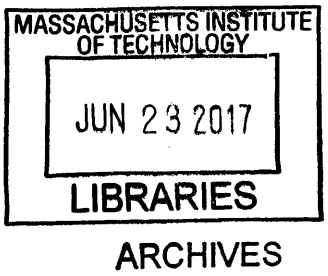
Identifying Patients at High Risk of Death with Novel Computational Biomarkers

by

Paul Daniel Myers

B.S.E. Electrical Engineering
University of Michigan, 2015



Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2017

Author . . Signature redacted
Department of Electrical Engineering and Computer Science
May 12, 2017

Certified by . . Signature redacted
Collin M. Stultz MD, PhD
Professor of Electrical Engineering and Computer Science
Institute for Medical Engineering and Science
Thesis Supervisor

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

Identifying Patients at High Risk of Death with Novel Computational Biomarkers

by

Paul Daniel Myers

Submitted to the Department of Electrical Engineering and Computer Science
on May 12, 2017, in partial fulfillment of the
requirements for the degree of
Master of Science in Electrical Engineering and Computer Science

Abstract

The accurate assessment of a patient's risk of adverse events remains a mainstay of clinical care for patients with cardiovascular disease. Commonly used risk metrics have traditionally been based on simple models that incorporate various aspects of the medical history, presenting signs and symptoms, and lab values. More sophisticated methods, such as those based on signal processing and machine learning, form an attractive platform to build improved risk metrics because they can offer deeper insights into aspects of clinical data that cannot be approached by simpler methods. In particular, generalized additive models can exhibit comparable or superior performance to conventional logistic regression models, while simultaneously providing potentially useful prognostic information on the relationship between patient outcomes and clinical variables. Moreover, artificial neural networks can provide a convenient formalism for combining rich time series information from physiological signals with more conventional patient features. In this work, models based on signal processing and machine learning techniques are developed and applied to two independent datasets consisting of post-acute coronary syndrome patients to predict cardiovascular death at various endpoints. The models are shown to successfully identify high-risk patients, thereby demonstrating the potential of these techniques to improve the management of patients with cardiovascular disease.

Thesis Supervisor: Collin M. Stultz MD, PhD
Title: Professor of Electrical Engineering and Computer Science
Institute for Medical Engineering and Science

Acknowledgments

The author would like to thank his advisor, Professor Collin Stultz, for his advice and support during the course of this work. Most of the results included in this thesis are the direct result of Collin's guidance, and thus, without his help, the completion of this work would not have been possible.

The author would also like to thank Professor John Guttag and Ms. Jen Gong for their useful suggestions, many of which significantly helped shape this work. Thanks are also due to Mr. Jeff Ashe, Dr. Benjamin Scirica, Professor Charlie Sodini, Dr. Maulik Majmudar, and Ms. Maggie Delano for their valuable input.

The author would like to express his gratitude to Dr. Yun Liu, who helped the author begin his work, as well as Harlin Lee for her valuable suggestions and help with implementing several of the algorithms in this thesis. The author also sends his gratitude to the other members of the Computational Biophysics Group for their support, as well as to Dr. Andrew Viterbi and General Electric for funding the work.

Finally, the author would like to thank his family for their love and support.

†Lord Jesus Christ, Son of God, have mercy on me the sinner.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Acute Coronary Syndrome and Patient Risk Stratification

The term acute coronary syndrome (ACS) refers to a spectrum of conditions that are primarily attributable to inadequate blood flow to the myocardium following an acute cholesterol plaque rupture or erosion leading to thrombus formation [1]. In 2010, more than 1.1 million unique hospitalizations were ascribed to ACS and it is estimated that in the United States more than 780,000 people will experience an ACS each year [1], [2].

The effective treatment of patients who have suffered an ACS necessitates an accurate assessment of their risk of subsequent adverse cardiovascular events, thereby permitting the delivery of appropriate therapy. Invasive procedures, for example, are inappropriate for patients whose risk of future adverse events is lower than the risk of the procedure itself. Similarly, patients who are deemed to be very high risk benefit from aggressive, higher risk strategies and closer follow-up [3], [4]. Therefore, the process of categorizing patients according to their risk - a process called risk stratification - is a core part of the management of patients post-ACS.

Early risk-stratification plays an especially important role in the management of patients post-ACS. Patients who are identified as high risk using traditional risk

metrics benefit from invasive therapies (such as percutaneous coronary intervention) within 48 hours of presentation [5], [6]. Risk scores such as the Thrombolysis in Myocardial Infarction (TIMI) and the Global Registry of Acute Coronary Events (GRACE) scores are widely used tools that are calculated using a patient's presenting signs and symptoms, historical data - information that is available at the time of presentation - and the results of laboratory studies that can be obtained within minutes to hours after presentation [7], [8], [9].

1.2 Machine Learning in Cardiovascular Medicine

Due to the heterogeneous nature of ACS, patients often have a wide spectrum of risk for adverse outcomes, thereby requiring the use of a range of clinical variables to accurately guage patient risk [7]. The aforementioned risk scores address this need in part, but it has been noted that such risk scores fail to capture a significant number of deaths in certain patient cohorts [10], [11]. A number of more sophisticated techniques based on signal processing and machine learning have been developed over the past several decades to meet such challenges in various aspects of clinical medicine. While machine learning techniques have been successfully applied in a number of domains of clinical medicine, applications in cardiovascular risk stratification have been somewhat limited [12]. Several studies have attempted to apply machine learning models to a number of clinical problems, including the diagnosis of myocardial infarction [13], risk stratification after myocardial infarction/unstable angina [12], [14] and ACS [15] but have found such techniques to offer few, if any, advantages over more conventional methods. What is lacking in these studies, however, is an attempt to use machine learning techniques to exploit aspects of clinical data that cannot be approached by conventional methods.

This work seeks to demonstrate the efficacy of applying machine learning methods to the risk stratification of post-ACS patients. In Chapter 2, generalized additive models trained on a rich set of features from two post-ACS patient populations are compared with conventional logistic regression models, and shown to offer com-

parable or improved performance and improved interpretability. In Chapter 3, an artificial neural network that incorporates features derived from the ST-segment of the electrocardiogram (ECG) is developed and shown to offer significantly improved performance over conventional logistic regression techniques. Finally, this document concludes with a discussion of possible future directions for the work.

Chapter 2

Risk Stratification Using Linear Models of Electronic Medical Record Data

2.1 Introduction

Prediction problems for clinical outcomes are often formulated as follows: Given a set of relevant clinical variables for a collection of patients who have presented with certain symptoms, assign a score to each patient that indicates how likely that patient is to suffer from an adverse event within a certain timeframe. If the occurrence of an adverse outcome may be represented as a binary variable, a natural model to use in this instance is the logistic model, which looks as follows:

$$y_i = \frac{1}{1 + e^{-\boldsymbol{x_i}^T \boldsymbol{\beta}}} \tag{2.1}$$

Here, y_i is the outcome for patient i, $\boldsymbol{x_i} = (1, x_{i1}, ..., x_{ip})^T$ is a column vector of observations of clinical variables for patient i, and $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_p)^T$ is a column vector of weights corresponding to each observation. In order to estimate y_i, it is necessary to determine the weights, which is difficult to do given the non-linear relationship between y_i and the weights in Eq. (2.1). A simpler approach is to instead

rewrite Eq. (2.1) so that the right-hand-side of the equation is linear in the weights, as follows:

$$ln(\frac{y_i}{1-y_i}) = \boldsymbol{x_i}^T\boldsymbol{\beta} \tag{2.2}$$

The form of Eq. (2.2) is analogous to that of a simple linear model, except that the term on the left-hand-side of the equation is a non-linear function of the response variable, rather than the variable itself; this model is called a generalized linear model (GLM). The weights may be estimated using a procedure similar to that used in an ordinary linear model, and y_i may then be estimated using Eq. (2.1). The logistic regression model of Eq. (2.1), which is a particularl class of GLMs, is commonly applied in clinical prediction problems.

2.2 Generalized Additive Models

While Eq. (2.2) is relatively simple to interpret, the assumption that the right-hand-side of the equation is linear in $\boldsymbol{x_i}$ is rather restrictive. An extension of the GLM is the generalized additive model (GAM), which allows the functional dependence on $\boldsymbol{x_i}$ to be determined by the data, as follows:

$$ln(\frac{y_i}{1-y_i}) = \sum_{j=1}^{p} f_j(x_{ij}) \tag{2.3}$$

Here, $\boldsymbol{f} = (f_1, f_2, ..., f_p)^T$ is a set of smooth functions of a single variable.

Deriving the set of smooth functions proceeds in the following manner [16]. Let $\boldsymbol{y} = (y_1, ..., y_n)^T$ be a response variable of a single predictor $\boldsymbol{x} = (x_1, ..., x_n)^T$; each instance of $\boldsymbol{y}$ is then related to each instance of $\boldsymbol{x}$ as follows:

$$y_i = f(x_i) \tag{2.4}$$

In order to estimate the function f, the following optimization problem may be solved:

$$\hat{f} = \underset{f \in C^2}{\arg\min} \sum_{i=1}^{n} (y_i - f(x_i))^2 + \lambda \int_a^b (f''(t))^2 dt \tag{2.5}$$

Here, λ is a regularization parameter, the function of which will be explained shortly, a and b are defined such that a is less than or equal to the minimum value of $\boldsymbol{x}$ and b is greater than or equal to the maximum value of $\boldsymbol{x}$, and C^2 is the space of all functions that have continuous second derivatives. The form of this optimization problem may be motivated by noting the role of each of the two summands. The first summand measures how closely the function f fits the data and is analogous to the mean-squared error used to fit linear regression models. Fitting the function f using the first summand alone could result in a function that fits the data nearly perfectly, but fails to generalize to unseen data. The role of the second summand is therefore to act as a regularization term. The second derivative of a function f measures the curvature of that function, so the regularization term penalizes the degree of curvature of the function. This penalty is consistent with the purpose of smoothing, which is to generate a function that follows the trend of the data, but sufficiently deviates from the data to allow the model to generalize to unseen data. The variable λ controls how much weight is given to the penalty term, with $\lambda = 0$ resulting in no regularization and $\lambda \to \infty$ resulting in a straight-line fit to the data, where no curvature is allowed. The second derivative is squared to ensure that the integrand is always positive. It may be shown that the solution to the optimization problem in Eq. (2.5) is a natural cubic spline with knots at each unique value of $\boldsymbol{x}$, as follows:

$$f(x_i) = \sum_{j=1}^{n} h_j(x_i)\theta_j \tag{2.6}$$

Here, $\boldsymbol{h} = (h_1, ..., h_n)^T$ is a set of basis functions and $\boldsymbol{\theta} = (\theta_1, ..., \theta_n)^T$ is a set of weights corresponding to each basis function. With this result, the regularization term in Eq. (2.5) may be rewritten as follows:

$$\begin{aligned}\lambda \int_a^b (f''(t))^2 dt &= \lambda \int_a^b (\boldsymbol{h}''^T \boldsymbol{\theta})^2 dt \\ &= \lambda \int_a^b (\boldsymbol{\theta}^T \boldsymbol{h}'')(\boldsymbol{h}''^T \boldsymbol{\theta}) dt \\ &= \lambda \boldsymbol{\theta}^T \int_a^b \boldsymbol{h}'' \boldsymbol{h}''^T dt \boldsymbol{\theta} \\ &= \lambda \boldsymbol{\theta}^T \boldsymbol{\Omega} \boldsymbol{\theta}\end{aligned} \tag{2.7}$$

