray opaque dye is injected into a patient's bloodstream near the coronary arteries using a catheter. High-speed x-ray imaging can then be used to identify narrowing of the blood vessels.

Patients with severe myocardial ischemia (a shortage of blood flowing to cardiac tissue) may require revascularization, which may refer to any of a number of surgical or catheter procedures to improve blood flow to the heart.

In percutaneous transluminal coronary angioplasty (PTCA, also called balloon angioplasty), a small balloon is threaded to the blockage site through a catheter. The

balloon is inflated so that it widens the artery. In a PTCA procedure, a patient may also receive a stent, a small wire mesh tube. The stent is unfolded in the artery to hold the artery open and is left there permanently. Some stents are coated with drugs that are released over time and prevent additional plaque formation from closing the artery again.

In the case of more extensive blockage, other procedures may be used to widen the artery by removing the plaque. In laser angiography, a laser is used to vaporize plaque at the blockage. In an atherectomy, a rotating drill bit is used to grind the plaque away. These procedures may be used in combination with balloon angioplasty or stents.

In some cases, a patient may receive a coronary artery bypass graft (CABG). A blood vessel from another part of the body is transplanted with one end attached to the aorta and the other end attached just downstream of a blockage. Blood flow can bypass the blocked blood vessel by way of the transplanted vessel.

### 2.3.2 Risk factors

Various attributes have been identified that are associated with increased risk of atherosclerosis. Older people, males, and people with a family history of coronary disease are at elevated risk. A number of other controllable risk factors can be mitigated by lifestyle changes and drug treatments, as described below.

Maintaining appropriate blood cholesterol levels is a major component of reducing risk. High LDL cholesterol levels are associated with increased risk of coronary artery disease, since excess LDL may form deposits in the intima as previously described. In contrast, high-density lipoproteins (HDL, also known as "good cholesterol") have been shown to protect against atherosclerosis, since HDL can remove excess cholesterol from circulation.

Obesity is associated with high LDL cholesterol levels and is a risk factor for atherosclerosis. Diet and exercise can reduce this risk. Many drug therapies also aim to reduce LDL levels. The most effective of these are the HMG CoA reductase inhibitors ("statins"), which accelerate the expression of pathways that remove LDL

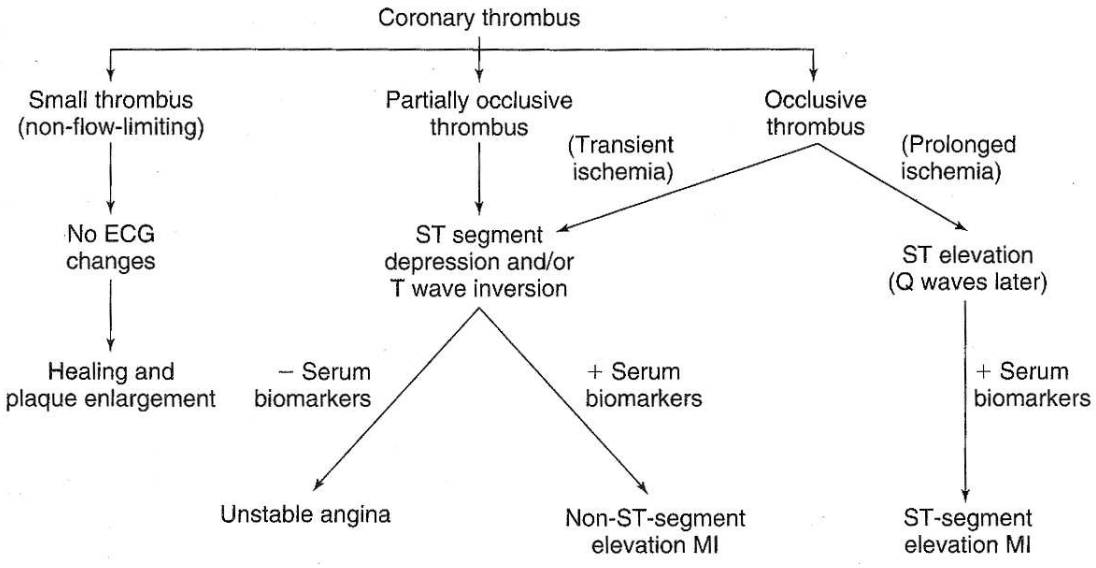



Figure 2-3: Consequences of coronary thrombosis. From Lilly [28].

from the bloodstream. Statins have been shown not only to lower LDL levels but also to reduce risk for strokes and ACS [12, 36].

Other factors have been implicated in the compromise of the endothelium, making it more permeable to lipids. Cigarette smoking weakens the endothelium by introducing various toxic chemicals. Hypertension (high blood pressure) directly puts physical strain on the endothelium and induces the production of foam cells. Diabetes is often associated with impaired endothelial function (as well as with reduced HDL).

## 2.4 Acute coronary syndromes

An acute coronary syndrome (ACS) is an event in which blood supply to part of the myocardium is blocked or severely reduced. The most common symptom of ACS is unusual and unprovoked chest pain. An ACS is usually caused by the rupture of an atherosclerotic plaque producing a blood clot within a coronary artery. This leads to ischemia or to cell death in the myocardium.

Various subclassifications of ACS are distinguished by the presence of myocardial necrosis (cell death) and by ECG diagnosis. An overview of these subclassifications is shown in Figure 2-3.

Unstable angina refers to an ACS event in which necrosis does not occur, while myocardial infarction (MI) refers to one in which it does. ACS's are also subclassified based on the extent of the occlusion, which can be inferred from ECG recordings. An ECG showing elevation in the ST segment is indicative of complete occlusion of an artery and necrosis (and therefore, myocardial infarction). Such patients are given a diagnosis of ST-elevation MI (STEMI).

Non-ST-elevation ACS (NSTEACS) is indicative of partial occlusion of an artery and is a less severe condition. NSTEACS may be diagnosed by the presence of certain ECG irregularities (ST depression or T wave inversion). Two subclasses of NSTEACS, unstable angina and a non-ST-elevation MI (NSTEMI), are distinguished by whether necrosis occurs. Blood tests are used to determine levels of two *serum biomarkers*—cardiac-specific troponin and creatine kinase MB (CK-MB)—which are chemicals released into the bloodstream when myocardial necrosis occurs.

Treatment for NSTEACS focuses on inducing the dissolution of blood clots by natural pathways (via aspirin or heparins), and on reducing ischemia by lowering the heart's oxygen demand and raising oxygen supply. Drugs that dilate blood vessels (nitrates) or lower heart rate ($\beta$-blockers) are commonly employed. STEMI patients may benefit from the same treatments, but they also receive more aggressive thrombolytic drugs to break down blood clots and restore normal blood flow. The revascularization procedures described in Section 2.3.1 may also be conducted, either in an emergency or after the patient's condition has stabilized.

## 2.5 Arrhythmias

An ACS may leave damaged or scarred heart tissue, which can interfere with the heart's electrical conduction system. This may lead to arrhythmias, or abnormal heart rhythms.

Arrhythmias can be benign, but the most severe of them may be fatal. One of the most serious of these is ventricular fibrillation (chaotic and rapid twitching of the ventricles). This may lead to cardiac arrest (failure of the heart to circulate blood

around the body effectively), and, if not promptly treated, death.

Arrhythmias are broadly classified into two groups: tachyarrhythmias (those associated with increased firing rate) and bradyarrhythmias (those associated with decreased firing rate). They may arise from irregularities in the generation of action potentials or in the conduction of action potentials through the myocardium. The generation of action potentials is usually the job of the SA node. In abnormal situations, other parts of the heart may start to spontaneously depolarize (leading to tachyarrhythmias) or impulse generation may be impaired (leading to bradyarrhythmias). Typically, a bradyarrhythmia stemming from impaired impulse generation is not a life-threatening situation, because the myocardium contains multiple regions of tissue that have the potential to spontaneously depolarize; these act as "backup" pacemakers if impulse generation at the SA node becomes too slow.

Myocardium that has been damaged by ischemia may contain small islands of tissue that is electrically unexcitable [19]. Such tissue may also merely conduct impulses slowly or only in one direction. Small damaged regions may lead to beat-to-beat variations in the conduction pathway, while larger damaged regions may cause arrhythmias.

Small regions of impaired tissue in the conduction pathway may lead to bifurcation of the conduction path, which is associated with slow and anisotropic conduction [24]. Such conduction patterns may vary from beat to beat for two reasons. First, they may be due to random and unstable variations in damaged myocardium. In addition, if a region of cells is slow to repolarize, then a subsequent wave of depolarization will be forced to take a different route around it. It is these beat-to-beat changes that we believe are measured by the MV method. The consequences of minor conduction inhomogeneity are not well understood, but there is some evidence that they may degenerate into, and have predictive value for, ventricular arrhythmias [8] and other adverse events.

More major conduction pathway alterations can lead to arrhythmias. A *conduction block* arises when a region of unexcitable tissue stops the wave of depolarization entirely, preventing part of the heart from contracting. *Reentry* is a phenomenon in

which a wave of depolarization travels around a closed-loop conduction path, sometimes around an island of unexcitable tissue. The wave of depolarization becomes self-sustaining, leading to a tachyarrhythmia.

Ventricular fibrillation occurs due to a reentrant conduction pattern in the ventricles. Reentry in the atria may lead to atrial flutter (rapid but regular atrial beats), which may degenerate into atrial fibrillation (chaotic atrial activity).

When arrhythmias are deemed dangerous, they may be treated by drugs that raise or lower the heart rate, or by other more invasive interventions. A persistent bradyarrhythmia may be treated by the implantation of an artificial pacemaker. An artificial pacemaker applies electrical stimulation to induce depolarization at a desired rate, preempting the heart's (slower) natural pacemaker.

A tachyarrhythmia caused by reentry may be an emergency situation as it may lead to cardiac arrest. Such a condition is treated by the application of an electrical current across the chest. This depolarizes the entire myocardium so that reentrant patterns are interrupted. The heart's natural pacemaker then assumes control of heart rhythm. This technique is called *defibrillation* in the case of ventricular fibrillation. In other cases, the discharge has to be synchronized with the QRS complex in order to avoid *inducing* ventricular fibrillation; in these cases, this technique is called *cardioversion*.

Patients at high risk of tachyarrhythmias may receive an implantable cardioverter-defibrillator (ICD), a device similar in appearance to a pacemaker that can detect aberrant heart rhythms and apply electrical shocks to restore normal rhythm.

# Chapter 3

# Post-ACS risk stratification

This chapter provides background information on post-ACS risk stratification methods. We consider the TIMI risk score (TRS), echocardiography, and long-term ECG-based techniques. The TRS [3, 31, 32] provides a general assessment of risk based on clinical variables that can easily be obtained at the time of admission. The variables considered by the TRS represent the most significant independent predictors of risk—selected from a larger set of 12 clinical variables—that were identified in the TIMI11B trial [3]. Echocardiography is a technique for imaging the heart using ultrasound; it yields information about blood flow in the heart as well as the shape of the heart. The electrocardiogram is a valuable tool for the diagnosis and prognosis of ACS patients. As described in Section 2.4, ECG data may be used to diagnose the severity of an ACS at the time of presentation and is typically used to guide immediate treatment.

A variety of methods have been proposed that assess risk based on automated analysis of long-term ECG data collected in the hours or days following admission. Such data is routinely collected during a patient's stay and therefore these additional risk assessments can be obtained at almost no additional cost. We discuss three ECG-based methods that have been proposed in the literature: heart rate variability (HRV) [30, 26], deceleration capacity (DC) [7], and morphologic variability [51, 50, 48]. Each of these measures has been shown to correlate with risk of various adverse events in the period following an ACS.

One additional long-term ECG-based risk stratification technique, T-wave alter-

nans (TWA) [44], has also received some attention. However, evaluating TWA requires the use of specialized equipment and requires patients to complete specific maneuvers in order to elevate their heart rate. Unlike the other long-term ECG risk measures we consider, TWA cannot be computed using regular Holter monitor data. It is unlikely that TWA (in its current form) could be used widely for risk stratification in general populations, and as such we do not consider it further in this thesis.

Each of the techniques considered here incorporates some information about a patient and yields a number that can be used to estimate the patient's risk. For example, higher values of the TRS are associated with higher risk. We evaluate the utility of these risk stratification techniques using two metrics. The c-statistic, or the area under the receiver operating characteristic (AUROC) [22], identifies the degree to which progressively higher values of the variable are associated with higher risk of adverse events. In practice, treatments may be chosen based only on the dichotomized value of a particular variable—whether its value is greater than or less than some threshold value. In this case, the Cox proportional hazards model [16] is used to estimate the *hazard ratio*, the ratio of the instantaneous rates of death in the two groups.

## 3.1 TIMI risk score

The TIMI risk score (TRS) [3, 31, 32] is a simple risk stratification technique that incorporates clinical variables easily acquired at the time of admission and is designed to be evaluated without the use of a computer. It can therefore be used in triage and immediate decision-making with regard to treatment options. The GRACE [21] and PURSUIT [10] risk scores perform similar functions.

The TRS considers the following 7 predictor variables:

- Age 65 years or older
- At least 3 risk factors for coronary artery disease among the following: hypertension, hypercholesterolemia, diabetes, family history of coronary artery

disease, or being a current smoker

- Prior coronary stenosis (narrowing of an artery) of 50% or more
- ST-segment deviation on ECG at presentation
- Severe anginal symptoms (at least 2 anginal events in prior 24 hours)
- Use of aspirin in prior 7 days
- Elevated serum cardiac markers (CK-MB or troponins)

One point is counted for each variable that is observed, and the TIMI risk score is the total number of points (between 0 and 7). The set of variables was obtained by selecting independent prognostic variables from a set of 12 prospective clinical variables after a multivariate logistic regression [3]. The other 5 variables that were considered but not included in the TRS were: prior MI, prior coronary artery bypass graft (CABG), prior angioplasty (PTCA), prior history of congestive heart failure, and use of IV heparin within 24 hours of enrollment.

The TIMI11B and ESSENCE trials [3] showed that a higher TRS is associated with higher rates of adverse events—defined as death, MI, or severe recurrent ischemia—in the 14 days following the initial event. Those trials also demonstrated that the TRS can be used to identify groups of patients who respond differently to a treatment. In particular, they showed that prescribing enoxaparin (an antithrombotic drug) instead of unfractionated heparin reduced risk more in patients with high TRS than in patients with low TRS.

## 3.2 Echocardiography

Echocardiography (often referred to as simply "echo") is the use of ultrasound techniques to create an image of the heart. An echocardiogram can yield structural information about the heart and its valves, and about blood flow through the heart. Magnetic resonance imaging (MRI), nuclear imaging, and the angiographic techniques described in Section 2.3.1 can provide some of the same information.

In particular, an echocardiogram is frequently used to assess left ventricular function [27]. The left ventricle is the largest chamber of the heart and is responsible for pumping oxygenated blood to the body, so proper left ventricular function is important to the health of the body. If the myocardium is damaged in an MI, then the heart is weakened and the left ventricle's capacity to pump blood may be reduced. This is associated with higher risk of congestive heart failure and fatal arrhythmias [34]. One measure of left ventricular function is the left ventricular ejection fraction (LVEF), the fraction of the blood volume ejected from the left ventricle during systole (the contraction phase of the heartbeat). An echocardiogram may be used to estimate the volume of blood in the heart before and after systole, and therefore the LVEF:

$$\text{LVEF} \equiv \frac{(\text{LV volume before systole}) - (\text{LV volume after systole})}{(\text{LV volume before systole})}$$

A healthy heart has an LVEF of between 0.55 and 0.75 [28]. Patients with an LVEF of below 0.40 are considered to have left ventricular dysfunction [2]. As discussed above, a weakened myocardium may stem from an MI, so low LVEF often correlates with other disorders of the heart. The MADIT II trial [34] demonstrated that those with low LVEF may benefit more from aggressive treatments—in particular, the implantation of a defibrillator.

## 3.3 Heart rate variability

One class of ECG-based risk stratification techniques that has been discussed extensively in the literature is based on measurements of heart rate variability (HRV) [30, 26]. The theory underlying HRV-based techniques is that in healthy people, the body should continuously compensate for changes in oxygen demand by changing the heart rate; therefore, a heart rate that changes little suggests that the heart or its control systems are not actively responding to stimuli. HRV-based measures attempt to quantify the amount of change in a patient's instantaneous heart rate in order to yield an estimate of risk.

Heart rate is primarily modulated by the autonomic nervous system, which comprises the the sympathetic and parasympathetic nervous systems. The parasympathetic nervous system's effects on heart rate are mediated by the release of acetylcholine by the vagus nerve, which lowers the heart rate. The sympathetic nervous system's effects are mediated by the release of epinephrine and norepinephrine, which raise heart rate. In particular, it is decreased vagal or parasympathetic modulation (i.e. reduced down-regulation of heart rate) that is thought to be linked to increased risk of death [9, 43]. One possible explanation for this is that the heart raises the heart rate in order to maintain a steady blood supply when its pumping capacity per beat has been reduced by ischemia or infarction. This fact is also consistent with the observation that tachyarrhythmias are often life-threatening but bradyarrhythmias are rarely so. However, there is currently no consensus on whether low HRV is simply a correlate of poor outcomes or whether it is actually part of some mechanism that leads to arrhythmias [30].

