Title
Mining Electronic Health Records to Improve Remote Health Monitoring

Permalink
https://escholarship.org/uc/item/6360z5q9

Author
Sideris, Konstantinos Sideris

Publication Date
2016

Peer reviewed|Thesis/dissertation
Mining Electronic Health Records to Improve Remote Health Monitoring

A Dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Computer Science

by

Konstantinos Sideris

2016
ABSTRACT OF THE DISSERTATION

Mining Electronic Health Records to Improve Remote Health Monitoring

by

Konstantinos Sideris

University of California, Los Angeles, 2016

Professor Majid Sarrafzadeh, Chair

Remote Health Monitoring Systems (RHMS) provide a continuous stream of patient physiological data that allows nurses and doctors to make timely decisions and help patients manage their chronic conditions. As the number of patients effectively managed by these informational systems increases, the need for more sophisticated approaches in health analytics becomes significant.

The purpose of this work is to provide methodologies and tools for a more systematic approach in building RHMS. New algorithms are presented that mine various informational sources, especially Electronic Health Records (EHR), to design more generalizable analytics models for RHMS. In addition, a prototype RHMS system is developed that collects and graphically represents health data from a variety of sources.
The dissertation of Konstantinos Sideris is approved.

Douglas Stott Parker

Miryung Kim

Jo-Ann Eastwood

Majid Sarrafzadeh, Committee Chair

University of California, Los Angeles

2016
To my beautiful wife Eleni, my parents and family . . .

who—among so many other things—

understood and supported me

throughout this endeavor
# Table of Contents

1 Introduction ................................................... 1  
   1.1 Background .................................................. 1  
   1.2 Statement of the Problem ................................. 4  
   1.3 Purpose of the Study ........................................ 4  

2 Review of the Literature ..................................... 6  
   2.1 Remote Health Monitoring ................................. 6  
   2.2 Electronic Health Records ................................. 9  
   2.3 Structured Knowledge Representation in Medicine ....... 10  

3 Building Severity of Condition Prediction Models from Electronic Health Records for Remote Health Monitoring Patients ................. 12  
   3.1 Background .................................................. 12  
      3.1.1 Electronic Health Records ............................... 12  
      3.1.2 Remote Health Monitoring System (RHMS) Clinical Trial ... 12  
   3.2 Methodology ................................................ 13  
      3.2.1 EHR Data Processing .................................... 14  
      3.2.2 Predicting Length of Stay in EHR ....................... 14  
      3.2.3 EHR-based Risk Model for First Readmission .......... 16  
      3.2.4 Applying First Admission Risk Model in BEAT- Clinical Trial Data ........................................... 21  
      3.2.5 EHR-based Risk Model for Readmission ................. 22  
      3.2.6 Applying Readmission Risk Model in BEAT-HF data .... 24
3.3 Results ............................................................. 24
  3.3.1 Predicting First Readmission in BEAT-HF ............... 24
  3.3.2 Predicting Readmission in BEAT-HF .................... 26
3.4 Conclusions .................................................... 28

4 Multiple Model Analytics for Adverse Event Prediction in Remote Health Monitoring Systems ........................................ 30
  4.1 Background .................................................. 30
  4.2 Methodology ................................................ 30
  4.3 Results & Conclusions ..................................... 33

5 Effects of Coaching on Adherence in Remote Health Monitoring Systems ......................................................... 35
  5.1 Introduction .................................................. 36
  5.2 Background and Related Works ............................ 37
  5.3 Study & RHMS design ...................................... 38
    5.3.1 Wanda-CVD RHMS .................................... 38
    5.3.2 Women’s Heart Health study ......................... 40
    5.3.3 Automated Wireless Coaching Messages ............ 40
    5.3.4 Clinician Messages .................................. 41
  5.4 Methodologies .............................................. 41
    5.4.1 Quantifying Adherence ................................ 41
    5.4.2 Data Analysis ....................................... 44
  5.5 Results and Discussion .................................... 46
    5.5.1 Analyzing Correlation of Messages ................. 46
    5.5.2 Predicting Weekly Participant Adherence .......... 48
7.3.2 Limitations .............................................. 81
7.4 Conclusions .............................................. 81
7.5 Abbreviations ............................................. 85

8 A Big-Data platform for Medical Knowledge Extraction from Electronic Health Records: Automatic Assignment of ICD-9 Codes ............. 88