The matrix $\boldsymbol{\Omega}$ is defined such that $\Omega_{jk} = \int_a^b h_j(t) h_k(t) dt$ for $j = 1, ..., n$ and $k = 1, ..., n$. Using the result from Eq. (2.7), the parameters $\boldsymbol{\theta}$ may be estimated from Eq. (2.5):

$$\hat{\boldsymbol{\theta}} = \underset{f \in C^2}{\arg\min} \; \|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{\theta}\|_2^2 + \lambda \boldsymbol{\theta}^T \boldsymbol{\Omega} \boldsymbol{\theta} \tag{2.8}$$

Here, $\boldsymbol{H}$ is a $n \times n$ matrix where $H_{ij} = h_j(x_i)$ for $i = 1, ..., n$ and $j = 1, ..., n$. Eq. (2.8) is a generalized ridge regression problem, which may be solved to yield the parameters $\boldsymbol{\theta}$:

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{H}^T \boldsymbol{H} + \lambda \boldsymbol{\Omega})^{-1} \boldsymbol{H}^T \boldsymbol{y} \tag{2.9}$$

The function f may then be calculated as:

$$\begin{aligned}\hat{f} &= \boldsymbol{H}\hat{\boldsymbol{\theta}} \\ &= \boldsymbol{H}(\boldsymbol{H}^T \boldsymbol{H} + \lambda \boldsymbol{\Omega})^{-1} \boldsymbol{H}^T \boldsymbol{y} \\ &= \boldsymbol{S}\boldsymbol{y}\end{aligned} \tag{2.10}$$

Here, the definition $\boldsymbol{S} = \boldsymbol{H}(\boldsymbol{H}^T \boldsymbol{H} + \lambda \boldsymbol{\Omega})^{-1} \boldsymbol{H}^T$ has been used; the matrix $\boldsymbol{S}$ is often called the smoother matrix. The set of basis functions $\boldsymbol{h}$ may be chosen to satisfy certain requirements; for computational efficiency, the B-spline basis is often used. To gain some insight into the functioning of the smoother matrix, however, it is instructive to choose a different basis. By applying a number of algebraic manipulations, $\boldsymbol{S}$ may be rewritten as:

$$\boldsymbol{S} = (\boldsymbol{I} + \lambda \boldsymbol{K})^{-1} \tag{2.11}$$

Here, the definition $\boldsymbol{K} = (\boldsymbol{H}^T)^{-1}\boldsymbol{\Omega}\boldsymbol{H}^{-1}$ has been used. Eq. (2.11) is said to be in Reinsch form. The eigendecomposition of $\boldsymbol{K}$, which is given as $\boldsymbol{K} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^T$, may be computed to yield:

$$\begin{aligned} \boldsymbol{S} &= \boldsymbol{U}(\boldsymbol{I} + \lambda \boldsymbol{D})^{-1}\boldsymbol{U}^T \\ &= \sum_{j=1}^{n} \frac{1}{1 + \lambda d_j} \boldsymbol{u_j} \boldsymbol{u_j}^T \end{aligned} \tag{2.12}$$

Here, the eigenvector $\boldsymbol{u_j}$ is the jth column of the matrix $\boldsymbol{U}$ and the eigenvalue d_j is the jth diagonal element of the matrix $\boldsymbol{D} = diag(d_1, d_2, ..., d_n)$. The columns of the matrix $\boldsymbol{U}$, form a basis called the Demmler-Reinsch basis that has the property, among others, that all of the basis functions are orthogonal to each other [17]. For the Demmler-Reinsch basis, $0 = d_1 = d_2 < d_3 \leq ... \leq d_n$, so the smoother matrix has the property that the first two basis functions are not penalized, while all other basis functions are penalized in in proportion to the magnitude of their eigenvalues. As an aside, it is conventional to define the number of degrees of freedom of the model to be the trace of the smoother matrix, as follows:

$$df = tr(\boldsymbol{S}) = \sum_{j=1}^{n} \frac{1}{1 + \lambda d_j} \tag{2.13}$$

The number of degrees of freedom is therefore determined by the value of λ.

Comparing Eq. (2.10) with Eq. (2.12), it may be seen that $\boldsymbol{H}$ may be replaced with $\boldsymbol{U}$ and $\boldsymbol{\Omega}$ may be replaced with $\boldsymbol{D}$; it therefore follows that the optimization problem in Eq. (2.8) may be rewritten as:

$$\hat{\boldsymbol{\theta}} = \underset{f \in C^2}{\arg\min} \ \|\boldsymbol{y} - \boldsymbol{U}\boldsymbol{\theta}\|_2^2 + \lambda \boldsymbol{\theta}^T \boldsymbol{D} \boldsymbol{\theta} \tag{2.14}$$

Returning to Eq. (2.7), the requirement that $d_1 = d_2 = 0$ implies that $u_1'' = u_2'' = 0$. If it is supposed that the basis functions $\boldsymbol{u}$ are arranged in order of increasing

functional complexity, it then follows that u_1 corresponds to a constant term and u_2 corresponds to a linear term. Setting $\theta_1 = \alpha_0$, $\theta_2 = \alpha$, and $\theta_j = \beta_j$ for $j = 3, ..., n$ allows Eq. (2.14) to be rewritten as:

$$(\hat{\alpha_0}, \hat{\alpha}, \hat{\boldsymbol{\beta}}) = \underset{\alpha_0 \in \mathbb{R}, \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^{n-2}}{\arg\min} \|\boldsymbol{y} - \alpha_0 - \alpha \boldsymbol{x} - \boldsymbol{U}_{3:n}\boldsymbol{\beta}\|_2^2 + \lambda \boldsymbol{\beta}^T \boldsymbol{D}_{3:n} \boldsymbol{\beta} \tag{2.15}$$

Here, $\boldsymbol{U}_{3:n}$ and $\boldsymbol{D}_{3:n}$ correspond to columns 3 through n of $\boldsymbol{U}$ and $\boldsymbol{D}$, respectively. From Eq. (2.15), it may be seen that the smoothing spline fits both linear and non-linear terms, with the non-linear terms being penalized according to increasing functional complexity.

The foregoing analysis may be generalized to the case when there is more than one predictor variable by modifying Eq. (2.5) as follows:

$$\underset{\{f_j \in C^2\}_1^p}{\text{minimize}} \sum_{i=1}^{n} (y_i - \sum_{j=1}^{p} f_j(x_{ij}))^2 + \lambda \sum_{j=1}^{p} \int_{a_j}^{b_j} (f_j''(t))^2 dt \tag{2.16}$$

Thus, there corresponds to each predictor variable $\boldsymbol{x_j} = (x_{j1}, x_{j2}, ..., x_{jn})^T$ a function f_j and a pair of limits of integration (a_j, b_j). The approach outlined above may be extended to yield the following optimization problem:

$$\underset{\alpha_0 \in \mathbb{R}, \{\alpha_j \in \mathbb{R}, \boldsymbol{\beta_j} \in \mathbb{R}^{n-2}\}_1^p}{\text{minimize}} \left\| \boldsymbol{y} - \alpha_0 - \sum_{j=1}^{p} (\alpha_j \boldsymbol{x_j} + \boldsymbol{U_j}\boldsymbol{\beta_j}) \right\|_2^2 + \lambda \sum_{j=1}^{p} \boldsymbol{\beta_j}^T \boldsymbol{D_j} \boldsymbol{\beta_j} \tag{2.17}$$

Each predictor variable $\boldsymbol{x_j}$ has its own set of α_j, $\boldsymbol{\beta_j}$, $\boldsymbol{U_j}$, $\boldsymbol{D_j}$ values, analogous to the single-predictor case, with the exception of the constant terms, which have been combined into a single α_0 term.

Solving the optimization problem in Eq. (2.17) allows the smoothing functions to be determined, which may then be visualized to obtain insight into the dependence of the clinical outcome on each variable [18]. Thus, in addition to generating models that derive their form from the data itself, the formalism described above offers additional intelligibility that logistic regression models do not.

2.3 Datasets

The Cohort-1 dataset, which was used in previous works [10], [19], [20] to evalaulte the performance of several computational biomarkers, contains records for 6354 patients, approximately half of which received a study drug and approximately half of which did not; the set of patients that received the study drug is referred to as the treatment arm, and the set that did not receive the drug is referred to as the control arm. A set of 21 features, categorized into demographics, health history, laboratory tests, and medications, was collected for each patient, though some features were empty for some patients. ECG waveforms were recorded for approximately two-thirds of the patients, with the duration of each recording lasting from 24 hours to 168 hours. Each patient was followed for approximately one year after the trial ended, and the following outcomes were noted: Cardiovascuar death (CVD), if it occurred, and the date of death, if it occurred. In addition to the features provided by the clinical trial, the morphologic variability [19], [20] was calculated for each patient from the first 24 hours of ECG recordings and included as an additional feature.

The Cohort-2 dataset [19] consisted of 990 patients with a smaller set of features than in Cohort-1. The features that are in common between the two datasets are age, gender, current smoker, history of hypertension, history of diabetes, previous MI, and previous angiography. All patients in Cohort-2 received a study drug, which was different from that used in Cohort-1.

2.4 Bootstrapping Results on the Cohort-1 Control Arm

Four models were developed on these two datasets: A logistic regression model with L1 regularization, a logistic regression model with L2 regularization, a GAM with no regularization, and a GAM with stepwise feature selection as a form of regularization (GFS) [21], [22]. Comparison of these models was facilitated by a number of experiments. In the first experiment, the control arm was used for both training and

testing; patients in the study arm were not considered. The set of the following 20 features was used:

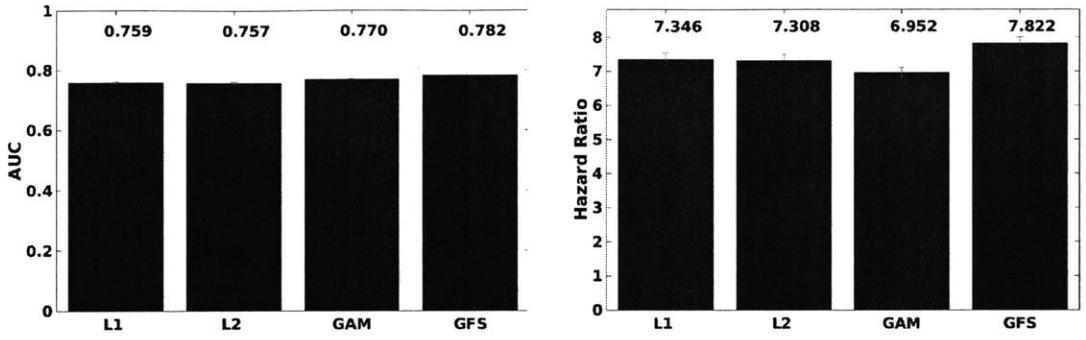
- Binary (16): Gender, history of diabetes, history of hypertension, current smoker, prior MI, prior PCI or CABG, history of CHF, history of ventricular arrhythmia, history of resuscitated sudden death, creatinine clearance > 60, index event, ST depression $\geq$ 1mm, prior angiography, aspirin administration in the hospital, beta-blocker administration in the hospital, and ACE-I/ARB administration in the hospital.

- Categorical (1): TIMI risk score group (TRG). The TRG is derived from the TIMI risk score (TRS) [7] as follows: TRS = 1 or 2 $\rightarrow$ TRG = 1; TRS = 3 or 4 $\rightarrow$ TRG = 2; TRS = 5 to 7 $\rightarrow$ TRG = 3.

- Continuous (3): Age, body mass index (BMI), and MV.

The three continuous-valued features were normalized within the range [0,1] by subtracting the minimum and dividing by the range of each variable. Patients who were missing any of the above 20 features were not considered in the analysis; under this constraint, the patient population consisted of 2130 patients with 79 CVDs (3.71% death rate).