Under resting conditions, changes in instantaneous heart rate are dominated by the effects of vagal modulation anyway [14, 30]. Therefore, total variation in heart rate is a good proxy for variation due to vagal stimulation.

In general, HRV-based techniques first compute the sequence of intervals between heartbeats, which may be determined from ECG tracings. These are typically obtained by counting from one QRS complex to the next [30] since the QRS complex is the most prominent feature of a heartbeat. If only heartbeats resulting from normal depolarization of the SA node are considered, then this sequence is termed the NN (for normal-to-normal) series. One of a number of methods is then used to summarize this series with a single number indicating the amount of heart rate variability. These HRV measures can be roughly divided into time-domain, frequency-domain, and nonlinear measures. Malik [30] provides a more complete overview of HRV metrics.

Frequency-domain HRV methods are believed to quantify primarily the effects of vagal stimulation. They rely on the fact that vagal and sympathetic nervous system activity can be distinguished because they are mediated by biochemical pathways that are associated with different timescales [30]. One frequency-domain metric, LF/HF, is

defined in terms of the power spectral density (PSD) of the NN series. In the analysis here, we compute the PSD using the Lomb-Scargle periodogram [29], a technique for estimating the frequency content of a signal that is sampled at irregular intervals. LF/HF is the ratio of the total power in two frequency bands of the spectrum, "low frequency" (LF) and "high frequency" (HF):

$$\text{HRV-LF/HF} \equiv \frac{\text{(Power between 0.04 and 0.15 Hz)}}{\text{(Power between 0.15 and 0.4 Hz)}}.$$

The HF band is believed to reflect vagal modulations, but there is disagreement as to whether the LF band reflects sympathetic nervous system modulations or some other phenomenon [30]. Because of this uncertainty, there is no widely accepted physiological interpretation of the LF/HF ratio. Although it is primarily decreased vagal (HF) activity that is thought to be dangerous, the sympathetic and parasympathetic nervous systems do inhibit each other to some degree. As a result, one interpretation is that the LF and HF energies, and the LF/HF ratio, all reflect the amount of vagal activity.

The LF/HF ratio is computed for 5-minute windows, as in [30], and the median value across windows is used as the LF/HF value for that patient. Patients with high HRV-LF/HF—that is, patients whose high frequency (vagal) heart rate modulations are small compared to their low frequency (sympathetic) heart rate modulations—are considered to be at risk.

Time-domain HRV methods give a measure of total variation in heart rate. Commonly considered time-domain HRV metrics include SDNN (standard deviation of NN intervals) and SDANN (standard deviation of mean NN interval over five-minute windows of the recording). Other time-domain measures include:

- ASDNN, the mean of the standard deviation of NN intervals within five-minute windows.
- RMSSD, the root-mean-square of differences of successive NN intervals.
- HRVI (HRV triangular index), the maximum number of items in a single bin

in a histogram of NN intervals (using a standard bin width of 1/128 s), divided by the total number of NN intervals.

- pNN50, the fraction of differences of successive NN intervals that exceeded 50 ms.

In our own experiments, we found that HRV-LF/HF performed better at identifying patients at high risk of death post-ACS than any of the time-domain metrics. These results are consistent with earlier findings reported by the Framingham Heart Study [52]. We believe that frequency-based methods may be more robust in general because high-frequency noise in ECG recordings does not interfere with measurements of the relatively low-frequency components that are physiologically relevant.

## 3.4 Deceleration capacity

Deceleration capacity (DC) is an extension of work on heart rate turbulence [6]. Like HRV, DC attempts to measure impaired vagal modulation of heart rate, which is believed to be associated with high risk. The theory underlying the DC technique is that vagal activity can be distinguished from sympathetic activity because vagal activation causes heart rate deceleration while the sympathetic nervous system causes heart rate acceleration [7].

To compute DC, we begin with the RR interval sequence $A[k]$. Each RR interval longer than the preceding one is identified as an *anchor*. Segments of intervals around each anchor are identified, so that if $k_i$ is the index of the $i$th anchor, then $S_i[k] \equiv A[k_i + k]$. (For example, $S_i[0]$ represents all the anchors, $S_i[1]$ represents every beat after an anchor, etc.) Let $X[k]$ be the average of the $S_i[k]$'s. Then DC is computed as follows:

$$\mathrm{DC} \equiv \frac{(X[0] + X[1]) - (X[-1] + X[-2])}{4}.$$

Roughly speaking, DC measures the magnitude of the typical beat-to-beat deceleration. Low values of DC have been linked to higher risk of death [7].

## 3.5 Morphologic variability

HRV and DC are designed to measure changes in the heart rate mediated by the autonomic nervous system. Morphologic variability (MV) is a technique designed to measure variations in the shape of the heartbeat that are related to myocardial instability [50]. MV measures beat-to-beat conduction path changes that may be indicative of myocardial damage [24], as discussed in Section 2.5.

At the core of the MV method is morphologic distance (MD), a measure of the difference in morphology of two heartbeats. MD is computed using a variant of dynamic time-warping (DTW) [39]. The DTW method, described in detail in Chapter 4, aligns the samples of the two heartbeats in a way that minimizes the total energy difference between aligned samples. This alignment process permits comparison of corresponding parts of the two beats even in the presence of timing inconsistencies. MD is the sum-of-squares energy difference under this optimal alignment and contains a term for each pair of aligned samples. Because the alignment is longer (leading to a larger morphologic distance) if the sequences differ in their timing characteristics, the MD measure reflects differences in amplitude as well as inconsistencies in timing.

To compute MV, the morphologic distance between each pair of consecutive beats is computed; the resulting series of morphologic distances is referred to as the MD time series. As with HRV, a frequency-domain measure can then be used to evaluate the amount of variability in this signal. The MV-LF/HF measure is defined analogously to HRV-LF/HF, but in terms of the spectrum of the MD series:

$$\text{MV-LF/HF} \equiv \frac{\text{(Power between 0.04 and 0.15 Hz)}}{\text{(Power between 0.15 and 0.4 Hz)}}.$$

As with HRV, MV-LF/HF is computed for each five-minute window and the median value is used to characterize the patient's morphologic variability. Patients with low values of MV-LF/HF (that is, patients who have significant high-frequency variations in their MD time series) have been shown to be at increased risk for adverse events following an MI [51, 50].

In Chapters 4 and 5 we develop an alternative to MV-LF/HF that quantifies

morphologic variability to provide a better measure of risk. Chapter 4 begins by examining alternatives to DTW for computing morphologic distance and the MD time series. Then, Chapter 5 introduces the MV-DF risk measure which (like MV-LF/HF) is based on frequency analysis of the MD time series.

# Chapter 4

# Quantifying morphology change

The MV risk measure [50] is an indicator of morphologic heterogeneity in the ECG signal; its computation requires, as an intermediate step, the computation of the morphologic distance time series. For each element of the MD time series, two successive heartbeats (represented as vectors of samples) are compared and the difference—or "distance"—between their shapes is quantified with a single number. In this chapter, we evaluate several different strategies for quantifying differences in the morphology of time series. In order to compute morphologic differences effectively, these strategies have mechanisms to account for variable time skew in the observed data. A compressed version of this material can be found in [46].

Simple metrics, such as Euclidean distance, are not suitable for signals that have variable amounts of time skew (or, indeed, for signals of different lengths). For example, the waves that appear in an ECG are caused by particular sequences of events. A delay in one phase of the ECG may displace subsequent parts by shifting them forward in time. In this case, comparing samples by their timing alone may cause parts of the signals associated with different phenomena to be compared, leading to a poor estimate of similarity. Methods that relate corresponding parts of two signals before measuring differences are useful in a variety of settings (e.g. [37, 50]).

We consider five morphologic distance measures for signals represented as real-valued sequences of time-amplitude pairs, $(t_1, v_1), (t_2, v_2), \ldots, (t_n, v_n)$. We consider two variations of dynamic time-warping [35], a technique conventionally applied to

sequential data such as time-series data. The other methods are adaptations of earlier methods for comparing shapes in a general metric space: Earth Mover's Distance (EMD) [41], Fréchet Distance (FD) [5], and Hausdorff Distance (HD) [23].

We first perform two general evaluations of each measure. We examine the distances produced when the measure is used to discriminate one shape from another. We then examine the robustness of the measure when its inputs are corrupted by noise. Both analyses are performed on two datasets: synthetic data from the cylinder-bell-funnel problem [42] and ECG data from the PhysioNet MIT-BIH dataset [20].

We then evaluate the quality of risk stratification of MV-LF/HF when the morphologic distance computation is replaced with each of the five techniques in turn.

The remainder of this chapter is organized as follows. Section 4.1 summarizes the five measures and the adaptations we propose to them. Section 4.2 describes the evaluation procedure and results. Section 4.3 concludes with a discussion.

## 4.1 Techniques

### 4.1.1 Dynamic Time-Warping

The Dynamic Time-Warping (DTW) distance metric [35] computes the distortion needed to align two time series. An *alignment* of two sequences $A$ and $B$, of length $m$ and $n$ respectively, is a sequence of integer pairs of some length $k$,

$$(\phi_A[1], \phi_B[1]), (\phi_A[2], \phi_B[2]), \ldots, (\phi_A[k], \phi_B[k]), \tag{4.1}$$

where $\phi_A$ and $\phi_B$ satisfy the following boundary conditions:

$$\phi_A[1] = 1, \qquad \phi_B[1] = 1,$$
$$\phi_A[k] = m, \qquad \phi_B[k] = n,$$

as well as the following continuity conditions:

$$\begin{aligned} \phi_A[j] \leq \phi_A[j+1] \leq \phi_A[j]+1 \qquad & \forall j : 1 \leq j < k, \\ \phi_B[j] \leq \phi_B[j+1] \leq \phi_B[j]+1 \qquad & \forall j : 1 \leq j < k. \end{aligned} \tag{4.2}$$

Intuitively, each ordered pair $(\phi_A[i], \phi_B[i])$ matches two elements to be aligned. The DTW distance between $A$ and $B$ is the minimum sum-of-squares difference between pairs of matched elements under all allowable alignments,

$$\mathrm{DTW}(A,B) \equiv \min_{\phi_A,\phi_B} \sum_{i=1}^{|\phi_A|} \left(A[\phi_A[i]] - B[\phi_B[i]]\right)^2, \tag{4.3}$$

where we use $|\phi_A|$ to denote the length of the sequence $\phi_A$—i.e. what we denote $k$ in (4.1).

DTW captures both amplitude and timing differences between the signals. Timing differences are captured by an increase in the length $k$ of the alignment (i.e. a larger number of summed terms).

DTW may be computed in an efficient manner using dynamic programming [15], as follows. Suppose we define the function $D(i,j)$ as the minimum cost of aligning the subsequences $A[i \ldots m]$ and $B[j \ldots n]$, and by convention let $D(i,j) = \infty$ if $i > m$ or $j > n$. If $i = m$ and $j = n$ then the two inputs only have one sample each, so $D(m,n) = (A[m] - B[n])^2$. In other cases, the continuity conditions (4.2) yield the following recurrence relation:

$$D(i,j) = (A[i] - B[j])^2 + \min \begin{cases} D(i+1,j) \\ D(i,j+1) \\ D(i+1,j+1) \end{cases}. \tag{4.4}$$

Intuitively, the cost of aligning two subsequences is the sum of two terms: the cost of aligning their first elements, and the cost of the the cheapest possible way of aligning the remainders of the two sequences. An implementation of Equation (4.4) can be used to compute the optimal total cost (namely, $D(1,1)$) in $O(mn)$ time.

### 4.1.2 Modified DTW Method

The recurrence (4.4) allows a single sample in either sequence to be aligned with an arbitrarily large number of consecutive samples in the other sequence. This is undesirable in many applications because such a situation usually represents an unphysical alignment: it is rare to see a single phenomenon manifest itself over a long period of time in one signal and a nearly instantaneous period in another signal. Syed *et al.* [50] avoid this problem by using a slightly modified formulation of DTW, which we refer to as DTW*. In DTW*, the recurrence (4.4) is replaced by the following recurrence:

$$D(i,j) = (A[i] - B[j])^2 + \min \begin{cases} D(i+1, j+1) \\ D(i+1, j+2) \\ D(i+1, j+3) \\ D(i+2, j+1) \\ D(i+3, j+1) \end{cases} . \tag{4.5}$$

The modified recurrence ensures that one sample in either sequence can never be aligned with more than 3 samples in the other sequence.[1] In this way, we never evaluate the cost of highly unphysical alignments.

### 4.1.3 Earth Mover's Distance

Earth Mover's Distance (EMD) [41] is a metric for comparing two non-negative signals, interpreted as distributions—a generalization of probability distributions in which the total mass need not sum to 1. The distance between the two signals is based on the amount of *work* needed to construct one signal by deforming the other. In order to obtain a non-negative distribution from an arbitrary time-series (e.g. ECG data), we consider the absolute value of the original signal and treat it as a distribution over time.

The signal with the larger total mass (the mass of $A$ is defined as $\sum_i A[i]$) is considered to be the *source* distribution ($A$) and the smaller signal is considered to

[1] The choice of the number 3 here, which reflects the specific form of (4.5), is somewhat arbitrary. We speculate any small integer $\geq 2$ will yield similar results for most applications.

be the *goal* distribution ($B$). The objective is to start with the configuration of mass specified by the source distribution and move mass in order to construct the goal distribution using as little work as possible.

Formally, suppose $f(t_A, t_B)$ is the amount of mass to be moved from point $t_A$ to point $t_B$. The movement specified by $f$ is allowable if it does not exhaust the mass available and satisfies the demand, for all values of $t$:

$$\begin{aligned} A[t] - \textstyle\sum_{t_B} f(t, t_B) \geq 0 \quad & \forall t \\ A[t] + \textstyle\sum_{t_A} f(t_A, t) \geq B[t] \quad & \forall t. \end{aligned} \tag{4.6}$$

EMD is the minimum amount of work needed (when all functions from the set of allowable functions $F$ are considered) to construct the goal, where one unit of work is needed to move a unit mass a unit distance. The result is normalized by dividing by the total mass of the goal signal:

$$\text{EMD}(A, B) \equiv \frac{1}{\sum_t B[t]} \left( \min_{f \in F} \sum_{t_A, t_B} |t_A - t_B| f(t_A, t_B) \right). \tag{4.7}$$

More generally, EMD may be defined for distributions over a general metric space, not just distributions over time. In either case, the problem of finding the minimum work constitutes a special case of the transportation problem and can be expressed as a linear program.

### 4.1.4 Fréchet Distance

Fréchet Distance (FD) [5, 18] is often described by analogy to a man walking a dog: a man and a dog are connected by a leash and each walks along a different path. The man and the dog may move at any speed they wish, but neither may move backwards. FD is the minimum leash length that allows the man and the dog to traverse their paths.

Formally, given two directed paths parameterized by functions $A$ and $B$ mapping the interval $[0, 1]$ to the set $S$, and a metric $d(s_1, s_2)$ on $S$, the Fréchet distance is

defined as

$$\min_{\phi_A,\phi_B} \max_{0\le t\le 1} d(A(\phi_A(t)), B(\phi_B(t))).$$

where $\phi_A$ and $\phi_B$ (intuitively, the positions of the man and dog as functions of time) may be any monotonically non-decreasing continuous functions from $[0, 1]$ to $[0, 1]$ satisfying the boundary conditions

$$\phi_A(0) = 0, \qquad \phi_A(1) = 1, \qquad \phi_B(0) = 0, \qquad \phi_B(1) = 1.$$

Here we consider the analogous notion for discrete time series, the *discrete Fréchet distance* [18]. Suppose we have two sequences of length $m$ and $n$ containing time-value pairs $(t, v)$ as well as a metric on time-value pairs, $d((t_1, v_1), (t_2, v_2))$. We consider alignments between the two sequences as we did for DTW in Section 4.1.1. For an alignment characterized by sequences $\phi_A$ and $\phi_B$, we may consider the maximum distance between aligned pairs of samples. The discrete Fréchet distance is the minimum value of this distance over all possible alignments:

$$\mathrm{FD}(A, B) \equiv \min_{\phi_A,\phi_B} \max_i d((\phi_A[i], A[\phi_A[i]]), (\phi_B[i], B[\phi_B[i]])) \tag{4.8}$$

This problem is structurally similar to that of DTW and, like DTW, can be solved with dynamic programming in $O(mn)$ time.

### 4.1.5 Hausdorff Distance

Hausdorff Distance (HD) [23] measures the difference between two closed and bounded sets of points. Given a metric $d(s_1, s_2)$, the distance from a single point $a$ to a set $B$ can be defined as the distance from $a$ to the nearest point in $B$, i.e. $\min_{b\in B} d(a, b)$. To generalize this notion to two sets $A$ and $B$, we consider the maximum of this minimum distance from any point in either set to the other set:

$$\mathrm{HD}(A, B) \equiv \max\left\{ \max_{a\in A} \min_{b\in B} d(a, b), \max_{b\in B} \min_{a\in A} d(a, b) \right\}. \tag{4.9}$$

To apply HD to time-series data, we consider the inputs as sets of points in time-amplitude space $(t, v)$ and define a metric $d((t_1, v_1), (t_2, v_2))$. A straightforward implementation of Equation (4.9) evaluates HD in $O(mn)$ time, where $m$ and $n$ are the lengths of the two input sequences.