8.1 Introduction .............................................. 88
8.1.1 Modeling & Mining Medical Knowledge .............. 89
8.1.2 Disease Diagnostic Codes ............................... 90
8.2 Methodology ............................................. 91
8.2.1 Architecture ........................................... 91
8.2.2 Data .................................................. 91
8.2.3 Machine Learning Pipeline for ICD-9 assignment .. 93
8.2.4 Evaluation ............................................. 93
8.3 Results .................................................. 94
8.4 Conclusions ............................................. 97

9 Building Continuous Arterial Blood Pressure Prediction Models Using Recurrent Networks ........................................... 98

9.1 Introduction .............................................. 98
9.2 Background & Related Works ............................ 101
9.2.1 Cuffless Blood Pressure monitoring .................. 101
9.2.2 Remote Health Monitoring Systems .................. 102
9.3 Methodology ............................................. 103
9.3.1 Data .................................................. 103
9.3.2 Long-Short Term Memory Neural Networks . . . . . . . . . . 103
9.4 Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 107
  9.4.1 Arterial Blood Pressure Estimation . . . . . . . . . . . . . . 107
  9.4.2 Comparison with Linear Regression . . . . . . . . . . . . . . 110
  9.4.3 Limitations . . . . . . . . . . . . . . . . . . . . . . . . . . . . 110
9.5 Conclusions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 117

10 Summary and Contributions . . . . . . . . . . . . . . . . . . . . . . . . 118
  10.1 Predictive Methodologies for Remote Health Monitoring . . . . . . . 118
  10.2 Tools and Frameworks for Mining Health data . . . . . . . . . . . . . 119
# List of Figures

1.1 Standard Remote Health Monitoring System Architecture .......................... 3

3.1 Hierarchical Clustering of patients according to their minimum systolic pressure, heart rate range during the first day of hospitalization. Colors represent average risk per group, with blue/green and red representing the lowest and highest risk respectively ............................................. 18

3.2 Age versus Length of Stay for CHF patients. It can be seen that there is an age bias on length of hospitalization. The LOS values are scaled between 0 and 1 .......................................................... 19

3.3 Hierarchical Clustering of patients with the risk factor normalized with regards to Age/Gender .......................................................... 20

3.4 Readmission risk clusters from EHR data normalized by gender, race and ethnicity .......................................................... 23

3.5 Predicted Risk .................................................................................. 25

3.6 3-day Readmission Rate .................................................................... 25

3.7 10-day Readmission Rate .................................................................. 25

3.8 30-day Readmission Rate .................................................................. 25

3.9 Root Mean Square Error of the predicted first day risk level vs the observed risk .......................................................... 26

3.10 Predicted Risk .................................................................................. 27

3.11 3-day Readmission Rate .................................................................... 27

3.12 10-day Readmission Rate .................................................................. 27

3.13 10-day Readmission Rate .................................................................. 27
3.14 Root Mean Square Error of the predicted last day risk level vs the observed risk ........................................... 28

4.1 Multiple model classification for medical adverse event prediction .......... 31
4.2 Clustering the data samples based on weight and blood pressure. .......... 32
4.3 ROC curves for single model and cluster-specific multiple model. .......... 34

5.1 Wanda-CVD architecture diagram ................. 39
5.2 Predictive pipeline in Wanda-CVD to estimate participant adherence the following week. ........................................... 49
5.3 Receiver operating characteristic curves for the Activity and Blood Pressure prediction tasks. ........................................... 50
5.4 Receiver operating characteristic curves for the Daily and Weekly Questionnaire prediction tasks. ........................................... 51
5.5 Accuracy, Area under the Curve and LOOCV standard deviation of error for the Activity and Blood Pressure prediction tasks and each of the classifiers ........................................... 53
5.6 Accuracy, Area under the Curve and LOOCV standard deviation of error for the Daily and Weekly Questionnaire prediction tasks and each of the classifiers ........................................... 54

6.1 Health Knowledge Graph Architecture .......................... 58
6.2 Patient Measurements in the Knowledge Graph. Timestamps are also represented as nodes (minutes, seconds etc) to facilitate retrieval of data based on time. ........................................... 59
6.3 Example from the disease subgraph ............................ 61
6.4 Performing health knowledge graph queries ................. 64
7.1 Comorbidity co-occurrence network. The nodes are scaled and colored according to their degree in the network. Such networks can explain how certain disease groups are more likely than others. They can also be used to analyze similarities between patients, diseases and different treatments.