The dataset was randomly partitioned 1000 times into training sets of 80% of the data and testing sets of 20% of the data in order to test the statistical robustness of the classifiers, as well as to generate confidence intervals for the data; this process is referred to as bootstrapping. The training and testing sets were required to contain the same percentage of CVDs. For each of the 1000 partitions, three-fold cross-validation was used to find the regularization parameters for the logistic regression models. These models were then trained on the entire training set to find the weights before being evaluated on the test set. For the GAM without feature selection, the smoothing functions were derived from the training sets before being evalulated on the testing sets. For the GFS model, features were selected in a stepwise fashion on the training sets, then evaluated on the testing sets.



(a) AUC

(b) Hazard Ratio

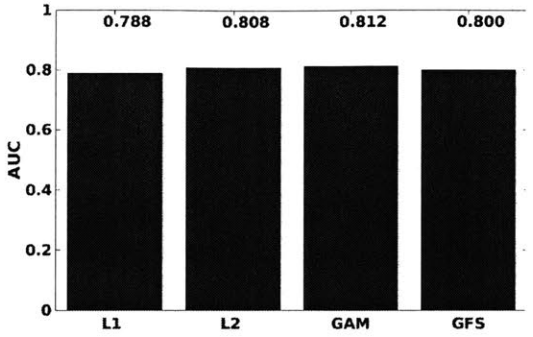
Figure 2-1: Performance metrics on the test set averaged over 1000 rounds of training and testing. The error bars measure one standard error and the numbers above the bars indicate the mean value for each model.

The results of the experiment are shown in Fig. 2-1. All four models show comparable performance when measured by the AUC, with the GAM with feature selection offering the highest AUC, as well as the highest hazard ratio.

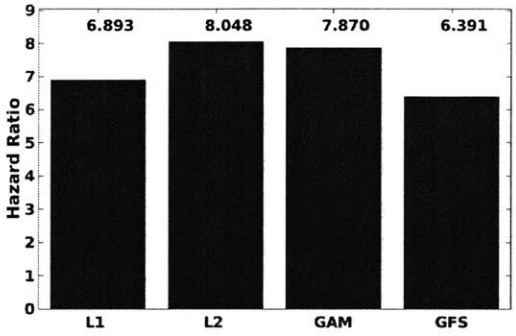
2.5 Results on Cohort-1 Treatment Arm

In the second experiment, the Cohort-1 treatment arm was used as the testing set, while the control arm was used as the training set. No bootstrapping was applied, so the results presented are for a single run of each model. The same set of features denoted above was used. The training set consisted of 2130 patients with 79 CVDs (3.71% deat rate), while the testing set comprised 2116 patients with 70 CVDs (3.31% death rate). For the training set, ten-fold cross-validation was used to find the regularization parameters for the logistic regression models. These models were then trained on the entire training set to find the weights before being evaluated on the test set.

The results of the experiment are shown in Fig. 2-2. All four models again show comparable performance when measured by the AUC, with the GAM offering the best performance; the GAM also displays the highest hazard ratio. The smoothing functions generated by the GAM without feature selection on the Cohort-1 treatment



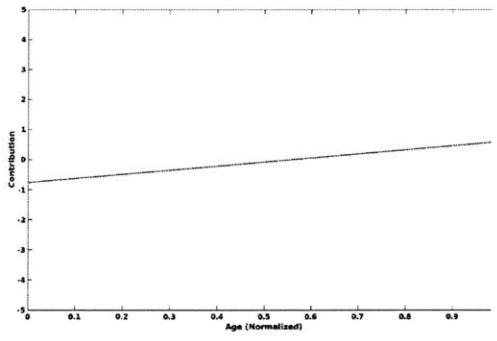
(a) AUC



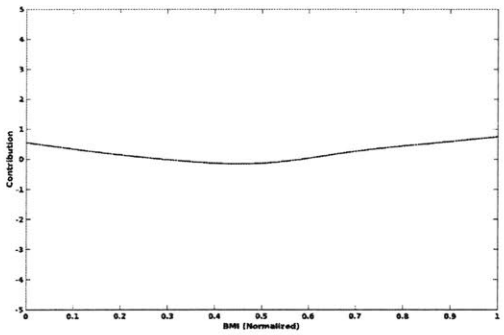
(b) Hazard Ratio

Figure 2-2: Performance metrics obtained from training the models on the Cohort-1 control arm and testing them on the Cohort-1 treatment arm.

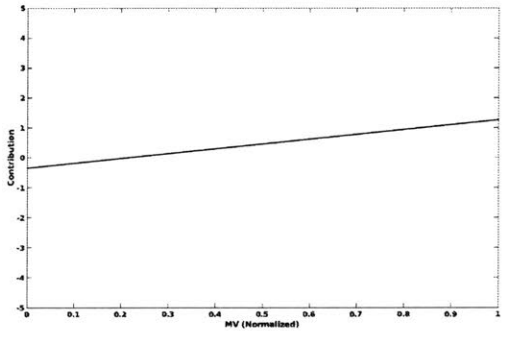
arm are shown in Fig. 2-3.



(a) Age



(b) BMI



(c) MV

Figure 2-3: Smoothing functions obtained from the Cohort-1 treatment arm.

2.6 Results on Cohort-2

The four models were then applied to an independent dataset, Cohort-2, to evaluate their generalizability. Because Cohort-2 contained fewer features than did Cohort-1, a reduced feature set was used for both training and testing; the following 7 features were used:

- Binary (6): Gender, current smoker, history of hypertension, history of diabetes, prior MI, and prior angiography.
- Continuous (1): Age.

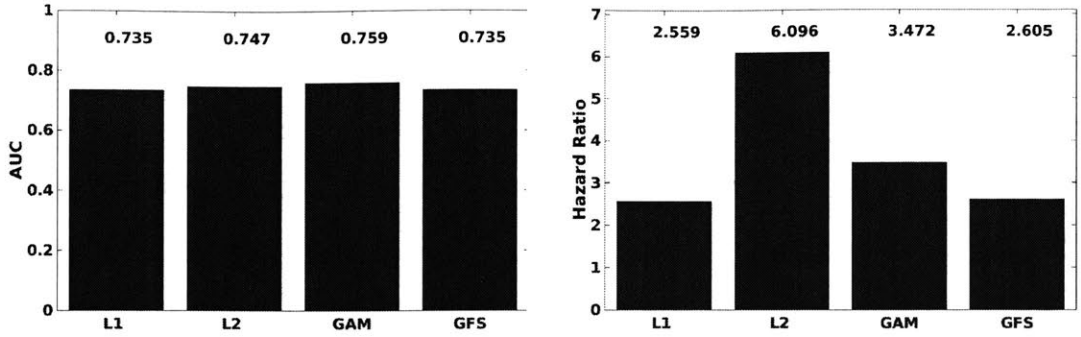
The age feature was normalized as described above.

The training set consisted of patients from the Cohort-1 treatment arm who had all of the above features; under this requirement, the training set contained 2259 patients with 89 CVDs (3.93% death rate). The testing set consisted of patients from Cohort-2 who had all of the above features; under this requirement, the testing set consisted of 909 patients with 15 CVDs (1.65% death rate). No bootstrapping was applied, so results are shown for a single run of each model. For the training set, ten-fold cross-validation was used to find the regularization parameters for the logistic regression models. The models were then trained on the entire training set to find the weights before being evaluated on the test set.

The performance metrics for each model are shown in Fig. 2-4. The GAM exhibits the highest AUC, while the L2 models offers the highest hazard ratio. The smoothing function obtained for the age feature from the GAM without feature selection is shown in Fig. 2-5.

2.7 Discussion

The foregoing results illustrate the strong association the proposed models have with one-year CVD. Moreover, the performance of the models on Cohort-2 demonstrate that they generalize well to unseen data.



(a) AUC (b) Hazard Ratio

Figure 2-4: Performance metrics obtained from training the models on the Cohort-1 control arm and testing them on Cohort-2.

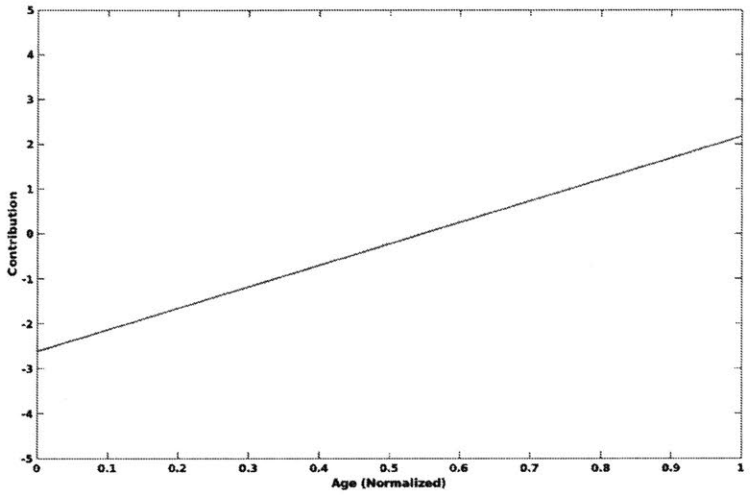


Figure 2-5: Smoothing function for the age feature obtained from the GAM without feature selection.

Although the four models perform comparably, the GAM and GFS models offer slight performance improvements over the conventional logistic regression models. One of the primary reasons, however, for using a GAM is to gain insight into the functional dependence of the outcome on individual features by studying the smoothing functions generated by the model. Although no definitive conclusions may be drawn from Fig. 2-3 and Fig. 2-5, the form of each function may be compared with intuition to provide a check on the output of the model. For each plot, the value of the feature is shown on the x-axis, while the contribution the feature makes to the

additive model, which is analogous to the product of the weight and the feature in a logistic regression model, is shown on the y-axis. Positive contribution values indicate that the feature increases the risk of an adverse outcome, while negative values indicate that the feature reduces the risk. For the age feature, the functional form appears to be approximately linear in both cohorts, and is monotonically increasing. This result is consistent with expectation, since the risk of an adverse outcome might be expected to increase approximately linearly with age. The BMI feature shows a slight downward bend at intermediate values of the BMI, which is reasonable given that very low and very high values of the BMI are often associated with poor health. The form of the MV feature is more difficult to assess, though it seems reasonable that it is monotonically increasing. High values of the MV suggest that there is a great deal of variability between heart beats, which may indicate heart trouble. Thus, the GAM offers an intuitive clinical interpretation of the data that is lacking in the logistic regression models, while maintaining or improving the performance of these conventional models.

Chapter 3

Computational Biomarkers Based on ST-Segment Morphology for Patient Risk Stratification

3.1 Introduction

In Chapter 2, it was shown that linear models based on features derived from the patient record can offer strong predictive performance for CVD given an appropriate set of features. However, not all clinical datasets have such rich sets of features. Moreover, the foregoing models offer limited incorporation of important prognostic information contained in time-varying physiological signals. A commonly used clinical tool for diagnosing cardiac dysfunction is the ECG, a graphical representation of the electrical activity of the heart. ECG signals are measured by placing electrodes at strategic locations on the body, primarily the chest, arms, and legs, and recording the electric potential difference between pairs of electrodes. A typical ECG signal consists of the following waveforms: P, Q, R, S, and T. Each waveform is associated with a different part of the cardiac cycle, with the P-wave corresponding to atrial contraction, the Q, R, and S-waves corresponding to ventricular contraction, and the T-wave corresponding to ventricular relaxation. Disorders resulting in reduced blood

flow to the heart, a condition called ischemia that is associated with ACS, are often manifested in the ST-segment, which is the isoelectric portion of the ECG signal between the S and T-waves corresponding to ventricular relaxation.