## 4.2 Evaluation and Results

### 4.2.1 Selecting scaling parameters for FD and HD

FD and HD are both parameterized by a distance metric $d((t_1, v_1), (t_2, v_2))$. Such a metric must make some trade-off in how it weights changes in time against changes in amplitude, and the particular choice of metric may affect performance on applications to which the distance measure is applied. To assess FD and HD, we evaluate their performance with a range of metrics. In particular, we consider metrics of the form

$$d((t_1, v_1), (t_2, v_2)) \equiv \sqrt{(v_2 - v_1)^2 + (\alpha(t_2 - t_1))^2} \tag{4.10}$$

where $\alpha$ (the "scaling parameter") is a positive number that sets the relative weight of unit changes in time and amplitude.

### 4.2.2 Cylinder-bell-funnel problem

#### Discrimination ability

To evaluate the utility of the various morphologic distance measures for distinguishing between shapes, we compared the distances yielded by each method on a set containing 500 randomly generated instances of each of the three eponymous classes (shapes) of the cylinder-bell-funnel (CBF) problem [42].

For each method, we computed the pairwise distances between all 1500 elements of the set. We then measured the ratio of the average distance in pairs containing different shapes to the average distance in pairs of the same shape. We refer to this ratio as the inter/intra distance ratio (IIDR). High values of the IIDR indicate

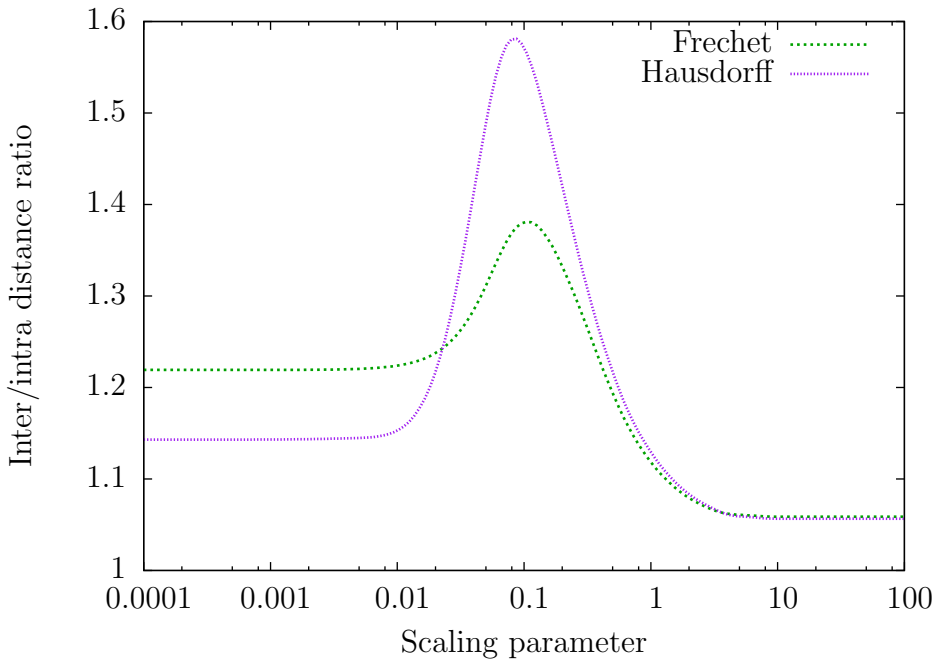



Figure 4-1: Discrimination performance on CBF data as a function of $\alpha$ for selected MD measures. FD achieved a maximum IIDR of 1.38 at $\alpha = 0.105$, and HD achieved a maximum of 1.58 at $\alpha = 0.087$.

| Method | | IIDR |
|---|---|---|
| DTW | | 2.06 |
| DTW* | | 2.14 |
| EMD | | 1.82 |
| Fréchet | $\alpha = 0.105$ | 1.38 |
| Hausdorff | $\alpha = 0.087$ | 1.58 |

Table 4.1: Discrimination performance of MD measures on CBF data.

that the method distinguishes between instances of different classes while ignoring differences between instances of the same class.

For FD and HD, we evaluated the IIDR under a range of scaling parameters as described in Section 4.2.1. The results are shown in Figure 4-1. For subsequent experiments on this dataset, we use the scaling parameters that yielded the highest IIDR here and report the results obtained.

The IIDR observed for each of the five methods is reported in Table 4.1. The DTW methods, followed by EMD, showed the best performance among the methods considered.

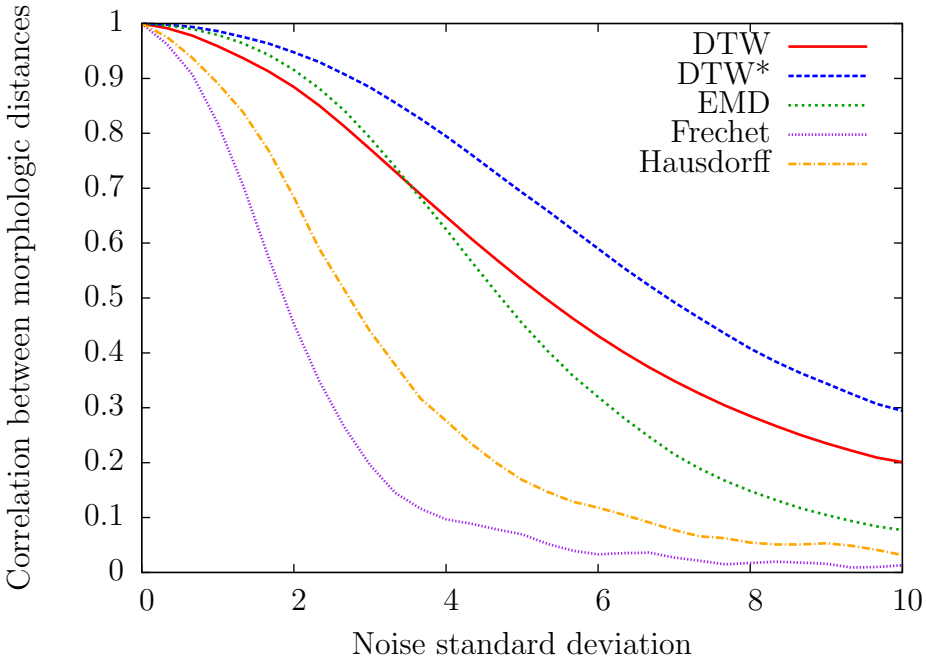



Figure 4-2: Robustness of MD measures on CBF data.

**Noise resistance**

We evaluated the noise resistance of each morphologic distance measure by comparing the morphologic distances obtained from given inputs to the morphologic distances obtained when those inputs were corrupted by noise.

We synthesized 100 pairs of random cylinder-bell-funnel instances (i.e. 200 instances) and computed the morphologic distance between the elements of each pair. We then corrupted each instance with Gaussian distributed white noise of a fixed variance. The computation of morphologic distances was repeated on the noisy data. The correlation coefficient of the two output sequences (before and after noise was added) was used as a measure of how robust the method was against noise. We evaluated this correlation for various values of noise variance.

The morphologic distance correlations observed for various noise levels are shown in Figure 4-2. DTW* showed the highest noise resistance at all noise levels, followed by DTW and EMD.

### 4.2.3 Electrocardiogram data

#### Discrimination ability

We performed experiments analogous to those in Section 4.2.2 on ECG data. We acquired ECG data for 23 patients from a subset of the MIT-BIH Arrhythmia Database [20] that is intended to be a representative sample of ECG waveforms. The data were selected randomly from a corpus of ambulatory ECG recordings obtained from patients at Beth Israel Hospital with normal heart activity. The recordings are relatively clean, and higher quality than typical Holter data (360 samples per second with 11-bit resolution). Two channels of ECG data were available; we considered only the first channel—a modified limb lead II (MLII), which shows the QRS complex clearly and in the usual orientation.

The data were annotated with cardiologist-supplied labels for each heartbeat (e.g. normal, premature ventricular contraction, etc.). We excluded 9 patients who did not have at least two differently labeled beats each of which occurred at least five times. For each of the remaining 14 patients, we randomly selected pairs of heartbeats with the same label and pairs of heartbeats with different labels. We then computed the IIDR—here, the ratio of the average distance between heartbeats of different labels to the average distance between heartbeats of the same label. The average IIDR across all patients was considered to be a measure of discrimination performance in this application.

For FD and HD, appropriate scaling parameters were selected as in Section 4.2.2; the average IIDR as a function of scaling parameter is shown in Figure 4-3. The average IIDR for each of the five methods is reported in Table 4.2. In this application, the DTW methods and EMD again showed the best discrimination performance. Surprisingly, DTW performed much better here than DTW*. We speculate that this is due to the relatively high variability of ECG morphologies. Even when instances have the same physiological label, they may have different morphologies and require significant warping to align. When significant warping is needed, DTW yields lower morphological distances than DTW* (where the amount of warping is restricted). We

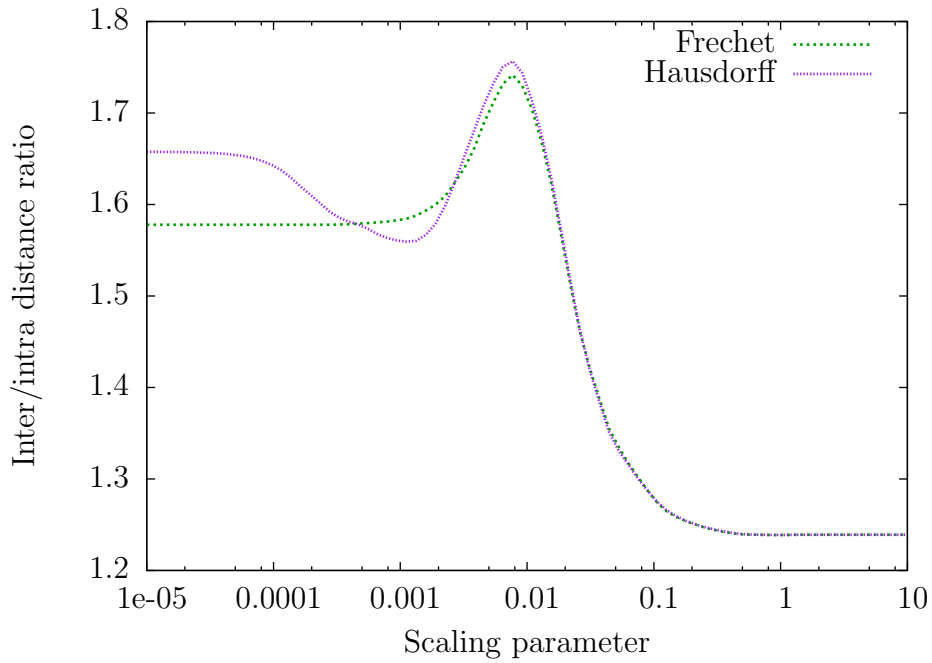



Figure 4-3: Discrimination performance on ECG data as a function of $\alpha$ for selected MD measures. FD achieved an IIDR of 1.74 at $\alpha = 0.0077$, and HD achieved an IIDR of 1.76, also at $\alpha = 0.0077$.

| Method | | IIDR |
|---|---|---|
| DTW | | 2.66 |
| DTW* | | 2.19 |
| EMD | | 2.16 |
| Fréchet | $\alpha = 0.0077$ | 1.74 |
| Hausdorff | $\alpha = 0.0077$ | 1.76 |

Table 4.2: Discrimination performance of MD measures on ECG data.

believe that this is true to a much smaller extent in the CBF instances.

### Noise resistance

We evaluated the noise resistance of each method as in Section 4.2.2, using the pairs of ECG signals from the previous experiment. Figure 4-4 shows the correlation between the morphologic distances on clean data and the morphologic distances when the data were corrupted by noise, as a function of the noise variance. DTW* showed the highest noise resistance at all noise levels, followed by DTW and HD.

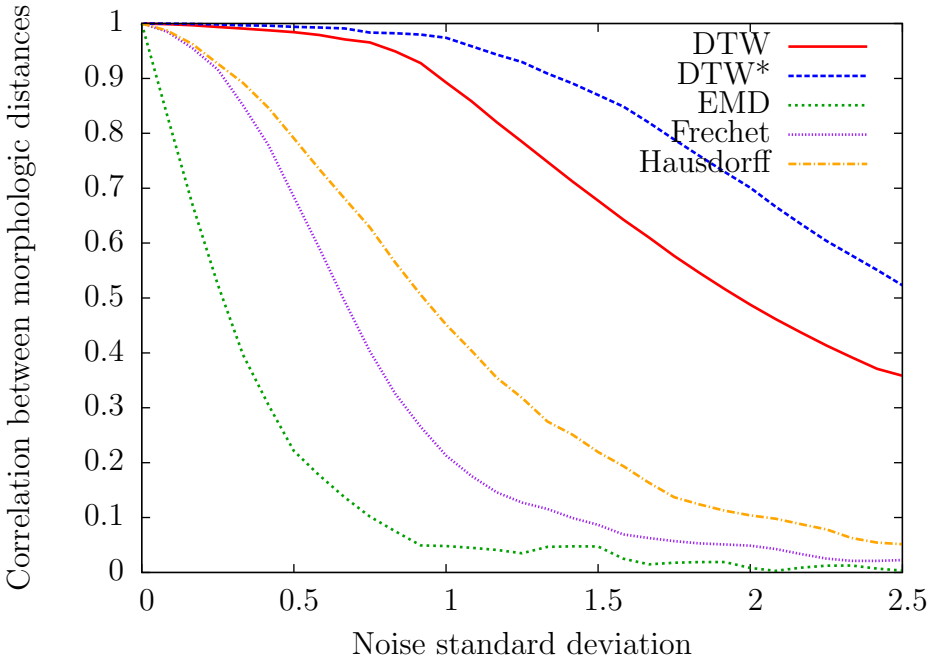



Figure 4-4: Robustness of MD measures on ECG data.

### 4.2.4 Risk stratification

Finally, we evaluated the utility of each method for risk stratification of patients following non-ST-elevation ACS (NSTEACS). We computed morphologic variability for 764 patients using data from the DISPERSE-2 (TIMI33) trial [13] for whom 24 hours of continuous ECG were recorded within 48 hours of a NSTEACS. More information about the trial data and our criteria for including patient data are given later in Section 5.1. MV-LF/HF was computed using DTW* as the morphologic distance measure, as described in Section 3.5 and in [50]. We then repeated the computation of MV while replacing DTW* with each of the other four morphologic distance metrics (DTW, EMD, FD, HD) in turn.

We evaluated each measure by computing the c-statistic of the resulting MV values for prediction of death within a 90-day followup period of the initial event. The resulting c-statistic for each measure is shown in Table 4.3. None of the morphologic distance measures under consideration yielded MV values that had prognostic value better than DTW*, the method currently proposed for use.

| Method | c-statistic |
|---|---|
| DTW | 0.619 |
| DTW* | 0.724 |
| EMD | 0.517 |
| Fréchet | 0.666 |
| Hausdorff | 0.656 |

Table 4.3: Prediction accuracy of 90-day death using MV with various MD measures.

## 4.3 Discussion

In the four general evaluations (discrimination performance and noise resistance on both datasets), the DTW methods consistently outperformed the other three metrics. EMD performed better than the remaining metrics on all but one of the evaluations—the exception being the noise resistance test on ECG data (illustrated in Figure 4-4).

We believe that the superior performance of the DTW methods and EMD is in part due to the fact that they sum a particular cost function across the entire signal, yielding a cumulative assessment of morphology differences. In contrast, FD and HD look at the input samples to see where some feature is maximized, and then use that as their computed distance. Consequently, those metrics are more sensitive to differences (including random noise) in limited regions of the two signals, differences that are not necessarily representative of morphologic differences across the entire signal. In many settings, including ECG recordings, measurements are accompanied by significant amounts of noise, so it is critical that signal processing techniques are robust against this noise.

In the risk stratification evaluation, DTW* showed the best performance among the MD metrics, by a significant margin, for identifying morphologic differences associated with risk of death. We conclude that DTW* is, reassuringly, well suited for use in MV-LF/HF.

Because the five methods showed somewhat inconsistent behavior across the different evaluations, we caution against interpreting the results of any one evaluation too generally. For example, although the EMD-based MV measure showed the worst performance among all the MD metrics, EMD performed relatively well in the noise re-

sistance and discrimination performance evaluations (even those on ECG data). This inconsistency may arise because the methods we considered may vary in their sensitivity to large-scale morphology changes (like those between heartbeats of different physiological classes) and small-scale morphology changes (like the subtle distinctions that we believe are quantified by MV techniques).

One additional consideration is that the use of FD and HD (as described here) is complicated somewhat by the need to select an appropriate metric. This choice of metric can have a significant effect on discrimination performance. We cannot exclude the possibility that the performance of FD or HD might improve substantially on the tests we conducted when they are computed using a different metric.

# Chapter 5

# Morphologic Variability

In this chapter, we develop a new technique for summarizing morphologic distance information to provide a measure of *morphologic variability* and an evaluation of post-ACS risk. Syed *et al.* showed that low levels of the MV-LF/HF measure are significantly associated with higher risk of death and MI following NSTEACS [50, 51]. As discussed in Section 2.5, there is evidence that MV measures beat-to-beat variations in the cardiac conduction path caused by intrinsic myocardial instability.

We propose changes to the MV-LF/HF measure that improve its accuracy of risk assessment. Our primary contribution in this part is the empirical identification of a *diagnostic frequency* band such that the total energy in that band correlates strongly with risk of death. We present a new risk measure called MV-DF that considers the energy in the diagnostic frequency band.