7.2 Classification Accuracy vs number of disease clusters. The classification accuracy increases with the number of generated features until over-fitting phenomena appear.

7.3 Predictive pipeline that utilizes the proposed comorbidity framework.

7.4 Receiver Operating Characteristic curves for the $I_1-I_2$ CHF outcomes.

7.5 Receiver Operating Characteristic curves for the $I_3-I_4$ CHF outcomes.

7.6 Receiver Operating Characteristic curves for the $I_5$ CHF outcome.

7.7 Receiver Operating Characteristic curves for the UCI dataset readmission tasks.

7.8 Classification Accuracy vs # disease clusters for the $I_1-I_2$ CHF outcomes. The accuracy value for each of the other comorbidity frameworks is displayed with a single line across all clusters for comparison.

7.9 Classification Accuracy vs # disease clusters for the $I_3-I_4$ CHF outcomes.

7.10 Classification Accuracy vs # disease clusters for the $I_5$ CHF outcome.

7.11 Classification Accuracy vs # disease clusters for the UCI dataset readmission tasks.

8.1 Architectural diagram of the proposed big-data platform.

8.2 Hierarchical machine learning pipeline.

8.3 Precision, recall, f-measure and classification accuracy for the most populated ICD-9 top level categories.
8.4  Precision recall curve for each level of the classifier .......................... 96

9.1  Standard architecture of an RHMS. Data collected from the patient is transmitted to a cloud server. The server analyzes and presents the data to clinicians and provides feedback to patients. .......................... 100

9.2  A depiction of the training and test sets used ............................ 104

9.3  The input of the network is the processed, windowed PPG signal and the output is the ABP value at the current time step. .......................... 106

9.4  Systolic and diastolic blood pressure validation mean square error over training epochs (single patient) .......................... 108

9.5  Systolic, diastolic and arterial blood pressure test mean square error over successive model updates (single patient) .......................... 109

9.6  Systolic Blood Pressure RMSE in linear regression vs our proposed methodology .......................... 113

9.7  Diastolic Blood Pressure RMSE in linear regression vs our proposed methodology .......................... 114

9.8  Arterial blood pressure signal prediction (mean subtracted) at the beginning of the training (subject 7) .......................... 115

9.9  Arterial blood pressure signal prediction (mean subtracted) at the end of the training (subject 7) .......................... 116
LIST OF TABLES

3.1 Attributes as ranked by the correlation-based feature subset selection . . 15
3.2 Average root mean square error of generalized linear regression with 
10-fold cross validation . . . . . . . . . . . . . . . . . . . . . . . . . . 16

5.1 Automated Messages - Inspirational & Blood Pressure Reminders . . . 42
5.2 Automated Messages - Questionnaire Reminders . . . . . . . . . . . . 43
5.3 Clinician Messages . . . . . . . . . . . . . . . . . . . . . . . . . . . . 43
5.4 Statistically significant interventions that result in different levels of 
adherence. Results are generated using two-tailed t-tests (Welch) . . . . 47
5.5 Clinician Positive Reinforcement and Inconsistent Adherence . . . . . 48
5.6 Most Predictive Features for Adherence . . . . . . . . . . . . . . . . . 52

7.1 Daily Threshold Alerts . . . . . . . . . . . . . . . . . . . . . . . . . . 74
7.2 Feature and outcome statistics in the Ronald Reagan CHF dataset . . . 75
7.3 CHF outcome prediction . . . . . . . . . . . . . . . . . . . . . . . . . . 77
7.4 UCI dataset readmission prediction . . . . . . . . . . . . . . . . . . . . 78
7.5 Disease Clusters for Dendrogram Height = 1.5 . . . . . . . . . . . . . 87

8.1 Classification Results For Diseases Of The Circulatory System . . . . 95

9.1 Regression Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . 111
9.2 Regression Results (continued) . . . . . . . . . . . . . . . . . . . . . . 112
I would like to express my immense gratitude to my advisor Professor Majid Sarrafzadeh, for his guidance and support throughout the Ph.D. During my time under his mentorship, I learned everything I know about academic research. I would also like to thank my committee members, Professor Parker, Professor Miryung and Professor Eastwood for serving in my committee and for their priceless feedback during the entire process. I would like to also acknowledge Professor Jo-Ann Eastwood for teaching me about academic etiquette. Finally, I would like to extend my gratitude to Professor Petros Faloutsos for welcoming me into UCLA and mentoring me during my first years in the Ph.D. program.