Patients who present with ACS are typically divided into two categories - those with ST segment elevation on their presenting electrocardiogram (STE-ACS) and those with non-ST segment elevation (NSTE-ACS). Risk assessment in patients with NSTE-ACS can be problematic as while their short-term in-hospital mortality is lower relative to patients who present with STE-ACS, they have similar long-term mortality rates [23], [24], [25], [26]. Traditional metrics that assess the risk of future adverse cardiovascular events after NSTE-ACS use a range of clinical variables to estimate the patient risk [7], [9]. Nevertheless, an accurate assessment of patient risk remains a difficult task. Many risk scores, for example, fail to capture a significant number of deaths in certain patient cohorts [10], [11].

Given the importance of the ST segment in identifying high-risk patient subgroups, the presence and extent of ST-segment deviation are often components in patient risk models [7], [25], [27]. However, while physicians are well trained to assess the extent of ST segment deviation by eye, subtle ST segment deviations that cannot be detected by eye alone may also carry important prognostic information. These considerations motivate the development of more sophisticated analysis techniques that exploit the detection of subtle ST segment changes that may be unrecognized by visual inspection alone.

Herein, we present a method for combining features from the patient record and the ECG to predict CVD within one year after a NSTE-ACS. A key feature of the method is that it not only utilizes traditional patient demographic and medical history information, but it also takes advantage of the continuous ECG waveform that is typically available from standard Holter monitoring. Unlike other risk assessment tools that only use a single feature of the surface ECG on presentation, such as the level of ST-segment depression, our method applies an Artificial Neural Networks (ANN) to the precise time series of ST segment changes after admission with a NSTE-ACS. The prognostic ability of the final model is evaluated on two independent datasets

composed of post-ACS patients. Using an ANN that is trained on a small amount of Holter data, we arrive at a model that outperforms traditional risk models and other risk metrics that use considerably more data.

3.2 Datasets

This work relied on the same two datasets used Chapter 2, but with different patient selection criteria. The first study included 4557 patients with interpretable continuous ECG data [20]. Continuous monitoring was done for up to seven days following randomization with a three-lead digital Holter monitor at a sampling rate of 128 Hz. Patients with fewer than 50 usable beats in the first day of ECG data or who were missing any of the seven patient features were not considered in the analysis; using these criteria, 4395 of the 4557 patients remained.

The second study included interpretable ECG data from 909 patients [19]. Continuous ECG monitoring was conducted during the first four to seven days after randomization. As with the Cohort-1 dataset, patients with fewer than 50 clean beats in the first day of ECG data or who were missing any of the seven patient features were not considered in the analysis; using these criteria, 861 of the 909 patients remained.

Continuous-valued patient features in both cohorts were normalized to fall within [0,1] by subtracting the minimum and dividing by the range of each variable.

3.3 Feature Extraction

Recently, a method for mathematically representing the morphology of the ST segment was proposed to distinguish between ischemic and non-ischemic ST segment events in an ECG tracing [28]. The use of orthogonal basis functions to represent the morphology of the ST-segment is based on the observation that the functional forms of appropriately chosen basis functions approximate the main categories of ST-segment morphology; these categories are the level with respect to the isoelectric line, slope, and curvature of the ST-segment. Mathematically, the first category corresponds

to a constant function, the second to a linear function, and the third to non-linear functions. The set of Legendre polynomials is known to possess such functional forms. The Legendre polynomials may be generated by applying the Gram-Schmidt orthonormalization procedure to the following matrix [29]:

$$\Omega_{ij} = (2\frac{i-1}{M-1} - 1)^{j-1},\ i = 1, 2, ..., M,\ j = 1, 2, ..., M \tag{3.1}$$

Here, M is the number of sample points in the range [-1,1], and is chosen based on the sampling rate of the ECG signal. In [28], 32 samples were used for a sampling rate of 250 Hz. In the present work, the sampling rate was 128 Hz, necessitating the use of a smaller number of samples; 16 samples were used. Applying the Gram-Schmidt orthonormalization procedure to $\mathbf{\Omega}$ generates a new matrix $\mathbf{\Phi_L}$, which may be used to transform a vector of ST-segment samples into its constituent Legendre polynomial basis functions. To do so, the ECG signal for a given patient is preprocessed for noise removal following the procedure outlined in [19]. Baseline wander is removed by applying a median filter to the ECG signal, and wavelet filtering with a soft threshold is used to remove high-frequency noise. The signal is then normalized by the mean R-wave amplitude and the Physionet Signal Quality Index [30] is used to remove noisy signal segments. Segmentation of each beat of the signal into its constituent components (P-wave, QRS-complex, ST-segment, T-wave) is done using a previously published method based on wavelet transforms [31]. The ST-segment of each beat is taken to be all sample points falling between the S-wave and T-wave labels generated by the segmentation algorithm; beats missing either an S-wave or T-wave label or both are rejected as being too noisy to use. The length of the ST-segment can vary widely between beats depending on the variation of the heart rate and condition of the patient over time; segment lengths were found to range between fewer than 16 to more than 32. Segments with fewer than 16 and more than 32 samples are rejected as either being too noisy to use or improperly labeled by the segmentation algorithm. For segments with lengths between 16 and 32 samples, the midpoint of the segment is found and eight samples are taken on either side of it. Doing so helps to ensure that

the samples used are part of the true ST-segment, even in the presence of small errors in the placement of the S-wave and T-wave labels. The set of 16 samples is stored in a vector called $\boldsymbol{x}$, which is used to generate the Legendre polynomial representation of the segment as follows:

$$s_j = \frac{\boldsymbol{\Phi}_L^T \boldsymbol{x}_j}{\sigma(\boldsymbol{\Phi}_L^T \boldsymbol{x}_j)} \tag{3.2}$$

Here, s_j is the Legendre polynomial expansion for beat j, x_j is the sample vector for beat j, $\boldsymbol{\Phi_L}$ is as defined above, and σ denotes the standard deviation; T indicates the matrix transpose operation. The length of s_j is equal to the number of ST segment samples taken, which is 16 in the present work; thus, 16 Legendre polynomial coefficients are generated. Although 16 coefficients are derived for each ST-segment, only the first two are used in subsequent analyses; this choice was made primarily for two reasons. The principal reason is that the first two coefficients, which represent the level and slope of the ST-segment, are of primary clinical significance when diagnosing conditions that affect the morphology of the ST-segment. The second reason is that the higher-order coefficients were found to capture the noise inherent in the signal due to the low sampling rate. It should further be noted that although all patients used in the training and testing of the network possessed at least one day of continuous ECG data, the number of clean beats varied markedly across patients, necessitating the use of a fixed number of beats. As will be discussed below, the coefficient time series is given as input to a recurrent neural network (RNN) for processing. Due to limits in the recurrent memory of the RNN, a window of the first 50 beats was used for each patient. The feature extraction procedure is shown schematically in Fig. 3-1.

3.4 Logistic Regression Models with ST-Segment Based Features

To test the discriminatory value of this approach we incorporated features derived from this representation of the ST-segment into logistic regression (LR) models de-

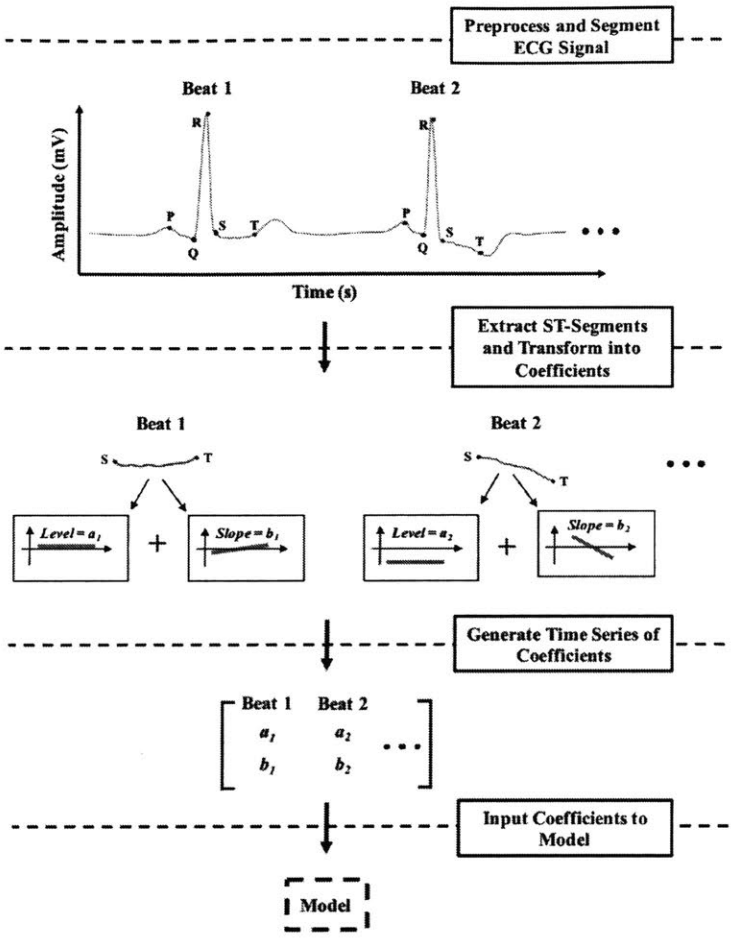


Figure 3-1: ST segment feature extraction process. The ECG signal is first preprocessed to remove noise and identify clean beats, then segmented to delineate the various waveforms comprising each beat. The ST segments are then isolated and a mathematical transformation is applied to extract the coefficients describing the morphology of the segment, which are then applied as input to the model.

signed to predict CVD one year after a NSTE-ACS. Three models were tested: LR_{Hx}, LR_{ST}, and LR_{Hx+ST}.

The LR_{Hx} model only uses information from the patient history and demographic data. The seven features used were age, gender, whether the patient was a smoker at the time of enrollment, history of hypertension, history of diabetes, previous myocardial infarction (MI), and previous angiography; i.e., features that were in common to both patient cohorts. All features, except age, were binary-valued; the age feature was continuous-valued.

The LR_{ST} model uses four features corresponding to the mean and standard deviation of the first two coefficients representing the level and slope of the ST segments of the first 50 beats from the Holter - a time scale corresponding to less than one minute in this patient population. Parenthetically, we note that since our method requires beats where the ST segment can be clearly identified in an automated manner, we often require more than one minute of data to obtain 50 usable beats. In general, approximately 45 minutes of data are required on average to ensure that 50 usable beats are obtained.

The LR_{Hx+ST} logistic regression model uses the seven features in the LR_{Hx} model plus the four ST-segment based features in the LR_{ST} model.

All three models were applied to the Cohort-1 dataset using a bootstrapping procedure in which the dataset was randomly partitioned into a training set of 80% of the data and a testing set of 20% of the data. This random partitioning was repeated 1000 times. For each partition, the three models were trained on the training sets (80% of the data) and then applied to the testing set (20% of the data). Results are reported as averages over the 1000 testing sets.

All three models were associated with CVD post-ACS, with the LR_{Hx+ST} offering the highest AUC Table 3.1. The small improvement in the AUC of the LR_{ST} model over the LR_{Hx} model was statistically significant with $p < 0.05$, and the improvement in the AUC of the LR_{Hx+ST} over the LR_{Hx} and LR_{ST} models was also significant, with $p < 0.001$ in both cases.