The remainder of this chapter is organized as follows. Section 5.1 gives information about the populations on which we evaluated our techniques. Section 5.2 describes the selection of an *output threshold*, which is a parameter of the summarization process. Section 5.3 describes the selection of the diagnostic frequency band and presents performance results for the MV-DF measure. Section 5.4 discusses the application of MV-DF to the prediction of myocardial infarction rather than death. Section 5.5 proposes a possible mechanism by which myocardial instability could yield high energy in the DF band of the MD time series. Section 5.6 concludes with a discussion.

## 5.1 Evaluation

We trained our methods on data from 764 patients from the DISPERSE-2 (TIMI33) trial [13]. The DISPERSE-2 trial compared the efficacy and safety of AZD6140 and clopidogrel in patients admitted following NSTEACS. Patients in the trial had continuous Holter monitor ECG data recorded at a sample rate of 128 Hz within 48 hours of admission. A total of 990 patients were enrolled in the trial. Patients with fewer than 24 hours of continuous ECG were excluded from the study. To exclude data with low signal quality, further preprocessing was conducted as described in [48]. Usable data from 764 patients remained after this step. Patients were followed up for a period of 90 days for the endpoints of death and MI; there were 15 deaths (2.0%) in this cohort during the follow-up period.

Subsequently, we validated our methods on data from an independent set of 2,302 patients in the placebo group of the MERLIN (TIMI36) trial [33], which compared the safety and efficacy of ranolazine versus placebo in patients post-NSTEACS. All patients in this study had at least 24 hours of continuous ECG data, and preprocessing was performed in a similar way as for DISPERSE-2. Patients were followed up for up to 24 months for the endpoints of death and MI. The median follow-up length was 348 days. There were 57 deaths (2.5%) in this cohort within the first 90 days of the follow-up period.

## 5.2 Output thresholds

The MV-LF/HF technique described in Section 3.5 computes a measure of morphologic variability for each of a patient's 5-minute windows and then uses the median value across windows as a summary—an indicator of the patient's overall morphologic variability. (HRV methods usually operate in the same way, except that it is more common to take the mean value across windows [30].)

We consider a generalization of this procedure by introducing a parameter called the *output threshold*, labeled $\theta$. Given a set of $n$ inputs—an MV-LF/HF value for

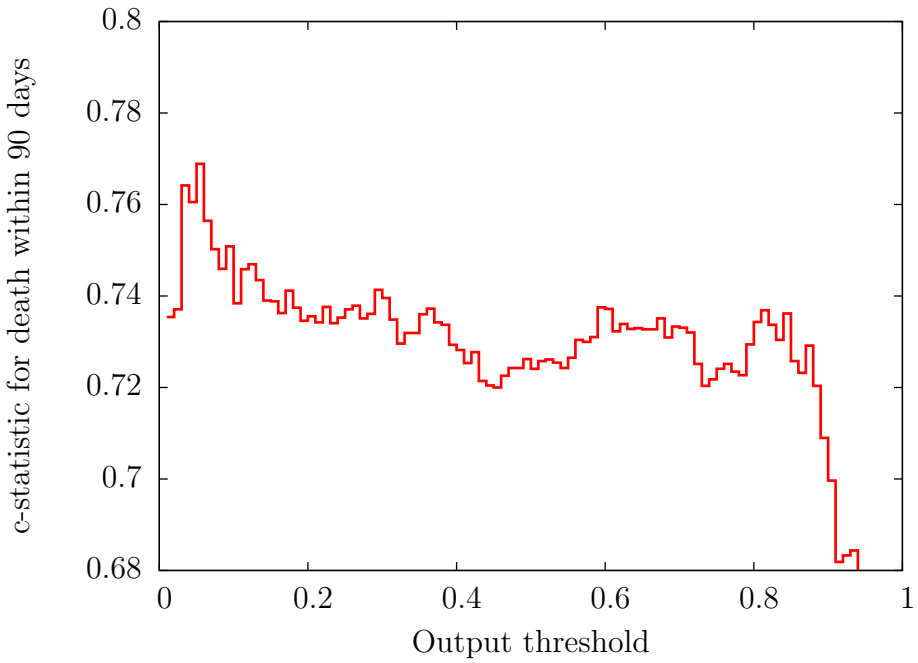



Figure 5-1: Prediction of 90-day death using MV-LF/HF with varying output thresholds ($\theta$).

each of $n$ windows—we sort them and return a value that is greater than $\lfloor \theta n \rfloor$ of them. The previous procedure, in which we returned the median, now corresponds to an output threshold of 0.5, because the median of a set is greater than half of its elements.

To determine whether a different output threshold might improve the risk stratification of MV-LF/HF, we evaluated the c-statistic for predicting death within 90 days on the DISPERSE-2 dataset as the output threshold $\theta$ was allowed to vary between 0 and 1. Figure 5-1 shows this discrimination quality as a function of output threshold. There is a trend towards improved discrimination as the output threshold is lowered. The c-statistic improved from our baseline of 0.724 (at $\theta = 0.50$) and achieved a maximum value of 0.769 (at $\theta = 0.05$).

Intuitively, although having low MV in general is dangerous, having even a small number of extremely low-MV periods is even more dangerous. Our results suggest that having this characterization of a patient's extreme behavior is useful for risk analysis. When using an output threshold, MV-LF/HF can incorporate information

about morphologic variability that is not captured in a measure of central tendency like the median.

The quality of discrimination drops off sharply as $\theta$ becomes very low or very high (e.g. $\theta < 0.03$ or $\theta > 0.90$). This may be because as $\theta$ nears 0 or 1, the method returns more extreme values from among the inputs. For example, $\theta = 0$ ($\theta = 1$) corresponds to returning the minimum (maximum) observed value over all windows—and windows with near-minimal (near-maximal) MV-LF/HF are more likely to contain noise, measurement anomalies, etc., and less likely to contain physiologically interesting information about the patient. (Happily, this noise effect seems to be insignificant on a surprisingly large portion of the $\theta$ space.)

When selecting an output threshold $\theta$ to produce a good summary, we need to take advantage of the trend towards improved discrimination while being careful to avoid the effects of noise. Guided by the results shown in Figure 5-1, we chose an output threshold for use in subsequent experiments. We selected a value of $\theta = 0.10$, for which the c-statistic on the DISPERSE-2 dataset is 0.738 (an improvement of 0.014 over the baseline). This $\theta$ is suboptimal for this dataset—in the sense that a smaller $\theta$ maximizes the c-statistic—but it lies far from the zone in which the effects of noise become significant.

We verified that similar trends were also observed on the MERLIN dataset. The baseline c-statistic was 0.675 ($\theta = 0.50$). At $\theta = 0.10$, it improved to 0.706, which is approximately the maximal c-statistic on the MERLIN dataset.

## 5.3 A diagnostic frequency (DF) band

The MV-LF/HF measure proposed in [50] was intended to be analogous to HRV-LF/HF [30] so that the utility of using the MD time series in place of the NN time series could be fairly evaluated. The frequency bands for the low frequency (0.04 Hz–0.15 Hz) and high frequency components (0.15 Hz–0.40 Hz) are those used for HRV. As described in Section 3.3, the LF and HF frequency bands are associated with the timescales on which the sympathetic nervous system and vagal activity regulate

| Measure | $\theta = 0.1$ | $\theta = 0.5$ |
|---|---|---|
| LF | 0.553 | 0.539 |
| 1/HF | 0.733 | 0.674 |
| LF/HF | 0.738 | 0.724 |

Table 5.1: Prediction accuracy (c-statistic) of 90-day death using LF, 1/HF, and LF/HF. Since the c-statistic is a non-parametric statistic, 1/HF and HF are equivalent here.

heart rate. Since morphologic variability is intended as a measure of a different phenomenon, electrical instability of the myocardium, there is little reason to believe that the frequency bands used for HRV would also be optimal for risk assessment with MV. We show here that there is, in fact, a range of better frequencies for the purposes of MV-based risk stratification.

The MV-LF/HF measure is computed as the ratio of energies in two separate frequency bands of the MD time series. Low LF/HF is an indicator of risk and is associated with either low LF energy, high HF energy, or both. We investigated whether energy in either of those frequency bands alone, or in *any* single frequency band, could be used effectively for prognostication.

We first examined the effectiveness of using the LF and HF energies alone (instead of the LF/HF ratio) for risk assessment on the DISPERSE-2 dataset. The results of this analysis, for output thresholds of $\theta = 0.5$ and $\theta = 0.1$, are shown in Table 5.1. For $\theta = 0.5$, neither the LF nor HF bands by themselves give as much information as LF/HF, but having high HF is much more indicative of adverse events than is having a low LF. For $\theta = 0.1$, HF and LF/HF are almost equivalent. This suggests that the HF energy is the major contributor of information in the LF/HF ratio. Indeed, for both $\theta$ values, considering the LF energy alone gives prognostication that is only slightly better than chance.

Based on these results, we decided to attempt to identify a frequency band near the HF range such that high energy in that band would be a strong predictor of death. We refer to this frequency band as the *diagnostic frequency* (DF) band.

To help choose the DF band, we evaluated all possible frequency bands within

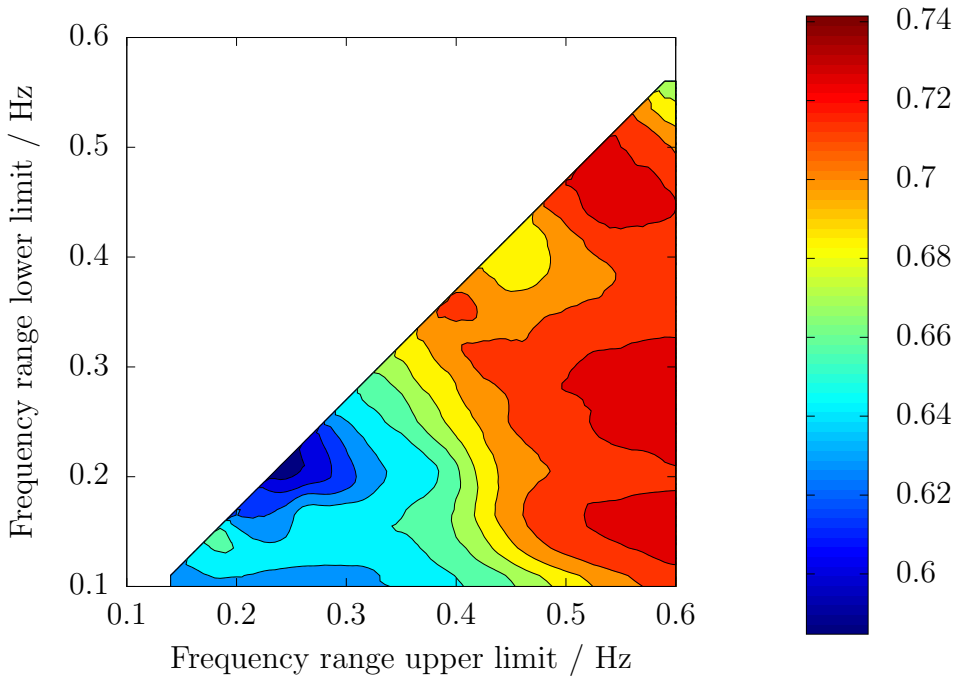



Figure 5-2: Prediction accuracy (c-statistic) of 90-day death using single band energy as a function of band location, $\theta = 0.5$.

the range 0.1 Hz–0.6 Hz to identify an optimal band. We excluded frequency bands with width less than 0.05 Hz in order to ensure that the energy within the band could be estimated robustly from time-series data. The frequency band endpoints were tested with a granularity of 0.01 Hz. For each band, we evaluated the c-statistic for predicting death within 90 days on the DISPERSE-2 dataset, using $\theta = 0.5$. Figure 5-2 shows the resulting c-statistic as a function of the band's location (i.e. we vary the lower and upper frequency limits). The maximum c-statistic observed within the search space was 0.741.

We repeated this analysis for an asymmetric output threshold. Because high risk is associated with low values of LF/HF but high values of DF, we reversed the direction in which we moved the output threshold in order to give an analogous metric for DF. Therefore, we use $\theta = 0.9$ rather than $\theta = 0.1$. Figure 5-3 shows the c-statistic as a function of band location when $\theta = 0.9$. The maximum c-statistic observed in the search space was 0.778.

Based on the results shown in Figures 5-2 and 5-3, we selected the DF band to be

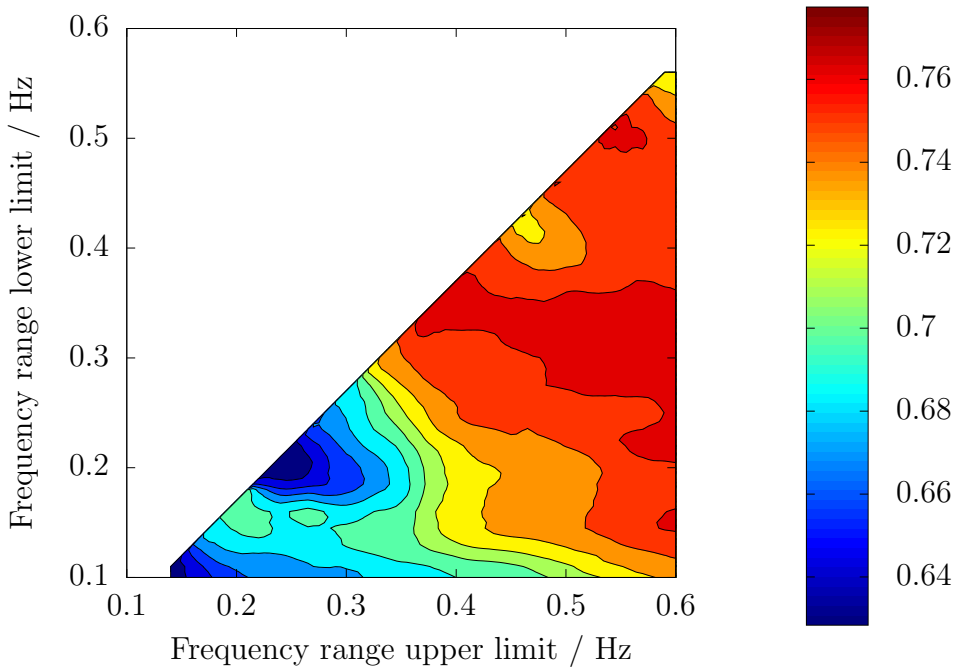



Figure 5-3: Prediction accuracy (c-statistic) of 90-day death using single band energy as a function of band location, $\theta = 0.9$.

0.30 Hz–0.55 Hz. This frequency range shows good discrimination for both $\theta = 0.5$ and $\theta = 0.9$; the respective c-statistics are 0.732 and 0.771. These are near optimal—for both $\theta$ values, the c-statistic of the DF band is within 0.01 of the best c-statistic observed within the search space for that $\theta$ (0.741 and 0.778, respectively). We refer to the risk measure obtained by using the DF band with $\theta = 0.9$ as the MV-DF measure.

We revisited the selection of the output threshold to see if the change of frequency bands would have had any effect on our choice of $\theta$. Figure 5-4 shows the quality of discrimination when the new DF band is used but the output threshold is once again allowed to vary. The results of this analysis support the use of our initial choice of $\theta = 0.9$ with the new DF band. The same general behavior is observed with MV-LF/HF (except, of course, that the effects of lowering and raising $\theta$ are reversed). The c-statistic rises steadily with $\theta$ and the trend is much more pronounced than it was under MV-LF/HF. As in Figure 5-1, there is a steep drop in quality of discrimination for values of $\theta$ very close to 0 and 1. The prominent peak between $\theta = 0.9$ and

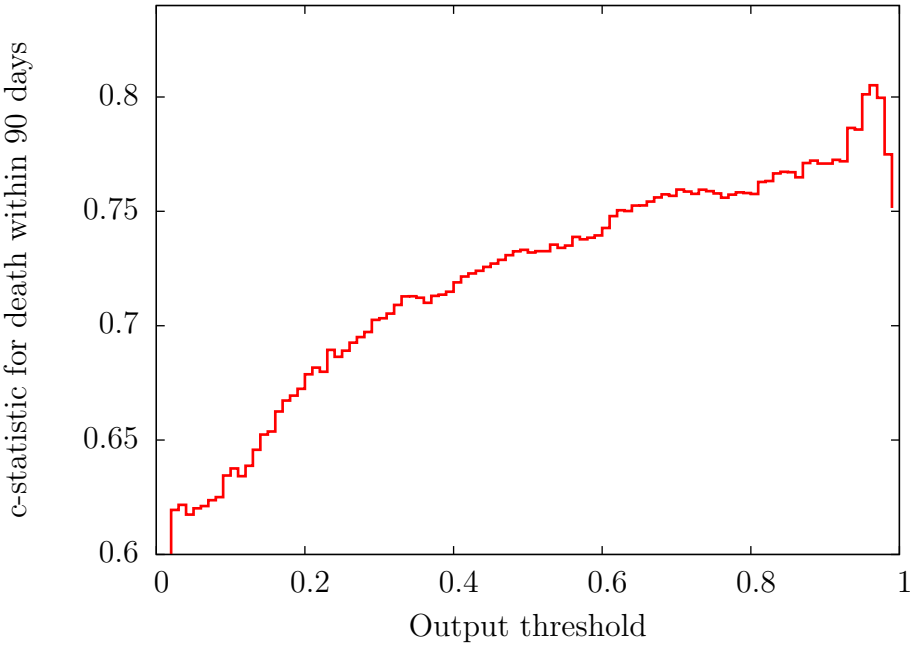



Figure 5-4: Prediction of 90-day death using DF band energy (0.30 Hz–0.55 Hz) with varying output thresholds ($\theta$).