Throughout the years in the Ph.D. program I had the opportunity and pleasure to work and collaborate with numerous researchers. Nabil Alshurafa was instrumental in pushing me to achieve my research potential with his meticulous feedback and guidance. Haik Kalantarian has been a great collaborator with whom I worked on several publications. Mohammad Pourhomayoun helped me formulate my dissertation ideas and involved me in a variety of exciting Wireless Health projects.

A version of chapter 3 was published in the Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing with the title "Using electronic health records to predict severity of condition for congestive heart failure patients". In this publication, I conducted the data analysis and methodology. M. Pourhomayoun contributed to methodology and writing of the paper. B Shahbazi helped with the data extraction and preparation and N. Alshurafa helped with the editing and writing of the paper. M. Sarrafzadeh was the PI.

Chapter 4 was published in the Proceedings of the 2014 Healthcare Innovation Conference (HIC) with the title "Multiple model analytics for adverse event prediction in remote health monitoring systems". In this publication, I contributed in the data analysis and methodology. M. Pourhomayoun worked on methodology and writing of the
paper. B. Mortazavi, H. Ghasemzadeh and N. Alshurafa helped with the editing and writing of the paper. M. Sarrafzadeh was the PI together with M. Ong. The rest of the authors participated in the study design and data collection.

Chapter 5 was published in the Proceedings of the conference on Wireless Health, 2015 with the title "Effects of coaching on adherence in remote health monitoring systems: analysis and prediction of participant adherence". In this publication, I conducted the data analysis and methodology. N. Alshurafa helped with the writing of the paper and was responsible for the technical aspects of the study design and data collection. H. Kalantarian contributed in the editing of the paper. JA Eastwood was the PI and M. Sarrafzadeh was the co-PI.

A version of chapter 7 was published in the Proceedings of the Engineering in Medicine and Biology Society (EMBC), 2015 with the title "A data-driven feature extraction framework for predicting the severity of condition of congestive heart failure patients.". The full chapter was also published in Elsevier’s Computers in Biology and Medicine journal. In this publication, I conducted the data analysis and methodology. N. Alshurafa helped with the writing of the paper. M Pourhomayoun helped with the design of the methodology. F Shahmohammadi processed the raw data of the comorbidity frequencies. L Samy contributed in the editing of the paper. H. Kalantarian participated in the writing of the journal version. M. Sarrafzadeh was the PI.

Chapter 8 was accepted for publication in the Proceedings of the 2016 ACM Petra conference with the title "A Big-Data Platform for Medical Knowledge Extraction from Electronic Health Records: Automatic Assignment of ICD9 Codes". In this publication, I conducted the writing and methodology design. H. Kalantarian helped with the design of the methodology. S Shaikh conducted the experiments and platform programming. M. Sarrafzadeh was the PI.

Chapter 9 was accepted for publication in the Proceedings of the 2016 SMART-COMP conference with the title "Using Recurrent Networks to Build Continuous Arterial Blood Pressure Prediction Models". In this publication, I conducted the method-
ology design and experiments. E. Nemati helped with the design of the methodology. H. Kalantarian contributed in the editing of the paper. M. Sarrafzadeh was the PI.
VITA

2009–2010 Junior Researcher, Decision Support Systems lab (DSS), National Technical University of Athens, Greece

2010 Electrical and Computer Engineering, Joint B.S. / M.S. (Diploma). National Technical University of Athens, Greece.

2010 UCLA Computer Science PhD fellowship.

2012 Honorable mention in the National Science Foundation Graduate Research Fellowship Program.

2011 Researcher, Division of Biomedical Informatics, University of California, San Diego

2012 Computer Science, M.S. University of California, Los Angeles.

2011–2016 Teaching Assistant, Computer Science Department, UCLA.

2011–2016 Research Assistant, Computer Science Department, UCLA.