For comparison, we evaluated the performance of three other logistic regression models that incorporate ECG information in various forms. In a previous work, we evaluated the performance of several ECG-based metrics for risk stratification post-NSTE-ACS using the Cohort-1 dataset [20]. Morphologic Variability (MV), Heart Rate Variability (HRV), and Deceleration Capacity (DC) are three ECG-based metrics that have been associated with adverse outcomes in patients with cardiovascular disease [19], [20], [32], [32], [33]. We therefore constructed three additional logistic regression models using these metrics. Each model combined all seven patient features with one of these metrics: i.e., LR_{Hx+MV}, LR_{Hx+HRV}, and LR_{Hx+DC}. It is important

to note that each of these ECG metrics use at least 24 hours' worth of ECG data while the LR_{ST} and LR_{Hx+ST} models only use 50 beats from the Holter monitor.

The LR_{Hx+MV} model offered the highest AUC of all the comparison models tested Table 3.1, followed by the LR_{Hx+HRV}, LR_{Hx+DC} models; however, all three AUCs are below that of the LR_{Hx+ST} model ($p < 0.001$). Lastly, the AUC associated with the TIMI Risk Score for NSTE-ACS (TRS) [7], which includes information about the presence of 1 mm ST segment depression on presentation, falls below all of the logistic regression models Table 3.1.

Model	AUC
LR_{Hx}	0.695
LR_{ST}	0.701*
LR_{Hx+ST}	0.734**
LR_{Hx+MV}	0.727
LR_{Hx+HRV}	0.720
LR_{Hx+DC}	0.705
TRS	0.670
RNN	0.689
LR_{Hx}-RNN	0.743***

Table 3.1: AUCs of different models (and the TIMI NSTE-ACS risk score) using the Cohort-1 dataset. *Improvement over LR_{Hx} significant with $p < 0.05$. **Improvement over LR_{Hx}, LR_{ST}, LR_{Hx+MV}, LR_{Hx+HRV}, and LR_{Hx+DC} significant with $p < 0.001$. ***Improvement over all other models significant with $p < 0.001$.

3.5 Recurrent Neural Networks

These results suggest that the ST-segment based features derived from an automated analysis of the ST-segment have discriminative power. However, since the logistic regression models do not exploit the fact that we obtain a time-series of ST-segments, we developed a neural network model that could effectively use these data. In particular, recurrent neural networks (RNNs) provide a natural formalism for modeling the time series data that we exploit in this work [36].

ANNs consist of a set of interconnected processing units that are loosely inspired by the supposed organization of the human brain [34]. An example of a processing

unit, which is referred to as an artificial neuron in analogy with biological neurons, is shown in Fig. 3-2 [35]. The neuron receives inputs from an arbitrary number of other neurons, and each input is multiplied by a weight, which represents the synapse between two neurons. The weighted inputs are then summed in a process called integration, and the resulting sum is given as input to a nonlinear function called an activation function, the output of which may then be sent to other neurons or used in other computations. The choice of activation function depends on the purpose for which the neuron is being used. Common choices include the hyperbolic tangent and sigmoid functions, since these functions are bounded and therefore help prevent computations from becoming unbounded in large networks. There are a number of different types of ANNs. A feedforward neural network (FNN), which is shown in Fig. 3-3, is constructed by arranging a number of these neurons into cascaded layers; the name feedforward indicates that information flows from the input of the network to the output, without any of the intermediate computations being stored for future use. Each layer of the network consists of an arbitrary number of neurons that receive inputs from the neurons in the preceding layer, if there is one. Each neuron may receive inputs from all or a subset of the neurons in the preceding layer; the term fully-connected is applied to layers in which the neurons receive inputs from all of the neurons in the preceding layer. The first layer of the network is called the input layer, and contains as many neurons as there are input variables. The final layer of the network is called the output layer, and contains as many neurons as there are output variables. All intermediate layers, if there are any, are called hidden layers and may contain any number of neurons. Increasing the number of hidden layers generally increases the ability of the network to identify complex patterns in the data, but also increases the risk of overfitting.

Recurrent neural networks (RNNs) are similar to FNNs in that they consist of cascaded layers of interconnected neurons, but unlike FNNs, RNNs store intermediate results for use in future computations; the term recurrent is used to denote this memory-like capability of RNNs. The RNN architecture we use is shown in Fig. 3-4. The essential difference between the FNN in Fig. 3-3 and the RNN in Fig. 3-4 is that

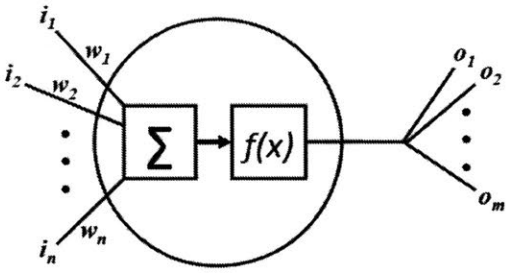


Figure 3-2: Individual processing unit or neuron. Inputs are multiplied by corresponding weights, summed, and passed to a nonlinear activation function. The i's correspond to inputs, the w's to weights, and the o's to outputs [35].

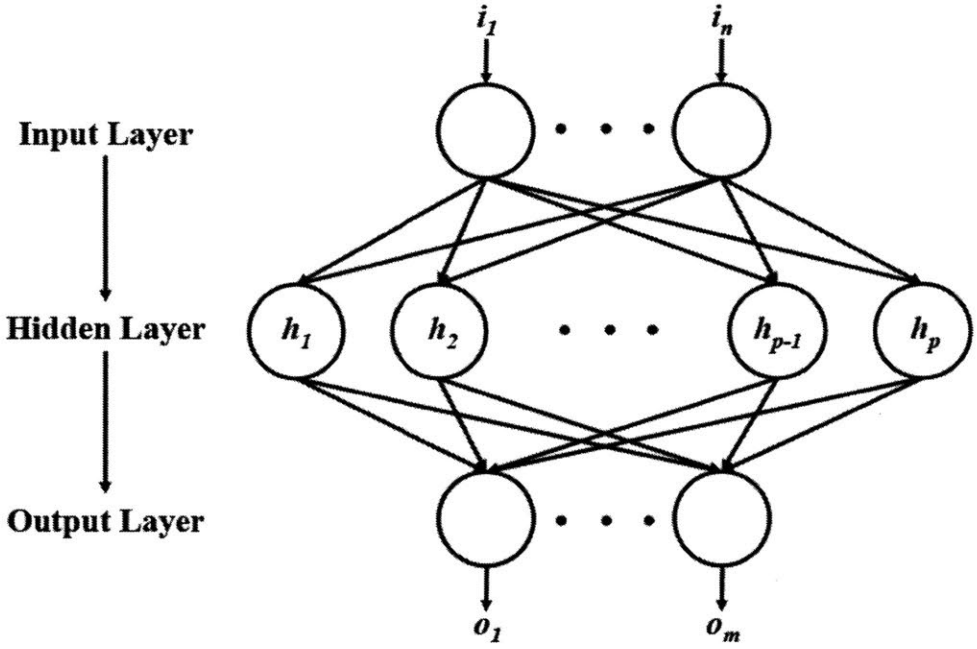


Figure 3-3: Feedforward neural network. Arrows indicate the direction of information flow. The i's correspond to inputs, the h's to hidden units, and the o's to outputs.

the output of the hidden layer in the RNN in one step of the computation is used as input to the hidden layer in the next step of the computation. Reuse of previous results allows the network to identify long-term dependencies in data that follow a sequence, making RNNs useful for the analysis of text, as well as time series data.

The training of a neural network is done in a supervised fashion, wherein the inputs are applied to the network and the resultant outputs are compared with the

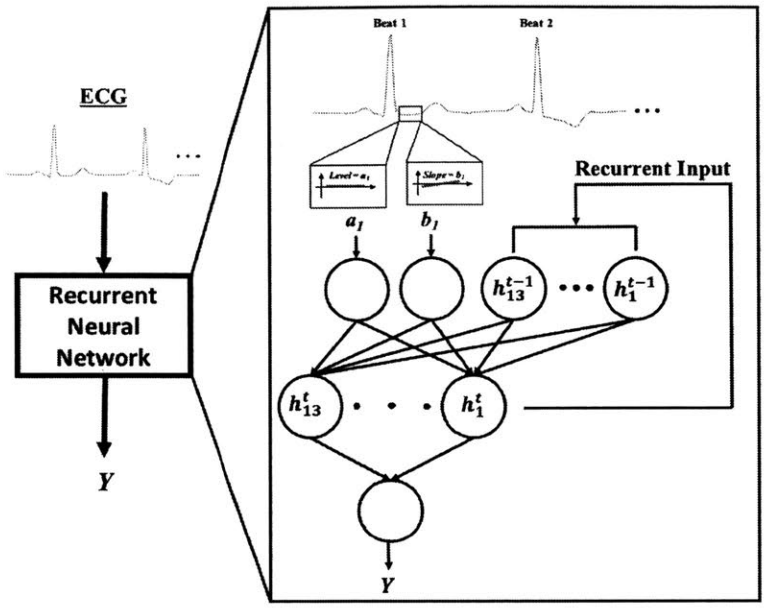


Figure 3-4: Schematic representation of the RNN model. The raw ECG signal is preprocessed and the first two Legendre polynomial coefficients are extracted for each beat, as shown in Fig. 3-1. The hidden units are denoted by $\boldsymbol{h}_i^t$ and the recurrent inputs are denoted by $\boldsymbol{h}_i^{t-1}$, where t is the current time step and $t-1$ is the previous time step. $\boldsymbol{Y}$ is the prediction.

expected outputs; the training process seeks to optimize the choice of the weights of the interconnections between neurons so as to minimize the error between the expected and computed output. The error is measured by an objective function that is chosen according to the use of the network, and is sent back through the network in order to adjust the weights; this process is called backpropagation. The adjustment of the weights is typically done using a method based on gradient descent, in which the gradients of the objective function with respect to the weights are computed using the errors calculated above and minimized. While the weights are updated after each training example is seen, the gradients may be updated after the network sees all examples, after it sees each example, or after seeing batches of examples; the first approach is called batch gradient descent, the second is called stochastic gradient descent, and the third is called mini-batch gradient descent. A single round of training

in which all examples in the training set are shown to the network is referred to as an epoch. The order of the training examples is usually shuffled at the beginning of each epoch, and the network is often trained using many epochs.

The time series nature of the Legendre polynomial coefficient data necessitates the use of a classifier capable of processing sequences of data; as discussed above, a RNN is well suited to this problem. The RNN employed consists an input layer of two units, a hidden layer of 13 units, and an output layer of one unit. Although 16 coefficients were generated from each ST-segment, only the first two were used in the final RNN classifier, since higher-order coefficients were found to capture too much noise in the data to offer accurate predictions. The number of hidden units was chosen so as to ensure that the number of free parameters, represented by the weights, of the model was limited to no more than 10% of the number of patients in the training set; with two input units, 13 hidden units, and one output unit, there are 208 weights, which is less than 10% of the patients in the entire Cohort-1 dataset. The classification problem to be solved by the network is binary, so only one output unit is needed. Training of the network was done using 100 epochs with a batch size of 32 patients, as this configuration was found to produce the maximal AUC. The objective function optimized was the binary cross-entropy loss function, which is commonly used for binary classification problems.

3.6 An Artificial Neural Network with ST-Segment Based Features

As our data suggest that combining features derived from the patient history with ST-segment based features leads to improved performance, we also evaluated a model that combines the output from a RNN and the LR_{Hx} models into one model (LR_{Hx}-RNN, Fig. 3-5). Both the RNN and LR_{Hx}-RNN models were associated with CVD post-ACS, with the LR_{Hx}-RNN model offering a markedly improved AUC over the RNN model ($p < 0.001$) and the best discriminatory value of all models tested ($p < 0.001$

for all pairwise comparisons) Table 3.1. The category-free net reclassification index (NRI) [37] of LR_{Hx}-RNN relative to the LR_{Hx+ST} model was 0.18 (18%), albeit the standard deviation of this value over 1000 bootstrapped datasets was 0.451.