$\theta = 1.0$ here is somewhat unexpected; however, it may not be indicative of any general phenomenon, given that only 15 patients died in the DISPERSE-2 population.

We validated the selection of the MV-DF measure (which was selected using the DISPERSE-2 dataset) on the MERLIN dataset. Figure 5-5 shows the c-statistic as a function of the band location for $\theta = 0.9$ (that is, this figure shows the same information for MERLIN that Figure 5-3 shows for DISPERSE-2). On this dataset, MV-LF/HF achieved a c-statistic of 0.675. The MV-DF measure achieved a c-statistic of 0.752. This was close to the maximum c-statistic observed in the search space in Figure 5-5, which was 0.766. These results provide corroboration on an independent dataset of the appropriateness of the MV-DF measure.

Table 5.2 and Figure 5-6 summarize the performance improvements from the two changes we made to the MV-LF/HF method—using an output threshold of 0.9 and using the new DF band—separately and then together. The baseline is the MV-LF/HF method as described in [50]. Interestingly, on both datasets, the improvement gained by using both changes together is greater than the sum of the improvements

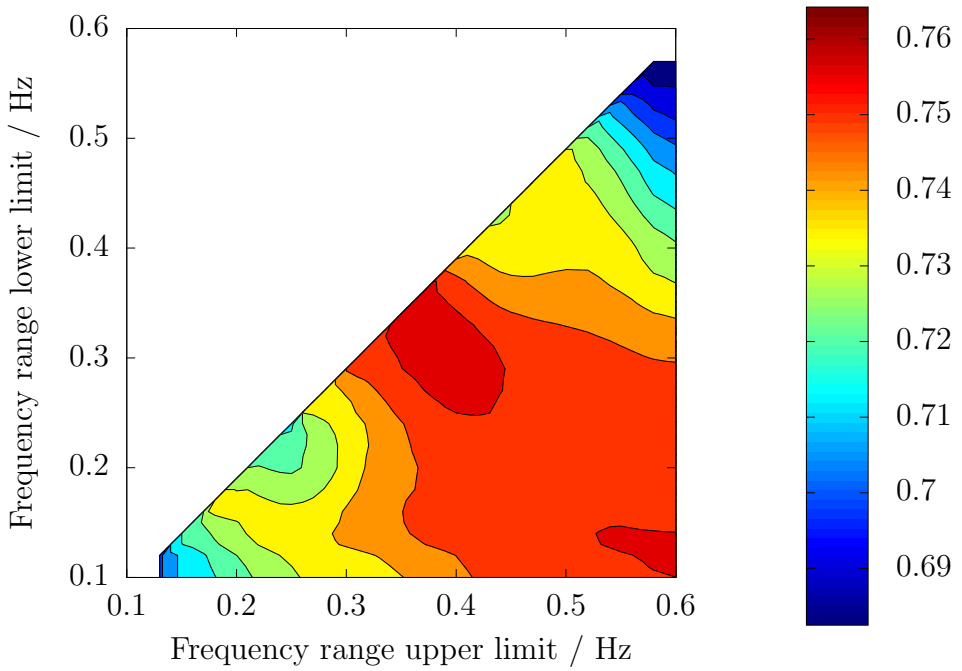



Figure 5-5: Prediction accuracy (c-statistic) of 90-day death using single band energy as a function of band location, $\theta = 0.9$, on the MERLIN dataset.

| Measure | DISPERSE-2 | MERLIN |
|---|---|---|
| Baseline (MV-LF/HF) | 0.724 | 0.675 |
| +Threshold | 0.738 | 0.706 |
| +DF Band | 0.732 | 0.718 |
| +Both (MV-DF) | 0.771 | 0.752 |

Table 5.2: Summary of MV method changes and their effects on method performance. The c-statistic for each prospective method is shown on both datasets.

gained by using the two changes separately.

We used a Cox proportional hazards model [16] to gauge the value of using dichotomized MV-DF values to separate the population into low- and high-risk groups. The Cox proportional hazards model estimates the *hazard ratio*, the ratio of the instantaneous rates of death for patients in the two groups. The quartile of patients with the highest MV-DF was identified as the high-risk group. On the DISPERSE-2 dataset, the cutoff MV-DF value for the top quartile was 48.0. The high-risk group (MV-DF $\geq$ 48.0) showed a hazard ratio of 8.21 ($p < 0.001$, CI $= [2.72, 27.48]$) for death within 90 days of NSTEACS, and a hazard ratio of 12.30 ($p = 0.002$) for death

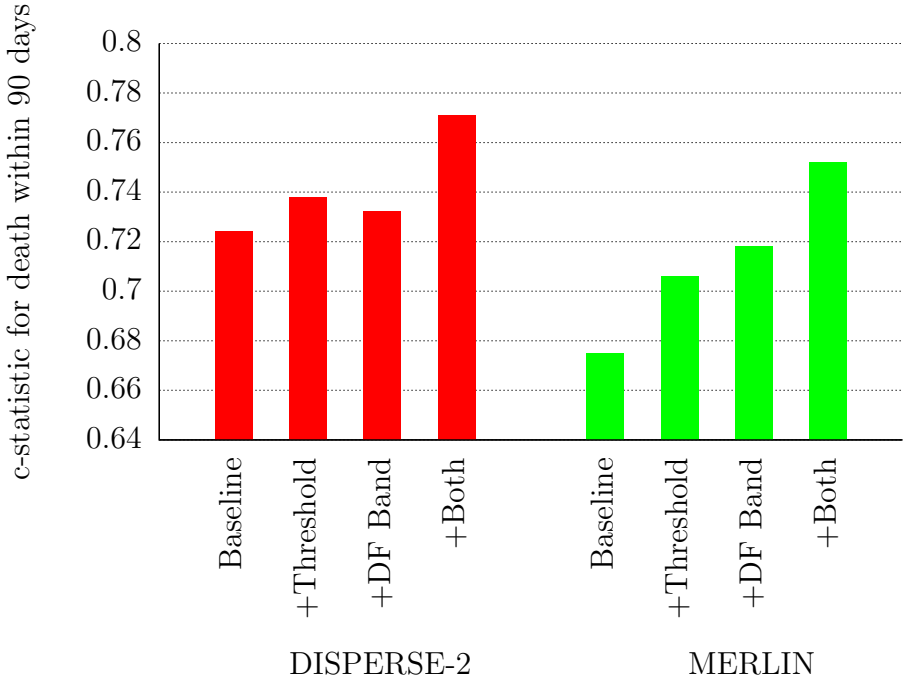



Figure 5-6: Summary of MV method changes and their effects on method performance.

within 30 days of NSTEACS.

When applied to the MERLIN dataset, this cutoff identified identified a high-risk group containing 647 patients (28.1% of the cohort). The high-risk group showed a hazard ratio of 5.12 ($p < 0.001$, CI $= [3.07, 9.39]$) for death within 90 days. This hazard ratio is lower than the hazard ratio on the DISPERSE-2 data. However, this could reasonably be expected because the DF frequency band was specifically chosen using the data from the DISPERSE-2 trial. In addition, the hazard ratio under MERLIN is still well within the 95% confidence interval obtained on the DISPERSE-2 data.

Subsequent to our work, Syed *et al.* compared the performance of the MV-DF metric to DC and HRV-LF/HF on data from the MERLIN trial [49]. In that study, there was a median follow-up period of 348 days post-NSTEACS. The following predictors were considered in a multivariate analysis: dichotomized values of the three ECG risk measures, LVEF $\leq 0.4$, TIMI risk score group, high levels of brain natriuretic peptide (BNP, an indicator of congestive heart failure), high BMI, and low

| Group | Adjusted HR | CI | P value |
|---|---|---|---|
| DC $\in [2.5, 4.5]$ | 1.55 | [0.93, 2.57] | 0.094 |
| DC $< 2.5$ | 1.97 | [1.04, 3.73] | 0.039 |
| HRV-LF/HF $< 0.95$ | 1.36 | [0.83, 2.23] | 0.221 |
| MV-DF $> 52.5$ | 1.65 | [1.00, 2.71] | 0.050 |

Table 5.3: Multivariate analysis of ECG risk measures ($n = 2122$; from [49]).

| Group | Adjusted HR | CI | P value |
|---|---|---|---|
| DC $\in [2.5, 4.5]$ | 1.76 | [1.00, 3.12] | 0.052 |
| DC $< 2.5$ | 1.88 | [0.86, 4.08] | 0.111 |
| HRV-LF/HF $< 0.95$ | 1.17 | [0.66, 2.09] | 0.594 |
| MV-DF $> 52.5$ | 2.31 | [1.28, 4.16] | 0.005 |

Table 5.4: Multivariate analysis of ECG risk measures for patients with LVEF $> 0.4$ ($n = 1943$; from [49]).

creatinine clearance (CrCl). Abbreviated results from this analysis are shown in Table 5.3. MV and DC, but not HRV, were independently and significantly associated with death, and showed comparable hazard ratios. However, MV's prognostic value increased (relative to the other two ECG risk measures) when only outcomes in the first 90 days following NSTEACS were considered. MV-DF therefore seems to be especially valuable for short-term risk stratification.

A separate analysis was conducted in which only patients with high LVEF were considered. The results of this analysis are shown in Table 5.4. MV-DF was significantly associated with death in this population, but DC and HRV were not. Because low LVEF is already a quite good predictor of risk, these results suggest that the MV-DF metric may be used to supplement the echocardiogram by identifying high-risk patients who would not have been identified as such on the basis of LVEF.

### 5.3.1 Graded response

The risk of death as MV-DF increases shows a graded response. The 90-day incidence of death in the MERLIN dataset, when the patients are divided into deciles by MV-DF value, is shown in Table 5.5. The 60% of patients with the lowest MV-DF (MV-DF $<$

| Decile | Number of deaths | Death rate |
|---|---|---|
| 0 (MV $<$ 26.4) | 1 | 0.4% |
| 1 (26.4 $\leq$ MV $<$ 28.9) | 4 | 1.7% |
| 2 (28.9 $\leq$ MV $<$ 31.5) | 1 | 0.4% |
| 3 (31.5 $\leq$ MV $<$ 34.3) | 2 | 0.9% |
| 4 (34.3 $\leq$ MV $<$ 37.6) | 1 | 0.4% |
| 5 (37.6 $\leq$ MV $<$ 41.3) | 3 | 1.3% |
| 6 (41.3 $\leq$ MV $<$ 46.9) | 5 | 2.2% |
| 7 (46.9 $\leq$ MV $<$ 54.2) | 7 | 3.0% |
| 8 (54.2 $\leq$ MV $<$ 79.9) | 13 | 5.7% |
| 9 (79.9 $\leq$ MV) | 20 | 8.7% |

Table 5.5: 90-day rate of death in each decile of MV-DF ($n = 2302$).

41.3) show a low (0.9%) risk of death in the 90 days following NSTEACS. This rate rises with increasing MV-DF in the remaining 4 deciles, with a maximum rate of 8.7% on the highest-MV decile (MV-DF $>$ 79.9).

### 5.3.2 Stability

To evaluate the sensitivity of the MV-DF measure to variations in the band parameters, we evaluated the stability of the high- and low-risk groups that were identified when the lower and upper band limits were perturbed. We compared the MV-DF metric to the metric obtained when either the DF band lower or upper frequency limit was raised or lowered by 0.05 Hz (a relative change in band size of 20%). For each of these 4 alternative techniques, we evaluated the number of patients who died who would have received different prognoses (high- or low-risk) under MV-DF and the alternative technique.

On the MERLIN dataset, for MV-DF and each of the other 4 metrics under consideration, we dichotomized the values at the top quartile, with high values indicating high risk. In each case this procedure identified a high-risk group containing 576 patients. Under MV-DF, 36 of the patients in this high-risk group died within 90 days.

Table 5.6 shows the results of this analysis. For each of the alternative bands, no

| New frequency band (Hz) | In | Out |
|---|---|---|
| 0.30–0.50 | 2 | 0 |
| 0.30–0.60 | 1 | 0 |
| 0.25–0.55 | 2 | 0 |
| 0.35–0.55 | 1 | 1 |

Table 5.6: Stability of prognosis when the DF frequency band is replaced with a nearby one. The "In" column shows the number of patients who died ($n = 57$) who were identified as high-risk under the alternative band but not under MV-DF. The "Out" column shows the number of patients who died who were identified as high-risk under MV-DF but not under the alternative band.

more than 2 (3.5%) of the patients who died would have received different prognoses under MV-DF and the alternative technique. Although the heatmaps in Figures 5-3 and 5-5 showed that the c-statistics were fairly stable as a function of the band location, they did not exclude the possibility that different metrics might single out different—but similarly sized—subsets of the patients who later died. The analysis here confirms that the risk assessment generated by MV-DF is stable (does not change substantially) at the level of individual patients when the band location is perturbed. Similar results were obtained when we counted all patients instead of only those who died.

### 5.3.3 Two-band measures

After refining the MV-DF metric, we returned to the question of whether a more complex risk measure that incorporated information from two frequency bands (as MV-LF/HF does) could provide significantly improved risk assessments. Naturally, a metric with more free parameters (four frequency band boundaries rather than two) can be expected to perform *no worse* on its training set than a simpler metric. The test of such a method is whether its performance is also improved on previously unseen test data.

Using techniques similar to those used in our previous investigations, we evaluated the prognostic value of an MV-LF/HF-like ratio, where the HF band was set to be the DF band we identified (0.30–0.55 Hz), and the LF band bounds were allowed

to assume values in the range 0.00–0.40 Hz. On the DISPERSE-2 dataset the best such LF/HF-like measure (with an LF band of 0.03–0.10 Hz) achieved a c-statistic of 0.789—an improvement over MV-DF. However, the same measure, when evaluated on the MERLIN dataset, produced worse risk stratification than MV-DF (a decrease in c-statistic from 0.752 to 0.725).

We fixed the HF band parameters here rather than searching the entire parameter space, and it is possible that allowing those parameters to vary could yield a better MV-LF/HF measure. However, our analysis here suggests that any improvement from such a change may be quite limited.

## 5.4 Predicting MI

Although low MV-LF/HF is significantly associated with risk of MI [50], we found no significant association between high MV-DF and MI. We evaluated the utility of using MV-DF for predicting MI within 90 days of NSTEACS on the DISPERSE-2 dataset. When dichotomized at the top quartile, high-DF patients showed a hazard ratio of 1.14, but this separation was not statistically significant ($p = 0.83$, CI $= [0.498, 2.624]$). Using the DF band in conjunction with $\theta = 0.5$ rather than $\theta = 0.9$ was no better.

As with the endpoint of death, we tried to determine whether energy in another frequency band might be used to predict MI. We found that in general, *low* energy in the given band was associated with high risk of MI. However, no frequency band provided particularly good risk assessment. Figure 5-7 shows the c-statistic for predicting MI within 90 days as a function of band location. The maximum c-statistic observed in this figure was 0.575.

In our investigations, we found that using unsmoothed MD data (that is, omitting the median filtering step when computing the MD series) yielded more accurate prediction of MI. However, this change still did not yield any statistically significant separation. In addition, we observed that using unsmoothed MD time series *decreased* discrimination performance for predicting death. Because smoothing is similar in ef-

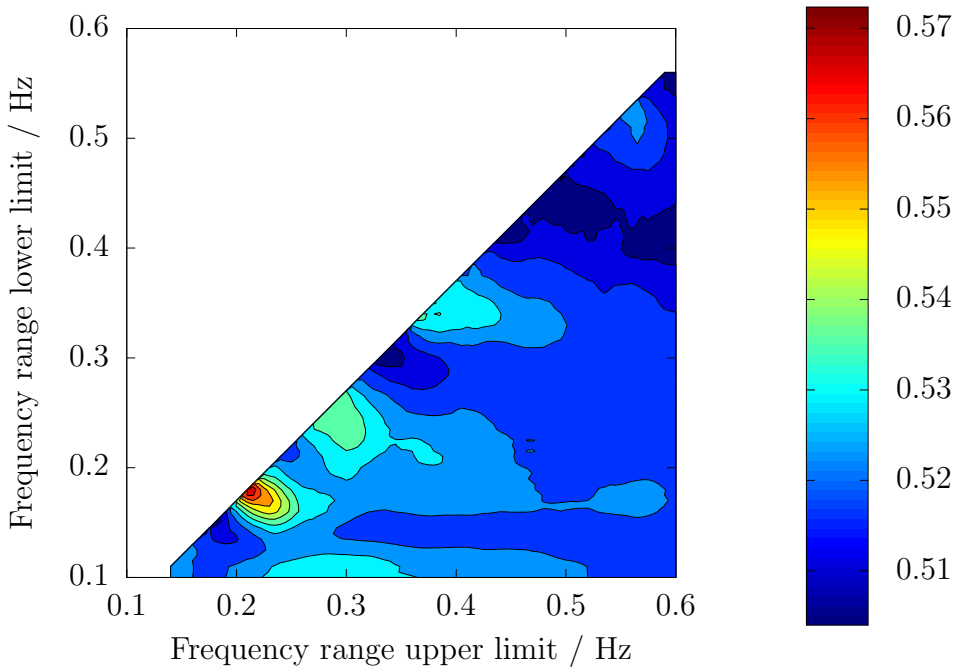



Figure 5-7: Prediction accuracy (c-statistic) of 90-day MI for single band energy as a function of band location, $\theta = 0.9$.

fect to a low-pass filter, this result suggests that morphologic differences that predict MI may be associated with higher frequencies than those that predict death.

Because the DF band was chosen for its accuracy in predicting risk of death, it is not entirely surprising that MV-DF does not predict MI well. It is believed that the DF band measures, specifically, conduction pathway irregularities that portend high risk for fatal arrhythmias. Work on MV-LF/HF has indicated that morphologic variability does indeed contain prognostic information that can be used in risk stratification for MI, but the mechanism for this is still not fully understood. Predicting risk for MI more accurately might require a new measure of morphologic variability that is associated with a different physiological mechanism and (at the very least) considers different frequency ranges.

## 5.5 Physiological interpretation

In this section, we propose a simplified model of myocardial activity that may provide a physiological interpretation of what the MV-DF metric measures. In particular, we attempt to provide a mechanism to explain conduction path changes that are associated with increased energy in the diagnostic frequency band (0.30 Hz–0.55 Hz) of the MD time series.

Some aberrant heart rhythms manifest as patterns with a fixed period. For example, in ventricular bigeminy and trigeminy, normal beats are interspersed with PVCs every two or three beats, respectively. PVCs (and other kinds of unusual beats) may also occur at larger intervals or only on an occasional and irregular basis. Such arrhythmias have, in certain cases, been observed to terminate as soon as heart rate goes below or above a certain threshold [25]. This suggests that the formation or conduction of aberrant beats is critically affected by some process that evolves with time, and on timescales that are on the order of the length of a heartbeat.

One possible mechanism for this is that a region of cells surrounding an ectopic focus (a group of cells, other than the SA node, that spontaneously generates a wave of depolarization) may be slow to repolarize completely and therefore cannot depolarize on every heartbeat, but only on average once every $N$ heartbeats (for some real number $N > 1$). It is on these heartbeats that a PVC occurs because the ectopic focus can actually excite the myocardium via the surrounding tissue.

When such an island of cells is located not at an ectopic focus but at some other part of the myocardium instead, it may lead to periodic conduction path changes rather than periodic ectopic beats. When all the cells of the myocardium can depolarize, the wave of depolarization takes some "preferred" course through the myocardium. However, if a region of cells is electrically unexcitable some of the time, then the wave is deflected around that region. Assuming that these two alternatives register as different heartbeat morphologies on the ECG, we might observe a sequence of morphologies like the following: $ABBBABBBABBB\ldots$, where $A$ represents a heartbeat in which the region of interest is depolarized, and $B$ represents one in which

it is not. This particular pattern is associated with $N = 4$.

Now assume, for simplicity, that the computation of morphologic difference captures the difference between the $A$ and $B$ morphologies, and no other differences. Then the ECG input above would be associated with a morphologic distance sequence $HLLHHLLHHLL\ldots$, where $H$ represents a high morphologic distance (between an $A$ beat and a $B$ beat) and $L$ represents a low morphologic distance (between $A$ and $A$, or $B$ and $B$). For an integer value of $N$, such a sequence will consist of alternating sequences of 2 $H$ values and $(N-2)$ $L$ values.

When $N$ is an integer $\geq 3$, the MD time series is periodic with period $N$, so the peak of its frequency content is given by $f/N$, where $f$ is the heart rate. However, when $N = 2$, the sequence is merely $HHHHHH\ldots$ and it has no energy content except at a frequency of 0. Therefore, $f/3$ represents an approximate ceiling on the frequency content for this phenomenon. For normal heart rates of 60–100 bpm, $f/3$ is no larger than 0.56 Hz. This frequency ceiling is essentially the same as the upper frequency limit of the DF band, 0.55 Hz. On the other end, the phenomenon we are considering could be associated with arbitrarily low frequencies, because as $N$ increases, $f/N$ can be arbitrarily close to 0. However, our empirical analysis showed that considering low frequencies does not improve the quality of discrimination. We speculate that the DF lower frequency limit we identified (0.30 Hz) arises because lower frequencies contain successively less and less information for prognostication—after all, even patients with completely healthy myocardium can show low-frequency changes in MD.

So far, we have modeled the myocardium with a single island of slowly repolarizing tissue. The model can easily be generalized to the more realistic situation of multiple such islands (possibly associated with different values of $N$) that produce conduction path changes independently. Assuming that the morphologic changes arising from two or more deflections are additive,[1] the MD time series can be considered to be a

[1]That is, the MD between a baseline conduction path and a path in which two deflections are taken is the sum of the MDs between the baseline path and the two paths in which the deflections are taken in isolation. Whether this condition holds may depend on the proximity of the two aberrant islands and the morphologic distance measure being used.

superposition of multiple time series, each of which is associated with a single island and behaves as described above. Because the power spectrum is linear in the MD time series, similar frequency-domain behavior should be observed under this more complex model.

## 5.6 Discussion

Morphologic variability identifies unstable beat-to-beat variations in the cardiac conduction path that portend high risk for cardiovascular death. In this chapter, we identified a diagnostic frequency (DF) band associated with the characteristic timescales of such variation. High DF energy in this band is strongly associated with risk of death. We also showed that having a small fraction (10%) of extremely high-MV activity predicts high risk for death, even more so than having generally high MV. We formulated the MV-DF method based on the results of analysis on ECG recorded during the DISPERSE-2 trial. Our confidence in the technique is bolstered by the fact that it provides good risk assessment on an independent test set of patient data from the MERLIN trial.

Not only does the MV-DF metric outperform the MV-LF/HF metric, it has the added advantage of having a simpler characterization, since it only relies on the energy in one frequency band rather than two. A preliminary investigation showed that using a two-band MV-LF/HF-like metric is unlikely to yield significantly better risk stratification. This was consistent with our observation that the LF/HF ratio itself only performed marginally better than the HF energy alone.

One weakness of MV-DF, as compared to MV-LF/HF, is that it does not do as good a job of risk stratifying for MI. Further research is needed to identify an MV-based metric that predicts risk for MI more accurately.

# Chapter 6

# Dynamics

In this chapter, we propose techniques for characterizing the dynamics of cardiac behavior using hidden Markov models (HMMs). We present preliminary results that suggest that some features of those models may have prognostic value for long-term risk stratification and short-term risk prediction.

Many interesting cardiac phenomena (for example, bigeminy and trigeminy, in which a premature ventricular contraction happens every second or third beat, respectively) can be naturally described in terms of particular sequences of classes of heartbeats. It has previously been shown that the presence of specific "motifs"—short and uncommon substrings in a discretized version of a stream of physiological data—can be used to separate clinically relevant classes of behavior [47]. Moreover, such motifs can be identified with no prior knowledge.

We believe that HMMs, because they are well suited for analysis of sequential data, could complement motif detection techniques for purposes of identifying physiologically interesting patterns. They are valuable because they could in principle identify certain patterns more robustly than with motifs in the presence of noise or other variations. HMMs use a probabilistic characterization of similarity, so they can efficiently recognize patterns that deviate from a given "template" and quantify the amount of deviation.

We developed and applied various approaches to quantify the dynamics of cardiac behavior, with two goals: identifying patients at high risk of death, and identifying

changes in dynamics that tended to occur shortly before death. In particular, we constructed models of two kinds of time series derived from ECG recordings: physiological labels of heartbeat types (as supplied by a cardiologist) and the MD time series (computed as described in Section 3.5).

This chapter begins with an overview of HMMs and their application to pattern detection. We then describe the procedures we used to train models from physiological data. We report on the results of experiments in which we trained HMMs and used various features of them to assess risk. Finally, we discuss possible future directions for this work.

## 6.1 Theory

Hidden Markov models are a generalization of Markov chains, and may be used to model the generation of both discrete- and continuous-valued time series. A given model provides a characterization of which output sequences (also called *emission sequences*) are likely to be generated by a particular process. More explicitly, for a fixed output length $m$, the probability that a particular output sequence $\mathbf{t} \equiv [t_1, t_2, \ldots t_m]$ is observed can be written as

$$\Pr(\mathbf{t}) = \Pr(t_1, t_2, \ldots, t_m) = \sum_{s_1,\ldots,s_m} \left( S(s_1)E(s_1, t_1) \prod_{j=2}^{m} T(s_{j-1}, s_j)E(s_j, t_j) \right), \quad (6.1)$$

where the functions $S$, $T$, and $E$ are parameters of the model. In the summation, each of the variables $s_1, \ldots, s_m$ may take on any value in $\{1, \ldots, n\}$, where $n$ is the number of *hidden states*.

This model may be interpreted in two stages, as follows. First, the values $[s_1, \ldots, s_m]$ (the *hidden state sequence*) are assumed to be generated by a Markov process where the probability of starting in state $s$ is given by $S(s)$ and the probability of transitioning to state $s'$ from state $s$ is given by $T(s, s')$. Then, for any $j$, the distribution of the emission $t_j$ is assumed to depend only on the corresponding state value $s_j$. In particular, for discrete-valued emissions, the probability that a particular value of $t_j$ is

observed is given by $E(s_j, t_j)$. Under these assumptions, the probability of observing a given output sequence $[t_1, \ldots, t_m]$ in conjunction with a given hidden state sequence $[s_1, \ldots, s_m]$ is

$$\Pr(s_1, \ldots, s_m; t_1, \ldots, t_m) = S(s_1)E(s_1, t_1) \prod_{j=2}^{m} T(s_{j-1}, s_j)E(s_j, t_j),$$

and (6.1) sums over all hidden state sequences to yield the total probability of observing $[t_1, \ldots, t_m]$. Although there are $n^m$ possible state sequences, a dynamic programming algorithm—the *forward algorithm*—can be used to evaluate (6.1) in $O(mn^2)$ time [40].

For discrete-valued emissions, $E$ simply gives a probability distribution over the $t$'s for each $s$. That is, we require that

$$E(s, t) \geq 0, \quad \forall s, t \qquad \text{and} \qquad \sum_t E(s, t) = 1, \quad \forall s.$$

For continuous-valued emissions, the form of Equation (6.1) is unchanged. However, $E$ now instead specifies a probability density function (PDF) over the values of $t$, satisfying

$$E(s, t) \geq 0, \quad \forall s, t \qquad \text{and} \qquad \int_{-\infty}^{\infty} E(s, t)\, dt = 1, \quad \forall s,$$

and the output of (6.1) is not the probability of observing $[t_1, \ldots, t_m]$ but a probability density instead.

An HMM simulates both the production of an underlying hidden state sequence over time and the generation of an emission sequence from the hidden states. Hidden Markov models are so named because the emission sequences are typically assumed to be directly observable but the hidden states are not (and therefore must be recovered from the emissions). Consequently, algorithms developed for HMMs can be used to analyze time-series signals that are considered to be noisy or indirect indicators of some underlying phenomenon that cannot be directly measured. For example, in HMMs used in speech recognition, the emission sequence might represent the recorded acoustic data while hidden states might represent phonemes.

In models of cardiac behavior, the emission sequences we consider are streams of symbols or numbers derived from ECG recordings. In this work, we analyze two kinds of emission sequences—sequences of cardiologist-supplied labels indicating the type of each heartbeat (discrete-valued) and the MD time series (continuous-valued)—that provide information at the level of individual heartbeats. In this context, the hidden states may be interpreted as encoding information about the state of the heart that is relevant to individual heartbeats. Different physiological states have varying propensities to produce specific emissions (e.g. MD values or different types of heartbeats), and this variation is encoded in the emission probabilities. Meanwhile, the evolution of the hidden state is governed by various physiological factors. For example, the depolarization and repolarization characteristics of the myocardium lead to dependencies between consecutive heartbeats. Moreover, the behavior of the heart is influenced by control systems external to the myocardium.

We use two general approaches for the analysis of cardiac HMMs. These approaches are described briefly here, and our specific experiments are described in more detail in Section 6.3.

First, features of the HMMs acquired during training may be used to characterize behavior. An HMM is a succinct but rich characterization of the kinds of patterns that are prevalent in the training data. We show that there is a specific sense in which the "complexity" of cardiac activity increases in the period immediately preceding death. This measure of complexity is obtained by training models of various sizes for a single input sequence and determining at what point larger models cease to explain the data better.

Second, because HMMs give a measure of the likelihood of a sequence of observations, one can interpret a model as identifying a class of similar observation sequences. We can use the likelihood given in (6.1) as a measure of *consistency* between the model and a new observation sequence, that is, a measure of how likely it was that the model was responsible for the new output. We explore one example of measuring this consistency over time. A change in dynamics over time for a single patient could be used to identify when that patient is deteriorating from a stable condition. We

note that such measures could be used in various other ways for prognosis and risk assessment. For example, testing consistency between patients might be useful for identifying common classes of pathological cardiac behavior or clustering patients by their patterns of physiological activity.

## 6.2 Training

The goal of HMM training is to find, given a set of training sequences[1] $\{\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_l\}$, a set of model parameters that maximize the likelihood of observing the training data,

$$\text{Likelihood} = \prod_{i=1}^{l} \Pr(\mathbf{t}_i), \tag{6.2}$$

while possibly respecting certain constraints on the model (such as fixing the number of hidden states $n$).

For both emission types (cardiologist-supplied labels and the MD time series), we used the Baum-Welch algorithm [40] to train HMMs from a single input sequence (i.e. $l = 1$). The Baum-Welch algorithm takes as input one or more emission sequences, the number of hidden states $n$, and an initial guess for the model parameters $S(s)$, $T(s, s')$, and $E(s, t)$. It then returns a model with $n$ hidden states that is a local maximum of the likelihood (6.2).[2] Because the Baum-Welch algorithm is only guaranteed to converge to a local maximum, we ran it multiple times with randomly chosen initial conditions and selected the output model that yielded the highest likelihood.

To train HMMs to model sequences of cardiologist-supplied heartbeat labels, we used data from the Sudden Cardiac Death Holter Database (SDDB) in the MIT-BIH Physionet dataset [20]. Each patient in that database died of sudden cardiac death and had a continuous ECG recording leading up to the time of death. The recordings were between 3 and 26 hours long. For the purposes of our experiments we used data from 9 patients who both had cardiologist-supplied annotations for each beat and

[1] We'll identify sequences of symbols using boldface letters like $\mathbf{t}$.

[2] In fact, there is no known analytic way to determine the parameters that globally maximize the likelihood [40].

suffered ventricular fibrillation while under monitoring.

Each beat was labeled with one of 18 symbols indicating the type of the beat (normal, premature ventricular contraction, bundle branch block beat, etc.). We used the beat-to-beat intervals to augment each label with one of the following three indicators of instantaneous heart rate: tachycardia (HR $>$ 100 bpm), normal rhythm (60 bpm $<$ HR $\leq$ 100 bpm), or bradycardia (HR $\leq$ 60 bpm). Thus, each beat was labeled with one of 54 possible symbols indicating the beat type and rhythm. We trained HMMs using these sequences of symbols.

It was common for not all possible symbols to be represented in the training data. This can cause problems because Baum-Welch and other maximum likelihood estimation techniques estimate the probability of an event to be its observed frequency in the training data—that is, $m/n$ where $m$ is the number of times the event was seen to occur and $n$ is the number of times it could have occurred. Baum-Welch can therefore yield models that assign a zero $(0/n)$ or undefined $(0/0)$ probability to a test data sequence; this can happen, for example, when the test data contains emitted symbols that were never observed in the training data. We avoided this problem by adding pseudocounts [45] to the data. In the presence of pseudocounts, the probability of an event is estimated to be $(\alpha + m)/(k\alpha + n)$ where $k$ is the number of different outcomes and $\alpha$ is some positive number, typically 1. Consequently, no probability is ever estimated to be 0, and in the absence of any data ($m = n = 0$), the estimated probability distribution defaults to a uniform distribution over the possible outcomes.

To train HMMs to model the MD time series, we used ECG data from 18 patients selected from the placebo group of the MERLIN (TIMI36) trial [33]. The set included 6 patients who died within 7 days of a NSTEACS and had ECG data starting within 48 hours of the initial event and continuing to within 24 hours of death. For comparison, 12 additional patients were randomly selected from those who did not die during the 90-day followup period. The MD time series for each patient was computed as described in Section 3.5. We trained HMMs in which, for each hidden state $s$, $E(s, t)$ (the distribution of emissions from $s$) was constrained to take the form of a Gaussian distribution. The Baum-Welch algorithm was then used to estimate optimal values

of the parameters $\mu_s$ and $\sigma_s$ (the mean and standard deviation of the distribution, respectively) for each hidden state $s$, as well as the transition parameters. Since $E(s,t)$ is nonzero for all possible values of $s$ and $t$, there was no need to use pseudocounts in this setting.