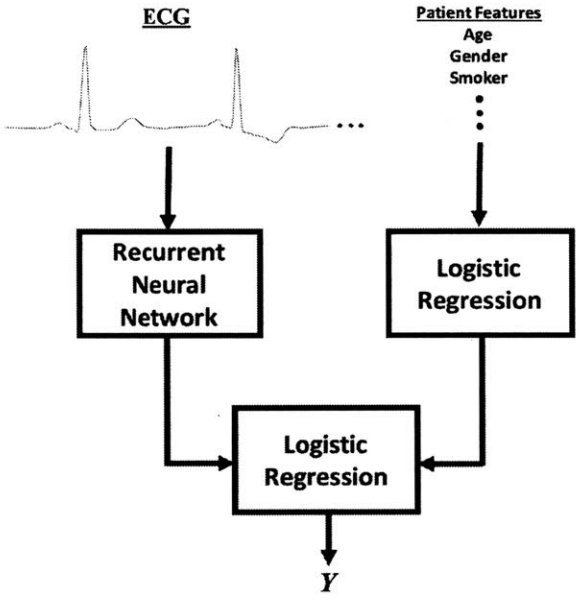


Figure 3-5: Schematic representation of the LR_{Hx}-RNN model. The predictions from the RNN and LR_{Hx} models are combined in a second level of logistic regression with weights $\boldsymbol{w_1}$ and $\boldsymbol{w_2}$ to form prediction $\boldsymbol{Y}$.

3.7 Univariate Association of Models with CVD in Cohort-1

We assessed the univariate association of the LR_{Hx+ST} and LR_{Hx}-RNN models with CVD at one year and 60, 30, and 14 days by calculating Hazard Ratios (HRs) using Cox proportional hazard models Table 3.2. A patient was deemed high-risk for the purpose of calculating the HR for each classifier if the raw prediction value for that patient and that classifier was in the upper-quartile of the raw predictions across all patients for that classifier. For comparison, we included results for the TIMI

NSTE-ACS risk score. HRs were deemed statistically significant if the lower 95% confidence interval remained above one. All HRs for the LR_{Hx+ST} and LR_{Hx}-RNN models were statistically significant with the LR_{Hx+ST} model showing the highest association with CVD. The TIMI risk score had significant HRs for CVD at one year, 60 days, and 14 days, although not at 30 days. One-year, 30-day and 14-day HRs for the LR_{Hx+ST} model were significantly higher than those associated with the TIMI risk score ($p < 0.05$), and one-year, 60-day, and 30-day HRs for the LR_{Hx}-RNN model were consistently higher than the TIMI risk score ($p < 0.05$).

	Hazard Ratio	**95% CI**
LR_{Hx+ST}		
1-Year	4.744	2.232 - 10.100
60-Day	5.918	1.841 - 19.892
30-Day	6.607	1.583 - 30.793
14-Day	6.606	1.229 - 39.907
LR_{Hx}-RNN		
1-Year	4.510	2.126 - 9.584
60-Day	4.943	1.580 - 16.202
30-Day	5.260	1.342 - 22.544
14-Day	5.546	1.068 - 32.413
TRS		
1-Year	3.667	1.761 - 7.638
60-Day	3.822	1.298 - 11.428
30-Day	3.309	0.920 - 12.512
14-Day	5.417	1.133 - 28.885

Table 3.2: Univariate Hazard Ratios (highest vs. other quartiles) calculated from the Cohort-1 dataset. Hazard Ratios and CIs represent averages over 1000 trials (each trial yields one HR and one 95% CI).

3.8 Multivariate Association of Models with CVD in Cohort-1

The HRs for the LR_{Hx+ST} and LR_{Hx}-RNN models for CVD at one year were adjusted for the TRS, left ventricular ejection fraction (LVEF), and brain natriuretic peptide (BNP), all of which are used clinically to identify high-risk patients [38], [39], [40],

[41]; the NSTE-ACS TRS model was also adjusted for LVEF and BNP only Table 3.3. For the multivariate HRs using the TRS, LVEF, and BNP, patients were grouped in the high-risk category using the following criteria: TRS > 4, LVEF $\leq 40\%$, and BNP > 80 pg/mL.

The LR_{Hx}-RNN model had the highest HR for the TRS+LVEF, TRS+BNP, and TRS+LVEF+BNP adjustments, while the LR_{Hx+ST} model had the highest HR for the TRS adjustment. More importantly, the HR for the LR_{Hx}-RNN model remains statistically significant at one year after accounting for all covariates and is larger than the HR for the TIMI risk score ($p < 0.05$).

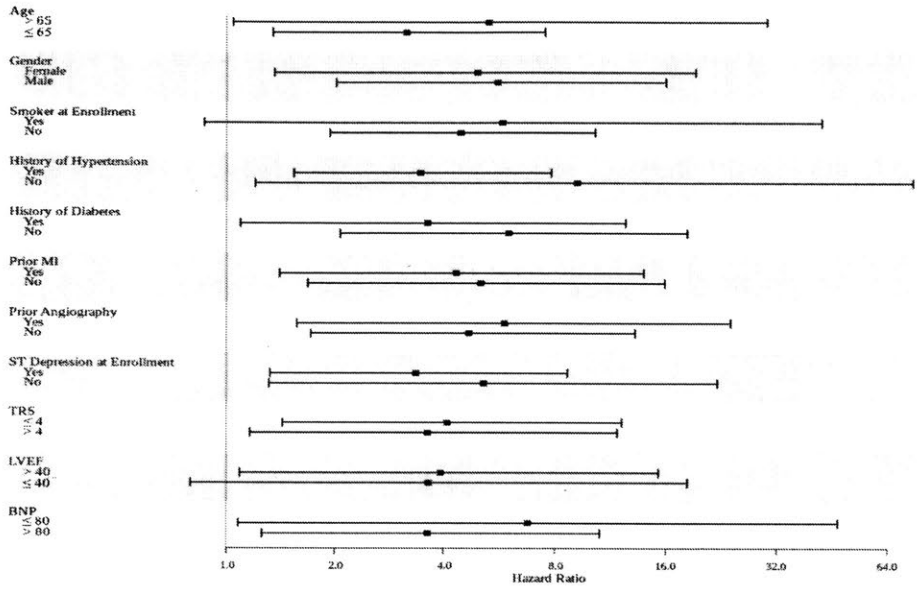
	1-Year HR	**95% CI**
LR_{Hx+ST}		
TRS	3.624	1.598 - 8.232
TRS+LVEF	2.823	1.016 - 7.924
TRS+BNP	3.430	1.243 - 9.565
TRS+LVEF+BNP	3.564	0.979 - 13.821
LR_{Hx}-RNN		
TRS	3.473	1.555 - 7.771
TRS+LVEF	2.956	1.075 - 8.221
TRS+BNP	3.431	1.267 - 9.419
TRS+LVEF+BNP	3.722	1.035 - 14.339
TRS		
LVEF	3.085	1.214 - 7.880
BNP	3.405	1.400 - 8.312
LVEF+BNP	3.118	1.001 - 9.875

Table 3.3: Multivariate Hazard Ratios on Cohort-1. HRs and CIs represent averages over 1000 trials (each trial yields one HR and one 95% CI).

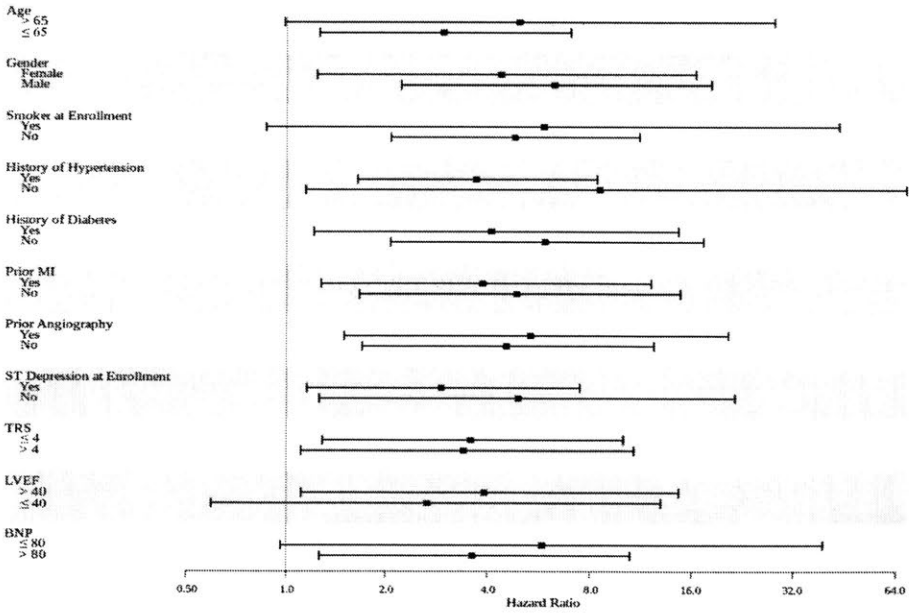
3.9 Univariate Association of Models with CVD in Specific Patient Populations

We further evaluated the performance of the LR_{Hx+ST} Fig. 3-6a and LR_{Hx}-RNN Fig. 3-6b models on specific patient populations. Univariate HRs were statistically significant in all subgroups examined except for Smoker-Yes and LVEF $\leq 40\%$ for the LR_{Hx}-RNN model, and Age > 65, Smoker-Yes, LVEF $\leq 40\%$, and BNP ≤ 80 pg/ml

for the LR_{Hx+ST} model. Most notably, the one-year HR for LR_{Hx+ST} and LR_{Hx}-RNN are significant in patients who do not have at least 1 mm ST segment depression at enrollment.



(a) LR_{Hx+ST}



(b) LR_{Hx}-RNN

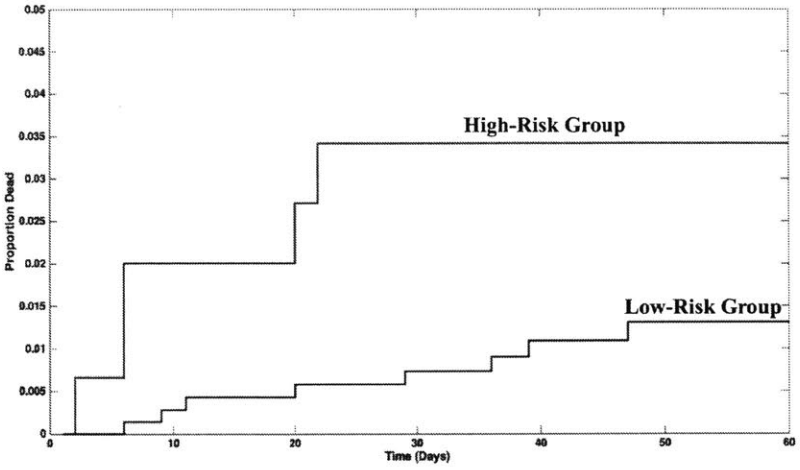
Figure 3-6: Univariate HRs in various subpopulations of Cohort-1. The black squares show mean values, and the horizontal bars represent 95% confidence intervals.