## 6.3 Experiments

### 6.3.1 Complexity

The Baum-Welch algorithm requires as input the number of hidden states $n$ to use in the resulting model. One important consideration in HMM training is the problem of *model selection*—selecting an appropriate number of states. Although using a larger number of states always yields a model with better explanatory power for the training data (that is, the likelihood on the training data is at least as large as with a smaller number of states), one runs the risk of overfitting the training data if the number of states is too large. The Akaike information criterion (AIC) [1] is one criterion that has been proposed for selecting the number of states in discrete-emission HMMs. For a training sequence $\mathbf{t}$, the criterion is to choose the number of states $n$ that maximizes the quantity $(\log \Pr_n(\mathbf{t}) - Q(n))$, where $\Pr_n(\mathbf{t})$ is the likelihood of the data under a model with $n$ hidden states, and $Q(n)$ is the number of free parameters in a model with $n$ hidden states (which increases as $O(n^2)$). By penalizing more complex models, the AIC ensures that the use of a larger model is justified by its contribution of additional explanatory power.

For each of the 9 patients in the SDDB dataset, we used the data from the first 60 minutes of the recording to train an HMM. For each patient, we then used the data from the 60 minutes of the recording immediately preceding ventricular fibrillation to train another HMM. For all models, we used the AIC to select an appropriate number of states between 2 and 16. The model sizes selected by the AIC are shown in Table 6.1. We found that the models trained on the last 60 minutes of data usually required larger numbers of hidden states. This suggests that data collected

| Patient ID | First 60 min. | Last 60 min. |
|---|---|---|
| 30 | 4 | 3 |
| 32 | 2 | 2 |
| 34 | 5 | 6 |
| 36 | 2 | 2 |
| 41 | 5 | 13 |
| 45 | 7 | 5 |
| 46 | 5 | 10 |
| 51 | 5 | 6 |
| 52 | 5 | 9 |

Table 6.1: Number of states selected by the AIC for each patient for models of the first 60 minutes and the last 60 minutes preceding ventricular fibrillation.

during those periods was more unpredictable, or complex, than data collected at the beginning of the recording, in the sense that more transition and emission parameters are needed to characterize the behavior of the training data.

### 6.3.2 Consistency

We attempted to quantify the consistency of patient dynamics over time in order to obtain useful information for short-term risk prediction. Obviously, an arrhythmic event like ventricular fibrillation represents a dramatic change in dynamics. We speculate that these large changes might be preceded by much subtler changes in dynamics in the minutes or hours before the arrhythmic event. If such changes could be detected early on, they could be used to warn patients or caregivers so that preventive measures could be taken.

For each of the 9 patients in the SDDB dataset, we evaluated the consistency between an HMM trained on the first 60 minutes of data and windows of data taken from the same patient at other times. For each of these windows of data, we computed a measure of agreement with the model, which we refer to as *normalized log likelihood*, as follows. For an emission sequence $\mathbf{t} = [t_1, t_2, \ldots, t_m]$, we first use the forward algorithm to evaluate $\Pr(\mathbf{t})$. Then the normalized log likelihood is defined as

$$\text{Normalized log likelihood} = \frac{\log \Pr(\mathbf{t})}{m}. \tag{6.3}$$

As with the likelihood $\Pr(\mathbf{t})$, a high normalized log likelihood indicates that the output sequence being evaluated is similar to the data on which the model was trained.

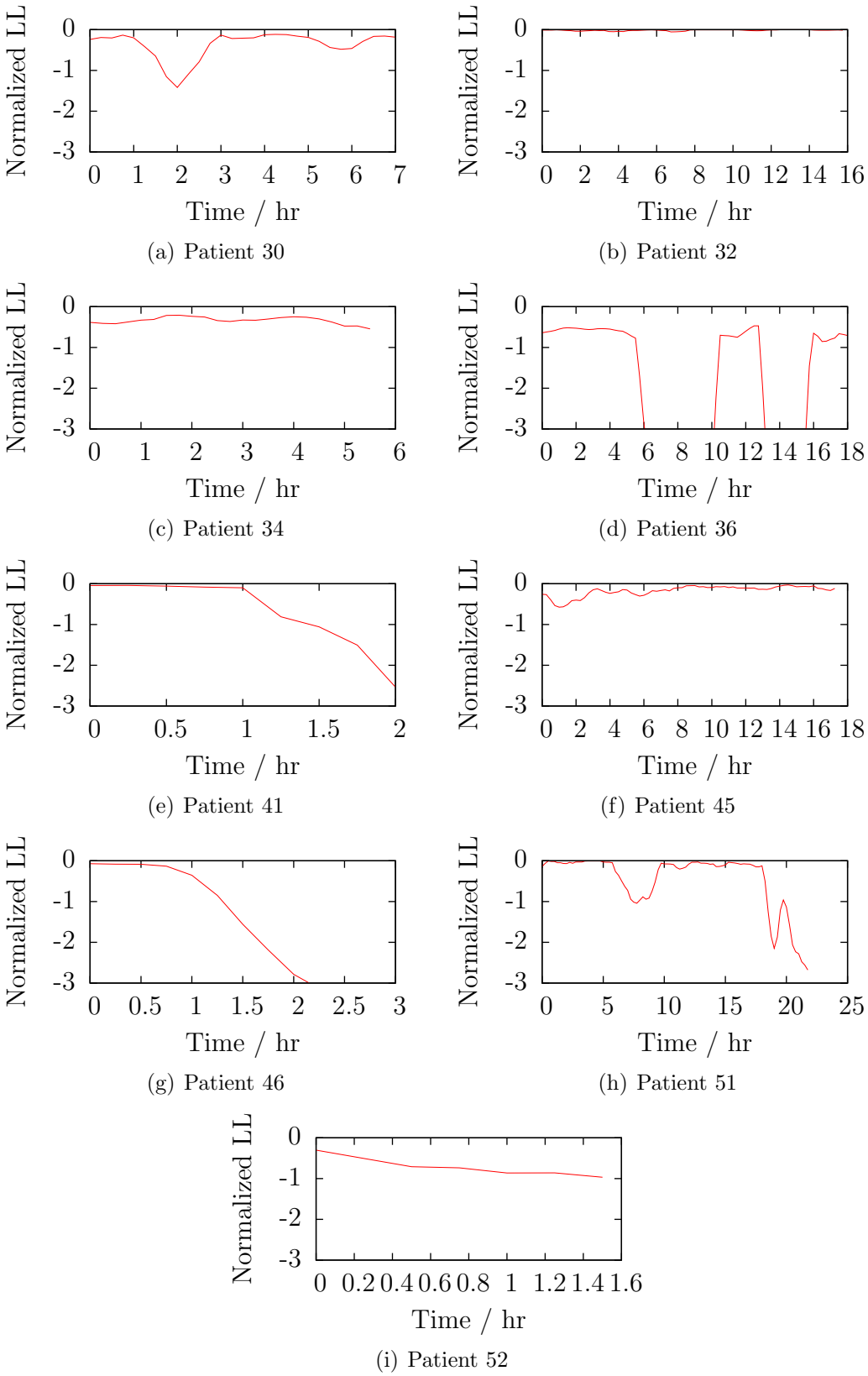
The purpose of computing normalized log likelihood rather than using the likelihood itself is to ensure that comparable figures can be obtained from emission sequences of different lengths. Ordinarily, under a given discrete-emission HMM, longer sequences with similar composition are assigned a lower probability of occurring. More precisely, the average probability of an emission sequence of length $m$ is $1/b^m$, where $b$ is the number of possible emitted symbols. Taking the logarithm and dividing by the length of the sequence, $m$, corrects for this effect to yield a quantity that is related to the likelihood and is comparable between sequences of different lengths.[3]

The resulting normalized log likelihoods for each patient are shown in Figure 6-1. Each plot represents data from a different patient and shows the normalized log likelihood with respect to time. Notably, 4 of the 9 patients (41, 46, 51, and 52) started to show deviation from the original model (normalized LL $\leq -1$) shortly—within 1 to 3 hours—before the time of death. (Patient 36's unusual-looking activity is the result of intermittently switching from a rhythm of primarily normal beats—on which the model was trained—to one of primarily premature ventricular contractions.)

By themselves, these results do not, by any means, prove conclusively that a dropoff in consistency portends cardiovascular death: after all, as long as there is any variation in dynamics over time—a reasonable assumption—a model trained on data taken at a particular time is naturally expected to be less consistent with data collected at a later time. However, our results leave open the possibility that HMMs might be able to distinguish such inconsistencies in a clinically useful way. In order for this hypothesis to be validated, experiments comparing the consistency of dynamics in low-risk and high-risk patients should show that low-risk patients have more stable dynamics than high-risk patients like the ones we analyzed here.

We attempted to reproduce these results on models trained on the MD time series

[3]For continuous-emission HMMs the situation is slightly more complicated. Since the output of (6.1) represents a probability density in $\mathbb{R}^m$, the probability densities obtained for different sequences do not even have the same units unless the sequences have the same length $m$. However, the same normalization procedure converts these densities into comparable quantities.




(a) Patient 30

(b) Patient 32

(c) Patient 34

(d) Patient 36

(e) Patient 41

(f) Patient 45

(g) Patient 46

(h) Patient 51

(i) Patient 52

Figure 6-1: Stability of trained model over time for each of 9 patients.

of patients in the MERLIN study. We trained HMMs using data from 6 patients for whom ECG recordings were available until within 24 hours of death. However, no analogous results were observed in this population.

## 6.4 Summary and Future Work

In this chapter, we proposed, and reported preliminary results for, two classes of techniques for modeling cardiac dynamics and using those models to perform risk assessment. First, the models themselves (i.e. the set of transition and emission parameters associated with an HMM) represent a concise characterization of the kinds of patterns that typically appear in a particular training sequence. Various properties of these models may provide valuable information for purposes of risk stratification and short-term risk prediction. As a simple example, the transition and emission parameters might encode some new notion of morphologic variability (in models of the MD time series) or heart rate variability (in models of the RR interval series). We showed that one measure of the "complexity" of a model increases for cardiac data taken near the time of death. Future analysis of cardiac HMMs may identify other features that may correlate with short-term or long-term risk.

Second, we defined and applied the normalized log likelihood, a notion of consistency that may be used to decide how probable it is that a given model could have generated a given sequence of previously unseen data. Our results provide support for the idea that the consistency of a patient's dynamics over time may be used to determine when the patient is deteriorating from a stable condition. Studying changes in dynamics at other critical times—such as right after an ACS—might also yield important information. In addition, consistency between the dynamics of multiple patients may be used to cluster patients with similar physiological activity or similar risk profiles.

The line of research described here is based on the observation that many cardiac irregularities manifest themselves as certain kinds of patterns in the time domain on the timescales of individual heartbeats. Bigeminy (and trigeminy, and other kinds

of rhythms), morphologic variability (under the interpretation in Section 5.5), and deceleration capacity can all be interpreted this way. Moreover, in some of these cases (bigeminy, and possibly morphologic variability) specific beat-to-beat interactions, mediated by repolarization and depolarization characteristics of the myocardium, have been implicated in these phenomena. Consequently, sequence-based approaches such as HMMs and motif detection may be seen as complementary to previous risk stratification approaches—such as HRV and MV—in which frequency-domain and time-domain analyses have proven to be valuable. In addition, we believe that HMMs, because they use a probabilistic characterization of similarity between a sequence and a model, may recognize certain kinds of patterns more reliably than non-probabilistic techniques can, especially in the presence of noise and variations.

The HMM techniques we considered may be applied to various streams of physiological data. We worked with sequences of heartbeat type labels and the MD time series, but consideration of other sequences like the RR time series—which would serve as an extension of research on HRV and DC—might also be fruitful. We believe that a systematic examination of these techniques on additional data sources represents the most immediate and straightforward next step from the current work.

One significant limitation of this work is that the small amount of data we used (limited both by computational resources and the datasets themselves) makes it difficult to obtain statistically significant results. The SDDB dataset is useful because it provides class annotations for each heartbeat. However, its utility for running controlled experiments is limited because of its size (9 patients) and because it does not contain any negative examples (i.e. comparable patients who did not die). Progress in automatic beat classification techniques might allow similar beat annotations to be obtained on larger datasets (like the DISPERSE-2 or MERLIN databases) where they are at present unavailable, thus allowing the hypotheses we created on the SDDB dataset to be tested elsewhere.

In addition to the two approaches we proposed in this chapter, we suggest an additional possible way of interpreting HMM results. Given an HMM and a sequence of emissions $[t_1, t_2, \ldots, t_m]$, the Viterbi algorithm can be used to recover the most

likely hidden state sequence (or sequences) that might have generated that emission sequence,

$$\arg \max_{s_1,\ldots,s_m} S(s_1)E(s_1,t_1)\prod_{j=2}^{m} T(s_{j-1},s_j)E(s_j,t_j).$$

This hidden state sequence $[s_1,\ldots,s_m]$ can be interpreted in a number of ways. In a simple model, a particular state or sequence of states may be thought of as a "signature" for particular motifs. Then, the process of computing the hidden state sequence can convert a noisy and possibly continuous-valued input sequence to a cleaner discrete-valued sequence for subsequent analysis. In a more complex model—which could be either hand-tuned or trained using data from multiple patients—it is possible that states or groups of states could be identified with specific pathophysiological conditions. Since the hidden state at a particular time is thought to encode information about the physiological state of the heart, the event of entering or staying in particular hidden states may be clinically relevant. The utility of such information is not necessarily limited to risk stratification. It might also more generally be used as a tool for diagnosing specific conditions in an automated way.

# Chapter 7

# Summary and Conclusions

## 7.1 Summary

People who have suffered an acute coronary syndrome remain at elevated but varying levels of risk of death and reinfarction following the acute event. One factor contributing to this risk is the unusual conduction characteristics of ischemic myocardial tissue, which may lead to fatal arrhythmias; it is believed that morphologic variability correlates with extent and effect of ischemic tissue. Morphologic variability, when combined with other prognostication techniques, provides information that is useful for identifying the best treatment options for specific patients.

The overall goal of this project was to improve short- and long-term risk assessment post-ACS using morphologic variability-based techniques. Our first goal was to develop an MV-based risk stratification technique that significantly outperformed existing MV-based measures. We believed that such a goal was attainable because previous research in morphologic variability (resulting in the MV-LF/HF measure) was somewhat artificially constrained in order to facilitate initial comparison with HRV-based measures. The result of our work, and the first main contribution of this thesis, was a new risk measure, MV-DF (for *diagnostic frequencies*), which not only identifies patients at heightened risk of death more accurately than MV-LF/HF but, to the best of our knowledge, outperforms all previously published Holter ECG-based risk stratification techniques. MV-DF provides significant information even

after controlling for other risk factors, and it is especially valuable for short-term risk stratification (assessing risk within the first 90 days after an ACS).

We developed the MV-DF technique empirically in a number of stages. First, we evaluated various ways of measuring morphologic differences—a computation that is at the heart of MV-based methods. We adapted three previously known techniques (EMD, Fréchet, and Hausdorff distances) to the task of quantifying morphologic differences between ECG signals, comparing them to the existing method based on dynamic time-warping (DTW). We concluded that DTW is not only the best suited of the methods for identifying the kinds of morphologic changes that are relevant to assessing risk for cardiovascular death, but also that there is a sense in which DTW is more stable than the other measures in the presence of noise.

Second, we identified a better way of integrating multiple measurements of a patient's morphologic variability to produce a final assessment of risk. We discovered that patients showing extremely high MV, even if only for a small fraction of the time, were at higher risk than patients who had generally elevated but stable MV. In order to take advantage of this fact to improve risk assessment, we introduced the notion of an *output threshold* in the computation of the MV risk measure. One unanticipated challenge of this approach was that using a patient's most extreme behavior to characterize his or her risk makes the measure sensitive to noise—after all, the windows with highest MV often reflect noise rather than physiologically interesting phenomena. We found that selecting more conservative output thresholds allowed us to reap most of the performance gains while limiting the potential of noise to corrupt the prognosis.

Finally, we identified a distinctive frequency band (the *diagnostic frequency band*, 0.30–0.55 Hz) such that increased MD energy within that band was significantly associated with higher risk. Incorporating the DF band and the improved output threshold yielded a risk measure (MV-DF) that identified patients at higher risk of death post-ACS, with a c-statistic of 0.771 on a dataset of 764 patients from the DISPERSE-2 study. For comparison, the c-statistic was 0.724 for the previous MV-LF/HF risk measure. Patients with the highest quartile of MV showed more than

five-fold elevated risk of death when compared to patients in the lower three quartiles.

Our confidence in the MV-DF technique was bolstered by the fact that similar improvements in prognostic accuracy were observed on a dataset of 2,302 patients from the placebo group of the MERLIN trial, even though the MV-DF measure was trained independently of this dataset. We presented the results of analysis showing that the MV-DF measure provides fairly stable prognoses and demonstrates a graded response.

The second main contribution of this thesis was the presentation of a framework for analyzing cardiac behavior using hidden Markov models. The heart can be considered to be a dynamical system in which dependencies from beat to beat are mediated by the depolarization and repolarization characteristics of the myocardium and by other cardiac control systems external to the myocardium. We suggest that HMMs, as a method of analyzing sequential data, may complement existing frequency-domain and time-domain techniques for characterizing cardiac activity.

We proposed three ways of obtaining dynamical data from HMMs and using it in risk assessment. We presented preliminary results indicating the viability of the first two. First, an HMM can serve as a characterization of the sequence(s) on which it was trained, and the idea of extracting features directly from the HMM for use in risk stratification is similar to the principle behind ECG-based risk stratification techniques like MV and HRV. Second, the log likelihood of a model on new data can be used as a measure of how similar that data is to the data on which the model was trained. We speculate that these similarity measures may be used to determine when a patient's condition deviates from some stable state—which would have implications for short-term risk assessment. Finally, with enough training data, one could imagine developing an HMM in which specific hidden states are identified with specific physiological conditions. The fact that a patient entered specific states in such a model (either once or repeatedly) might be used to modify that patient's risk estimate or to diagnose certain conditions.