3.10 Association of Models with CVD in Cohort-2

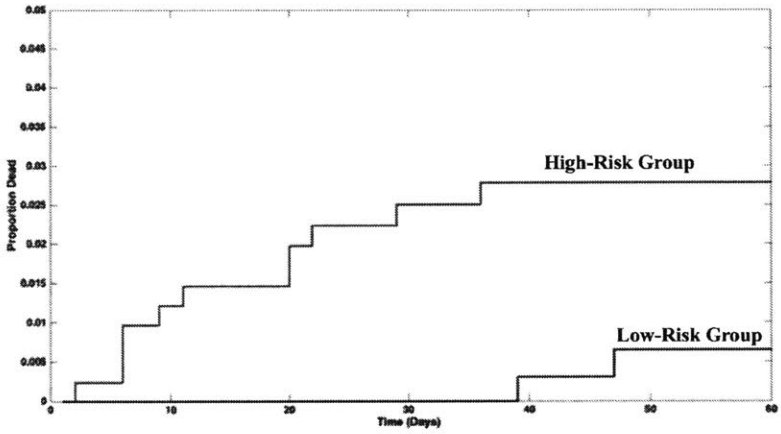
When applying the models to Cohort-2 the LR_{Hx+ST} and LR_{Hx}-RNN models were trained on the entire Cohort-1 dataset and then applied to the entire Cohort-2 dataset. In this sense, the Cohort-2 dataset was used as an independent holdout set to evaluate the ability of the models developed on Cohort-1 to generalize to unseen data. The LR_{Hx}-RNN model achieved an AUC of 0.767 on Cohort-2, as opposed to an AUC of 0.758 for the LR_{Hx+ST} model. Univariate one-year, 60-day, 30-day, and 14-day HRs were calculated for both models, and the LR_{Hx}-RNN model was found to have statistically significant results for all time periods considered; by contrast, the LR_{Hx+ST} model only had statistically significant results for one-year and 30-day outcomes Table 3.4. HRs for Cohort-2 were calculated using the upper-quartile of the raw predictions from Cohort-1. Patients in Cohort-2 were deemed high-risk if their raw prediction value for a given classifier exceeded the upper-quartile cutoff derived from Cohort-1. Overfitting of the models to the Cohort-1 dataset was mitigated by truncating the upper-quartile cutoff for each model to one significant digit. Multivariate HRs were not computed, as TRS, LVEF, and BNP measurements were not available in the Cohort-2 dataset. Kaplan-Meier survivor curves were generated the LR_{Hx}-RNN Fig. 3-7b and LR_{Hx+ST} Fig. 3-7a models for a follow-up period of 60 days.

	Hazard Ratio	**95% CI**
LR_{Hx+ST}		
1-Year	3.661	1.269 - 10.559
60-Day	2.992	0.979 - 9.146
30-Day	4.764	1.379 - 16.457
14-Day	4.750	0.959 - 23.534
LR_{Hx}-RNN		
1-Year	4.420	1.232 - 15.853
60-Day	4.773	1.313 - 17.349
30-Day	6.885	1.780 - 26.630
14-Day	15.711	3.170 - 77.855

Table 3.4: Univariate Hazard Ratios (highest vs. other quartiles) for Cohort-2.



(a) LR_{Hx+ST}



(b) LR_{Hx}-RNN

Figure 3-7: Kaplan-Meier survival curves on Cohort-2. The high-risk group consists of those patients with a raw prediction value on the Cohort-2 holdout set above the value representing the upper-quartile on the Cohort-1 training set.

3.11 Discussion

Our results demonstrate 1) the utility of combining easily obtained clinical history with independent information extracted from the ECG and 2) that signal processing and machine learning can be used to generate patient risk models with improved performance compared to traditional logistic regression techniques.

In our study population, the TIMI NSTE-ACS risk score has an AUC of 0.67 and therefore remains a useful discriminatory metric that can identify patients at

elevated risk of death one year post NSTE-ACS. However, only 20% of the patients in the Cohort-1 study belonged to the highest TRS category, and 54% of the deaths in Cohort-1 dataset occurred in patients in the low- and moderate-risk categories. Hence, the TRS by itself is insufficient to accurately risk-stratify the majority of patients in this population.

Comparing the relative performance of the first three models demonstrates Table 3.1 both the predictive power of the feature set derived from the ECG, as well as the utility of combining these features with those from the patient record. The LRST model uses only mean and standard deviation values derived from the first 50 usable beats of the ECG record for each patient, but offers a statistically significant improvement in performance over the LR_{Hx} model, which only uses patient features that are strongly associated with adverse cardiovascular outcomes. While both models perform comparably individually, clear improvements in performance can be achieved by combining the two types of clinical data into the LR_{Hx+ST} model, thereby demonstrating the utility of exploiting disparate pieces of information in a single model. Indeed, the LR_{Hx+ST} model offers the highest univariate HRs of all models considered.

Further gains in the AUC can be achieved by replacing the simple summary statistics for the ECG, which we use in the LR_{Hx+ST} model, with a RNN that is trained on the ECG time series alone; indeed, the LR_{Hx}-RNN achieves the highest AUC of all models considered, thereby offering the best predictive accuracy. Moreover, the LR_{Hx}-RNN model is the only model considered that exhibits statistically significant HRs when adjusted for all combinations of the TRS, LVEF, and BNP; it is also offers univariate HRs that are statistically significant in more subgroups than the LR_{Hx+ST} does. Moreover, the LR_{Hx}-RNN has statistically significant HRs at one year, 60 days, 30 days, and 14 days when applied to an independent dataset, while the LR_{Hx+ST} model does not.

The LR_{Hx}-RNN model utilizes the raw time series information while the LR_{Hx+ST} model uses simple summary statistics derived from this same time series, thereby losing this temporal information. Thus, the added predictive power of the ANN model

appears to come from the ability of the RNN to take into account the evolution of the ST segment morphology over time.

The LR_{Hx}-RNN model uses a relatively small amount of data - the first 50 beats in the ECG recording - to achieve this improved performance. It should be stressed, however, that while the ST segment morphology features were derived from 50 beats of ECG data corresponding to less than one minute of data in this patient population, we generally require more time to secure 50 beats that have ST segments that can be clearly identified using existing automated methods. Overall, approximately 45 minutes of data were required on average to obtain the requisite number of beats. Other risk metrics, like the TIMI and GRACE scores require the results of lab values that generally are available within an hour after the patient presents. Consequently, all of the data needed to assess patient risk with the LR_{Hx}-RNN model can be acquired in a similar time frame as the data needed to calculate more traditional risk metrics. Risk assessment within the first few hours after presentation plays an important role in the management of ACS as patients who are identified as high risk may benefit from invasive strategies within 24-48 hours after the diagnosis of ACS is made [6]. Interestingly, although our method uses a relatively small amount of ECG data, it has greater discriminatory power relative to some other ECG-based methods such as MV [19], HRV [42], and DC [33], that require at least 24 hours of data.

One limitation of this study is that a single sampling rate of 128 Hz was used for collecting the ECG data in both cohorts; thus, no claims can be made about the performance of the model on data sampled at different rates. However, only the first two Legendre polynomial coefficients, which represent the level and slope of the ST-segment, were used in the analysis, and it is unlikely that changing the sampling rate will drastically change the values of these coefficients. A second limitation of this study is that an automated algorithm was used to extract the ST-segments from the ECG data, so the criteria used by the algorithm to identify the endpoints of the ST-segment may influence the final values of the coefficients obtained. Nonetheless, if the ECG data are relatively clean, small changes in the ST-segment delineation are not expected to significantly affect the final results.

Finally, the strong association with CVD of the LR_{Hx}-RNN model on two independent datasets argues that this model generalizes well to unseen data. These observations demonstrate that machine learning provides a framework for fruitfully combining disparate pieces of data to yield models with improved performance.

Chapter 4

Conclusion

The foregoing discussions demonstrated how analysis techniques based on signal processing and machine learning can be used to effectively identify high-risk patients after a NSTE-ACS. In particular, the results of Chapter 3 showed that combining classifiers based on disparate pieces of clinical data can result in models with greater discriminative value than any of the individual classifiers alone. This formalism could potentially be extended to incorporate additional variables beyond simple patient record features and features derived from the ECG. For example, there is evidence to indicate that there is a genetic basis for ACS [43], so incorporating features based on genetic information into this formalism might further improve its predictive performance.

Identifying high-risk patients forms an important first step in managing patients who present with ACS, and assigning appropriate therapies to such patients constitutes a necessary second step in treating them. Thus, an interesting future direction might be to develop models capable of deriving a treatment regimen for a given patient based on that patient's medical history and vital signs. These two steps could then be combined to form an automated recommendation system that could assist clinicians in identifying high-risk patients and assigning appropriate therapy to them uniquely, thereby enabling personalized medicine. As demonstrated by the foregoing results, machine learning techniques show the potential to enable the efficient development of such systems.

Bibliography

[1] E. A. Amsterdam, N. K. Wenger, R. G. Brindis, D. E. Casey, Jr., T. G. Ganiats, D. R. Holmes, Jr., A. S. Jaffe, H. Jneid, R. F. Kelly, M. C. Kontos, G. N. Levine, P. R. Liebson, D. Mukherjee, E. D. Peterson, M. S. Sabatine, R. W. Smalling, S. J. Zieman, American College of Cardiology, American Heart Association Task Force on Practice Guidelines, Society for Cardiovascular Angiography Interventions, Society of Thoracic Surgeons, American Association for Clinical Chemistry, 2014 AHA/ACC Guideline for the Management of Patients with Non-ST-Elevation Acute Coronary Syndromes: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J. Am. Coll. Cardiol.* **64**, e139-228 (2014).

[2] D. Mozaffarian, E. J. Benjamin, A. S. Go, D. K. Arnett, M. J. Blaha, M. Cushman, S. R. Das, S. de Ferranti, J. P. Despres, H. J. Fullerton, V. J. Howard, M. D. Huffman, C. R. Isasi, M. C. Jimenez, S. E. Judd, B. M. Kissela, J. H. Lichtman, L. D. Lisabeth, S. M. Liu, R. H. Mackey, D. J. Magid, D. K. McGuire, E. R. Mohler, C. S. Moy, P. Muntner, M. E. Mussolino, K. Nasir, R. W. Neumar, G. Nichol, L. Palaniappan, D. K. Pandey, M. J. Reeves, C. J. Rodriguez, W. Rosamond, P. D. Sorlie, J. Stein, A. Towfighi, T. N. Turan, S. S. Virani, D. Woo, R. W. Yeh, M. B. Turner, American Heart Association Statistics Committee, Stroke Statistics Subcommittee, Heart Disease and Stroke Statistics-2016 Update A Report From the American Heart Association. *Circulation* **133**, E38-E360 (2016).

[3] C. P. Cannon, W. S. Weintraub, L. A. Demopoulos, R. Vicari, M. J. Frey, N.

Lakkis, F. J. Neumann, D. H. Robertson, P. T. DeLucca, P. M. DiBattiste, C. M. Gibson, E. Braunwald, TACTICS (Treat Angina with Aggrastat and Determine Cost of Therapy with an Invasive or Conservative Strategy) - Thrombolysis in Myocardial Infarction 18 Investigators, Comparison of early invasive and conservative strategies in patients with unstable coronary syndromes treated with the glycoprotein IIb/IIIa inhibitor tirofiban. *New Engl. J. Med.* **344**, 1879-1887 (2001).

[4] E. Diderholm, B. Andren, G. Frostfeldt, M. Genberg, T. Jernberg, B. Lagerqvist, B. Lindahl, L. Wallentin, 2nd, ST depression in ECG at entry indicates severe coronary lesions and large benefits of an early invasive treatment strategy in unstable coronary artery disease; the FRISC II ECG substudy. The Fast Revascularisation during InStability in Coronary artery disease. *Eur. Heart J.* **23**, 41-49 (2002).

[5] M. Roffi, C. Patrono, J. P. Collet, C. Mueller, M. Valgimigli, F. Andreotti, J. J. Bax, M. A. Borger, C. Brotons, D. P. Chew, B. Gencer, G. Hasenfuss, K. Kjeldsen, P. Lancellotti, U. Landmesser, J. Mehilli, D. Mukherjee, R. F. Storey, S. Windecker, H. Baumgartner, O. Gaemperli, S. Achenbach, S. Agewall, L. Badimon, C. Baigent, H. Bueno, R. Bugiardini, S. Carerj, F. Casselman, T. Cuisset, C. Erol, D. Fitzsimons, M. Halle, C. Hamm, D. Hildick-Smith, K. Huber, E. Iliodromitis, S. James, B. S. Lewis, G. Y. Lip, M. F. Piepoli, D. Richter, T. Rosemann, U. Sechtem, P. G. Steg, C. Vrints, J. Luis Zamorano, 2015 ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: Task Force for the Management of Acute Coronary Syndromes in Patients Presenting without Persistent ST-Segment Elevation of the European Society of Cardiology (ESC). *Eur. Heart J.* **37**, 267-315 (2016).

[6] A. A. Bavry, D. J. Kumbhani, A. N. Rassi, D. L. Bhatt, A. T. Askari, Benefit of Early Invasive Therapy in Acute Coronary Syndromes: A Meta-Analysis of

Contemporary Randomized Clinical Trials. *J. Am. Coll. Cardiol.* **48**, 1319-1325 (2006).

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

[8] D. A. Morrow, E. M. Antman, L. Parsons, J. A. de Lemos, C. P. Cannon, R. P. Giugliano, C. H. McCabe, H. V. Barron, E. Braunwald, Application of the TIMI risk score for ST-Elevation MI in the National Registry of Myocardial Infarction 3. *JAMA-J. Am. Med. Assoc.* **286**, 1356-1359 (2001).

[9] C. B. Granger, R. J. Goldberg, O. Dabbous, K. S. Pieper, K. A. Eagle, C. P. Cannon, F. Van de Werf, A. Avezum, S. G. Goodman, M. D. Flather, K. A. Fox, Global Registry of Acute Coronary Events Investigators, Predictors of hospital mortality in the global registry of acute coronary events. *Arch. Intern. Med.* **163**, 2345-2353 (2003).

[10] Y. Liu, Z. Syed, B. M. Scirica, D. A. Morrow, J. V. Guttag, C. M. Stultz, ECG Morphological Variability in Beat Space for Risk Stratification After Acute Coronary Syndrome. *JAHA-J. Am. Heart Assoc.* **3**, (2014).

[11] P. de Araujo Goncalves, J. Ferreira, C. Aguiar, R. Seabra-Gomes, TIMI, PURSUIT, and GRACE risk scores: sustained prognostic value and interaction with revascularization in NSTE-ACS. *Eur. Heart J.* **26**, 865-872 (2005).

[12] R. Bigi, D. Gregori, L. Cortigiani, A. Desideri, F. A. Chiarotto, G. M. Toffolo, Artificial neural networks and robust Bayesian classifiers for risk stratification following uncomplicated myocardial infarction. *Int. J. Cardiol.* **101**, 481-487 (2005).

[13] H. Haraldsson, L. Edenbrandt, M. Ohlsson, Detecting acute myocardial infarction in the 12-lead ECG using Hermite expansions and neural networks. *Artif. Intell. Med.* **32**, 127-136 (2004).

[14] M. Ennis, G. Hinton, D. Naylor, M. Revow, R. Tibshirani, A comparison of statistical learning methods on the GUSTO database. *Stat. Med.* **17**, 2501-2508 (1998).

[15] R. F. Harrison, R. L. Kennedy, Artificial neural network models for prediction of acute coronary syndromes using clinical data from the time of presentation. *Ann. Emerg. Med.* **46**, 431-439 (2005).

[16] A. Chouldechova, T. Hastie, Generalized additive model selection. *Ann. Appl. Stat.*, submitted (2015).

[17] A. Demmler, C. Reinsch, Oscillation matrices with spline smoothing. *Numerische Mathematik* **24**, 375-382 (1975).

[18] R. Caruana, Y. Lou, J. Gehrke, P. Koch, M. Sturm, N. Elhadad, Intelligible models for health care: Predicting pneumonia risk and hospital 30-day readmission. *Knowledge Discovery and Data Mining* (2015).

[19] Z. Syed, B. M. Scirica, S. Mohanavelu, P. Sung, E. L. Michelson, C. P. Cannon, P. H. Stone, C. M. Stultz, J. V. Guttag, Relation of Death Within 90 Days of Non-ST-Elevation Acute Coronary Syndromes to Variability in Electrocardiographic Morphology. *Am. J. Cardiol.* **103**, 307-311 (2009).

[20] Z. Syed, C. M. Stultz, B. M. Scirica, J. V. Guttag, Computationally generated cardiac biomarkers for risk stratification after acute coronary syndrome. *Sci. Transl. Med.* **3**, 102ra195 (2011).

[21] T. J. Hastie, Generalized Additive Models. Ch. 7 of *Statistical Models in S*, Wadsworth and Brooks/Cole (1992).

[22] T. Hastie, R. Tibshirani, Generalized Additive Models. London: Chapman and Hall (1990).

[23] S. Savonitto, D. Ardissino, C. B. Granger, G. Morando, M. D. Prando, A. Mafrici, C. Cavallini, G. Melandri, T. D. Thompson, A. Vahanian, E. M. Ohman, R. M. Califf, F. Van de Werf, E. J. Topol, Prognostic value of the admission electrocardiogram in acute coronary syndromes. *JAMA-J. Am. Med. Assoc.* **281**, 707-713 (1999).

[24] C. P. Cannon, C. H. McCabe, P. H. Stone, W. J. Rogers, M. Schactman, B. W. Thompson, D. J. Pearce, D. J. Diver, C. Kells, T. Feldman, M. Williams, R. S. Gibson, M. W. Kronenberg, L. I. Ganz, H. V. Anderson, E. Braunwald, The electrocardiogram predicts one-year outcome of patients with unstable angina and non-Q wave myocardial infarction: results of the TIMI III Registry ECG Ancillary Study. Thrombolysis in Myocardial Ischemia. *J. Am. Coll. Cardiol.* **30**, 133-140 (1997).

[25] P. Kaul, Y. Fu, W. C. Chang, R. A. Harrington, G. S. Wagner, S. G. Goodman, C. B. Granger, D. J. Moliterno, F. Van de Werf, R. M. Califf, E. J. Topol, P. W. Armstrong, for the PARAGON-A and GUSTO IIb Investigators, Prognostic value of ST segment depression in acute coronary syndromes: insights from PARAGON-A applied to GUSTO-IIb. PARAGON-A and GUSTO IIb Investigators. Platelet IIb/IIIa Antagonism for the Reduction of Acute Global Organization Network. *J. Am. Coll. Cardiol.* **38**, 64-71 (2001).

[26] P. Damman, L. Holmvang, J. G. Tijssen, B. Lagerqvist, T. C. Clayton, S. J. Pocock, F. Windhausen, A. Hirsch, K. A. Fox, L. Wallentin, R. J. de Winter, Usefulness of the admission electrocardiogram to predict long-term outcomes after non-ST-elevation acute coronary syndrome (from the FRISC II, ICTUS, and RITA-3 [FIR] Trials). *Am. J. Cardiol.* **109**, 6-12 (2012).

[27] W. R. Hathaway, E. D. Peterson, G. S. Wagner, C. B. Granger, K. M. Zabel, K. S. Pieper, K. A. Clark, L. H. Woodlief, R. M. Califf, Prognostic significance of the

initial electrocardiogram in patients with acute myocardial infarction. GUSTO-I Investigators. Global Utilization of Streptokinase and t-PA for Occluded Coronary Arteries. *JAMA-J. Am. Med. Assoc.* **279**, 387-391 (1998).

[28] M. Amon, F. Jager, Electrocardiogram ST-Segment Morphology Delineation Method Using Orthogonal Transformations. *PLOS One* **11**, (2016).

[29] L. N. Trefethen, D. Bau, *Numerical Linear Algebra*, SIAM: 1997.

[30] Q. Li, R. G. Mark, G. D. Clifford, Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter. *Physiol. Meas.* **29**, 15-32 (2008).

[31] J. P. Martinez, R. Almeida, S. Olmos, A. P. Rocha, P. Laguna, A wavelet-based ECG delineator: Evaluation on standard databases. *IEEE T. Bio-Med. Eng.* **51**, 570-581 (2004).

[32] A. Erdogan, M. Coch, M. Bilgin, M. Parahuleva, H. Tillmanns, B. Waldecker, N. Soydan, Prognostic value of heart rate variability after acute myocardial infarction in the era of immediate reperfusion. *Herzschrittmacherther Elektrophysiol* **19**, 161-168 (2008).

[33] A. Bauer, J. W. Kantelhardt, P. Barthel, R. Schneider, T. Makikallio, K. Ulm, K. Hnatkova, A. Schornig, H. Huikuri, A. Bunde, M. Malik, G. Schmidt, Deceleration capacity of heart rate as a predictor of mortality after myocardial infarction: cohort study. *Lancet* **367**, 1674-1681 (2006).

[34] M. De Beule, E. Maes, O. De Winter, W. Vanlaere, R. Van Impe, Artificial neural networks and risk stratification: A promising combination. *Math. Comput. Model.* **46**, 88-94 (2007).

[35] P. A. Jansson, Neural networks: An overview. *Anal. Chem.* **63**, 357A-362A (1991).

[36] A. C. Tsoi, A. Back, Discrete time recurrent neural network architectures: A unifying review. *Neurocomputing* **15**, 183-223 (1997).

[37] M. J. Pencina, R. B. D'Agostino, Sr., E. W. Steyerberg, Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat. Med.* **30**, 11-21 (2011).

[38] R. A. Nishimura, G. S. Reeder, F. A. Miller, Jr., D. M. Ilstrup, C. Shub, J. B. Seward, A. J. Tajik, Prognostic value of predischarge 2-dimensional echocardiogram after acute myocardial infarction. *Am. J. Cardiol.* **53**, 429-432.

[39] J. E. Moller, G. S. Hillis, J. K. Oh, G. S. Reeder, B. J. Gersh, P. A. Pellikka, Wall motion score index and ejection fraction for risk stratification after acute myocardial infarction. *Am. Heart J.* **151**, 419-425.

[40] J. A. de Lemos, D. A. Morrow, J. H. Bentley, T. Omland, M. S. Sabatine, C. H. McCabe, C. Hall, C. P. Cannon, E. Braunwald, The Prognostic Value of B-Type Natriuretic Peptide in Patients with Acute Coronary Syndromes. *New Engl. J. Med.* **345**, 1014-1021 (2001).

[41] A. M. Richards, M. G. Nicholls, E. A. Espiner, J. G. Lainchbury, R. W. Troughton, J. Elliott, C. Frampton, J. Turner, I. G. Crozier, T. G. Yandle, B-Type Natriuretic Peptides and Ejection Fraction for Prognosis After Myocardial Infarction. *Circulation* **107**, 2786-2792 (2003).

[42] A. J. Camm, M. Malik, J. T. Bigger, G. Breithardt, S. Cerutti, R. J. Cohen, P. Coumel, E. L. Fallen, H. L. Kennedy, R. E. Kleiger, F. Lombardi, A. Malliani, A. J. Moss, J. N. Rottman, G. Schmidt, P. J. Schwartz, D. H. Singer, Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. *Eur. Heart J.* **17**, 354-381 (1996).

[43] M. Franchini, Genetics of the acute coronary syndrom. *Ann. Transl. Med.* **4** (2016).