## 7.2 Conclusions

The MV-DF measure proposed in this thesis represents a significant improvement in prognostication quality over previously known long-term Holter-based risk measures. We showed that MV-DF performs better than HRV and DC at identifying patients at risk of death following an ACS. In addition, MV-DF identifies patients at high risk even after controlling for other significant risk factors such as poor left ventricular ejection fraction and age.

MV-DF is most valuable when used in conjunction with other risk measures. MV is a good complement to DC because the techniques are believed to assess different pathophysiological mechanisms. These ECG risk measures, as well as other techniques—catheterization procedures, biomarker tests, and non-invasive imaging techniques such as the echocardiogram—all yield information about specific aspects of cardiac health. This information can be used to guide the selection of long-term treatment options and to refine an initial prognosis obtained using techniques like the TIMI risk score.

Among the long-term risk measures, long-term ECG stands out because it is both inexpensive and safe. Biomarker tests require samples to be sent to a lab for analysis; imaging studies and catheterization procedures can only be performed and interpreted by highly skilled clinicians. Catheterization procedures are also invasive and pose considerable risk. In contrast, Holter monitor data is already routinely recorded for many patients during hospitalization, so long-term ECG risk measures like MV-DF can be computed at essentially no additional cost or risk. Moreover, no special training is needed to interpret the results of long-term ECG risk measures.

Because long-term ECG risk measures can be acquired essentially for free and at no risk to patients, we believe that they should play an important role in the selection of appropriate treatments. Patients who show low risk based on tests that include an ECG measure like MV-DF may be spared other invasive tests and invasive or expensive treatments. Patients who show high risk may be referred for further tests (like catheterization procedures) or aggressive treatments that would not have been

considered in the absence of supporting information.

It is unfortunate that long-term ECG risk measures are currently not widely used to guide clinical decisions, despite their value and the ease with which they can be acquired. The MV-DF technique is computationally simple enough that there would be no difficulty in running it on a typical personal computer or implementing it in hardware or software on an ECG monitoring device. The use of MV-DF—or indeed, of any of the widely known ECG risk stratification techniques—represents a cost-effective and practical way to improve clinical care for the millions of people who are hospitalized following acute coronary syndromes each year. We hope that the increasing weight of the evidence in support of ECG risk measures leads to their inclusion in common clinical practice.

# Bibliography


[1] H. Akaike. A new look at the statistical model identification. In *IEEE Transactions on Automatic Control*, volume 19, pages 716–723, December 1974.

[2] J. L. Anderson, C. D. Adams, E. M. Antman, C. R. Bridges, R. M. Califf, D. E. Casey, W. E. Chavey, F. M. Fesmire, J. S. Hochman, T. N. Levin, A. M. Lincoff, E. D. Peterson, P. Theroux, N. K. Wenger, R. S. Wright, S. C. Smith, A. K. Jacobs, C. D. Adams, J. L. Anderson, E. M. Antman, J. L. Halperin, S. A. Hunt, H. M. Krumholz, F. G. Kushner, B. W. Lytle, R. Nishimura, J. P. Ornato, R. L. Page, and B. Riegel. ACC/AHA 2007 Guidelines for the Management of Patients With Unstable Angina/Non-ST-Elevation Myocardial Infarction: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to Revise the 2002 Guidelines for the Management of Patients With Unstable Angina/NonST-Elevation Myocardial Infarction) Developed in Collaboration with the American College of Emergency Physicians, the Society for Cardiovascular Angiography and Interventions, and the Society of Thoracic Surgeons Endorsed by the American Association of Cardiovascular and Pulmonary Rehabilitation and the Society for Academic Emergency Medicine. *Journal of the American College of Cardiology*, 50:e1–e157, August 2007.

[3] E. M. Antman, M. Cohen, P. J. L. M. Bernink, C. H. McCabe, T. Horacek, G. Papuchis, B. Mautner, R. Corbalan, D. Radley, and E. Braunwald. The TIMI risk score for unstable angina/non-ST elevation MI: a method for prognostication and therapeutic decision making. *JAMA*, 284:835–842, 2000.

[4] P. W. Armstrong, Y. Fu, W.-C. Chang, E. J. Topol, C. B. Granger, A. Betriu, F. V. de Werf, K. L. Lee, and R. M. Califf. Acute Coronary Syndromes in the GUSTO-IIb Trial: Prognostic Insights and Impact of Recurrent Ischemia. *Circulation*, 98:1860–1868, 1998.

[5] B. Aronov, S. Har-Peled, C. Knauer, Y. Wang, and C. Wenk. Fréchet distance for curves, revisited. In *ESA'06: Proceedings of the 14th conference on Annual European Symposium*, pages 52–63, London, UK, 2006. Springer-Verlag.

[6] P. Barthel, R. Schneider, A. Bauer, K. Ulm, C. Schmitt, A. Schömig, and G. Schmidt. Risk Stratification After Acute Myocardial Infarction by Heart Rate Turbulence. *Circulation*, 108:1221–1226, August 2003.

[7] A. Bauer, J. W. Kantelhardt, P. Barthel, R. Schneider, T. Mäkikallio, K. Ulm, K. Hnatkova, A. Schömig, H. Huikuri, A. Bunde, M. Malik, and G. Schmidt. Deceleration capacity of heart rate as a predictor of mortality after myocardial infaraction: cohort study. *Lancet*, 367:1674–1681, 2006.

[8] S. A. Ben-Haim, A. Gil, and Y. Edoute. Beat-to-beat morphologic variability of the electrocardiogram for the evaluation of chest pain in the emergency room. *The American Journal of Cardiology*, 70(13):1139 – 1142, 1992.

[9] G. E. Billman, P. J. Schwartz, and H. L. Stone. Baroreceptor reflex control of heart rate: a predictor of sudden cardiac death. *Circulation*, 66:874–880, 1982.

[10] E. Boersma, K. S. Pieper, E. W. Steyerberg, R. G. Wilcox, W.-C. Chang, K. L. Lee, K. M. Akkerhuis, R. A. Harrington, J. W. Deckers, P. W. Armstrong, A. M. Lincoff, R. M. Califf, E. J. Topol, and M. L. Simoons. Predictors of Outcome in Patients With Acute Coronary Syndromes Without Persistent ST-Segment Elevation : Results From an International Trial of 9461 Patients. *Circulation*, 101(22):2557–2567, 2000.

[11] J. E. Calvin, L. W. Klein, B. J. VandenBerg, P. Meyer, J. V. Condon, R. J. Snell, L. M. Ramirez-Morgen, and J. E. Parrillo. Risk stratification in unstable angina. Prospective validation of the Braunwald classification. *Journal of the American Medical Association*, 273(2):136–141, January 1995.

[12] C. P. Cannon, E. Braunwald, C. H. McCabe, D. J. Rader, J. L. Rouleau, R. Belder, S. V. Joyal, K. A. Hill, M. A. Pfeffer, and A. M. Skene. Intensive versus Moderate Lipid Lowering with Statins after Acute Coronary Syndromes. *New England Journal of Medicine*, 350(15):1495–1504, April 2004.

[13] C. P. Cannon, S. Husted, R. A. Harrington, B. M. Scirica, H. Emanuelsson, G. Peters, and R. F. Storey. Safety, Tolerability, and Initial Efficacy of AZD6140, the First Reversible Oral Adenosine Diphosphate Receptor Antagonist, Compared With Clopidogrel, in Patients With NonST-Segment Elevation Acute Coronary Syndrome: Primary Results of the DISPERSE-2 Trial. *Journal of the American College of Cardiology*, 50(19):1844–1851, November 2007.

[14] G. F. Chess, R. M. K. Tam, and F. R. Calaresu. Influence of cardiac neural inputs on rhythmic variations of heart period in the cat. *American Journal of Physiology*, 228(3):775–780, 1975.

[15] T. H. Cormen, C. E. Leiserson, R. R. Rivest, and C. Stein. *Introduction to Algorithms*, chapter 10. MIT Press, 2nd edition, September 2001.

[16] D. R. Cox. Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220, 1972.

[17] C. J. DeFrances, K. A. Cullen, and L. J. Kozak. National Hospital Discharge Survey: 2005 annual summary with detailed diagnosis and procedure data. *Vital and Health Statistics*, 13(165), 2007. National Center for Health Statistics.

[18] T. Eiter and H. Mannila. Computing Discrete Fréchet Distance. Technical Report CD-TR 94/64, Christian Doppler Laboratory for Expert Systems, TU Vienna, April 1994.

[19] N. El-Sherif, B. J. Scherlag, R. Lazzara, and R. R. Hope. Re-entrant ventricular arrhythmias in the late myocardial infarction period. 1. Conduction characteristics in the infarction zone. *Circulation*, 55(5):686–702, 1977.

[20] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000 (June 13). Circulation Electronic Pages: http://circ.ahajournals.org/cgi/content/full/101/23/e215.

[21] C. B. Granger, R. J. Goldberg, O. Dabbous, K. S. Pieper, K. A. Eagle, C. P. Cannon, F. V. de Werf, A. Avezum, S. G. Goodman, M. D. Flather, and K. A. A. Fox. Predictors of Hospital Mortality in the Global Registry of Acute Coronary Events. *Archives of Internal Medicine*, 163(19):2345–2353, 2003.

[22] J. A. Hanley and B. J. McNeil. The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve. *Radiology*, 143(1):29–36, April 1982.

[23] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge. Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9):850–863, 1993.

[24] M. E. Josephson and A. L. Wit. Fractionated electrical activity and continuous electrical activity: fact or artifact? *Circulation*, 70(4):529–532, 1984.

[25] S. Kinoshita, F. Okada, G. Konishi, M. Kinoshita, and S. Ogawa. Bradycardia- and tachycardia-dependent termination of ventricular bigeminy: Mechanism of ventricular extrasystoles with fixed coupling. *American Heart Journal*, 129(3):557, 564 1995.

[26] R. E. Kleiger, J. P. Miller, and J. T. Bigger. Decreased heart rate variability and its association with increased mortality after acute myocardial infarction. *American Journal of Cardiology*, 59:256–262, 1987.

[27] H. M. Krumholz, P. S. Douglas, L. Goldman, and C. Waksmonski. Clinical utility of transthoracic two-dimensional and Doppler echocardiography. *Journal of the American College of Cardiology*, 24(1):125–131, 1994.

[28] L. S. Lilly, editor. *Pathophysiology of heart disease*. Lippincott Williams & Wilkins, 2003.

[29] N. R. Lomb. Least-squares frequency analysis of unequally spaced data. *Astrophysics and Space Science*, 39:446–462, 1976.

[30] M. Malik, J. Camm, J. T. Bigger, et al. Heart rate variability: standards of measurement, physiological interpretation, and clinical use. *Circulation*, 93(5):1043–1065, 1996.

[31] D. A. Morrow, E. M. Antman, A. Charlesworth, R. Cairns, S. A. Murphy, J. A. de Lemos, R. P. Giugliano, C. H. McCabe, and E. Braunwald. TIMI risk score for ST-elevation myocardial infarction: a convenient bedside clinical score for risk assessment at presentation. *Circulation*, 102:2031–2037, 2000.

[32] D. A. Morrow, E. M. Antman, L. Parsons, J. A. de Lemos, C. P. Cannon, R. P. Giugliano, C. H. McCabe, H. V. Barron, and E. Braunwald. Application of the TIMI risk score for ST-elevation MI in the National Registry of Myocardial Infarction 3. *JAMA*, 286:1356–1359, 2001.

[33] D. A. Morrow, B. M. Scirica, E. Karwatowska-Prokopczuk, S. A. Murphy, A. Budaj, S. Varshavsky, A. A. Wolff, A. Skene, C. H. McCabe, and E. Braunwald. Effects of Ranolazine on Recurrent Cardiovascular Events in Patients With Non-ST-Elevation Acute Coronary Syndromes: The MERLIN-TIMI 36 Randomized Trial. *JAMA*, 297(16):1775–1783, 2007.

[34] A. J. Moss, W. Zareba, W. J. Hall, H. Klein, D. J. Wilber, D. S. Cannom, J. P. Daubert, S. L. Higgins, M. W. Brown, and M. L. Andrews. Prophylactic Implantation of a Defibrillator in Patients with Myocardial Infarction and Reduced Ejection Fraction. *New England Journal of Medicine*, 346(12):877–883, March 2002.

[35] C. S. Myers and L. R. Rabiner. A comparative study of several dynamic time-warping algorithms for connected word recognition. *The Bell System Technical Journal*, 60(7):1389–1409, 1981.

[36] S. E. Nissen, E. M. Tuzcu, P. Schoenhagen, B. G. Brown, P. Ganz, R. A. Vogel, T. Crowe, G. Howard, C. J. Cooper, B. Brodie, C. L. Grines, and A. N. DeMaria. Effect of Intensive Compared With Moderate Lipid-Lowering Therapy on Progression of Coronary Atherosclerosis: A Randomized Controlled Trial. *JAMA*, 291(9):1071–1080, 2004.

[37] T. Oates, L. Firoiu, and P. R. Cohen. Clustering Time Series with Hidden Markov Models and Dynamic Time Warping. In *Proceedings of the IJCAI-99 Workshop on Sequence Learning*, 1999.

[38] J. F. Pombo, B. L. Troy, and R. O. Russell. Left Ventricular Volumes and Ejection Fraction by Echocardiography. *Circulation*, 43(4):480–490, 1971.

[39] L. Rabiner and B.-H. Juang. *Fundamentals of Speech Recognition*, chapter 4. Prentice Hall, 1993.

[40] L. R. Rabiner. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. In *Proceedings of the IEEE*, volume 77, pages 257–286, February 1989.

[41] Y. Rubner, C. Tomasi, and L. J. Guibas. A Metric for Distributions with Applications to Image Databases. In *Proceedings of the 1998 IEEE International Conference on Computer Vision*, pages 59–66, January 1998.

[42] N. Saito. *Local feature extraction and its application using a library of bases*. PhD thesis, Yale University, 1994.

[43] P. J. Schwartz, E. Vanoli, M. S. Badiale, G. M. D. Ferrari, G. E. Billman, and R. D. Foreman. Autonomic mechanisms and sudden death: new insights from analysis of baroreceptor reflexes in conscious dogs with and without a myocardial infarction. *Circulation*, 78:969–79, 1988.

[44] J. M. Smith, E. A. Clancy, C. R. Valeri, J. N. Ruskin, and R. J. Cohen. Electrical alternans and cardiac electrical instability. *Circulation*, 77:110–121, 1988.

[45] E. L. L. Sonnhammer, G. von Heijne, and A. Krogh. A hidden Markov model for predicting transmembrane helices in protein sequences. In J. Glasgow et al., editors, *Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology*, pages 175–182. AAAI Press, 1998.

[46] P. Sung, Z. Syed, and J. Guttag. Quantifying Morphology Changes in Time Series Data with Skew. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, April 2009.

[47] Z. Syed, J. Guttag, and C. Stultz. Clustering and Symbolic Analysis of Cardiovascular Signals: Discovery and Visualization of Medically Relevant Patterns in Long-Term Data Using Limited Prior Knowledge. *EURASIP Journal on Advances in Signal Processing*, 2007. Article ID 67938.

[48] Z. Syed, B. M. Scirica, S. Mohanavelu, P. Sung, C. P. Cannon, P. H. Stone, C. M. Stultz, and J. V. Guttag. Relation of Death Within 90 Days of Non-ST-Elevation Acute Coronary Syndromes to Variability in Electrocardiographic Morphology. *American Journal of Cardiology*, February 2009.

[49] Z. Syed, B. M. Scirica, D. A. Morrow, S. Mohanavelu, C. M. Stultz, and J. V. Guttag. ECG Markers to Predict Cardiovascular Death: Heart Rate Variability, Deceleration Capacity and Morphologic Variability in Non-ST-Elevation ACS from the MERLIN-TIMI 36 Trial. *Circulation*, 118:S_670, November 2008. Presentation at American Heart Association Scientific Sessions 2008.

[50] Z. Syed, B. M. Scirica, C. M. Stultz, and J. Guttag. Risk-stratification following acute coronary syndromes using a novel electrocardiographic technique to measure variability in morphology. *Computers in Cardiology*, 2008.

[51] Z. Syed, C. M. Stultz, B. M. Scirica, C. P. Cannon, K. Attia, I. O. Stebletsova, S. Mohanavelu, P. H. Stone, and J. V. Guttag. Morphological Variability: A New Electrocardiographic Technique for Risk Stratification After NSTEACS. *Circulation*, 116:II_634, October 2007.

[52] H. Tsuji, F. J. Venditti, E. S. Manders, J. C. Evans, M. G. Larson, C. L. Feldman, and D. Levy. Reduced heart rate variability and mortality risk in an elderly cohort. *Circulation*, 90:878–883, 1994.

[53] S. M. Zaacks, P. R. Liebson, J. E. Calvin, J. E. Parrillo, and L. W. Klein. Unstable angina and non-Q wave myocardial infarction: does the clinical diagnosis have therapeutic implications? *Journal of the American College of Cardiology*, 33(1):107–118, January 1999.