PUBLICATIONS

Multiple model analytics for adverse event prediction in remote health monitoring systems. M. Pourhomayoun et al. Healthcare Innovation Conference (HIC), 2014 IEEE


xix
CHAPTER 1

Introduction

1.1 Background

Medicine as a vocation has appeared in various forms during the course of human history [2]. For many centuries, it was practiced by a medicine man who would apply herbs and say prayers for healing. This and related forms of healing are now referred to as alternative medicine. Modern medical science appeared at the end of the 18th century with the discovery of the smallpox vaccine by Edward Jenner [77]. Its foundations can be traced back all the way back to ancient Greece [31], where the concept of medicine became formalized with the teachings of Hippocrates, who for the first time described many diseases and medical conditions [3, 23].

In the late 20th century, medicine was further formalized with the introduction of evidence based medicine (EBM) [85]. Evidence based medicine consists of the following steps: 1) A question is formulated. 2) Previous relevant evidence is examined with regards to the question. 3) The benefits, harms, and costs are compared. 4) The guideline is implemented. 5) The guideline is evaluated in practice. The introduction of EBM revolutionized both clinical practice and medical research. Randomized controlled trials became the norm when introducing a new therapy or drug.

Nowadays, genetic research and bioinformatics [39] support medical research with new insights on the etiology of diseases and the genetic predispositions that are important when treating a patient. In fact, new genetic insights and therapies are constantly being introduced in the battle against cancer and other chronic conditions such as heart
failure [114, 84]

Despite all these achievements, a limitation of medicine has been the fact that patients seek healthcare after a new condition has developed or an existing one has worsened. This limits the ability of clinicians to prevent illnesses and treat their patients effectively. It also limits the ability of medical researchers to study the development and progression of such conditions. In recent years, advances in computer science and engineering have made it possible for people to be effectively monitored outside traditional healthcare environments. This has led to novel software and hardware based approaches for diagnosing, treating, predicting and avoiding various medical conditions. A class of new technologies that encompass many of these advances are Remote Health Monitoring Systems (RHMS). RHMS provide a continuous stream of patient physiological data in real time that allows nurses and doctors to make timely decisions and help patients manage their chronic conditions. Since an individual’s condition is influenced by a variety of factors, RHMS rely on collecting as much data as possible from subjects during their everyday lives using smart phones and smartwatches embedded with or connected to accelerometers and other sensors such as pressure sensitive bedsheets, activity trackers etc. A common architecture for such a system is illustrated in figure 1.1

As shown in figure 1.1, a RHMS consists of several components. First, a set of sensors relevant to each monitored condition is distributed to each individual that is monitored. These devices capture data at predetermined intervals. Data come in many forms, such as activity measurements from motion sensors, location information from global positioning systems (GPS), questionnaire prompts and vital signs such as heart rate blood pressure etc. The data is then preprocessed and transmitted to a cloud or traditional backend which further processes the data and persistently stores it. Conventional RHMS rely on threshold-based alerts. The thresholds are predetermined based on medical expertise to alert clinicians when physiological data deviate from set thresholds. Analytics-based RHMS on the other hand employ machine learning algorithms
Figure 1.1: Standard Remote Health Monitoring System Architecture
to predict the risk of an adverse event.

1.2 Statement of the Problem

As the number of patients effectively managed by these informational systems increases, the need for more sophisticated approaches in health analytics becomes significant. Traditionally, the process of designing the necessary machine learning and analytics algorithms is guided by an initial, usually smaller, pilot study that provides sufficient data for design and validation. However, this approach suffers from several limitations. It requires substantial effort to collect such initial data, spanning months or even years depending on the chronic condition of interest. In addition, selecting which physiological data to collect as well as the collection frequency is based on guesswork and physician experience since there is limited data to support the decision. Since the learning algorithms are trained on a limited dataset with specific sensors, there are few guarantees that the designed algorithms are going to work on a different set of patients or different sensors. Finally, most of this body of work focuses on a single condition or class of conditions with little to no applicability to other conditions. While useful, this approach is not generalizable to broader communities.

1.3 Purpose of the Study

The purpose of this work is to provide methodologies and tools for a more systematic approach in building RHMS. More precisely, new algorithms will be developed that mine various informational sources, especially Electronic Health Records (EHR), to design more generalizable analytics models for RHMS. In addition, a prototype RHMS system is presented that collects and graphically represents health data from a variety of sources, including traditional RHMS data such as sensors and questionnaires as well as EHR, drug databases, etc. Combining these two sets of advancements provides clin-
icians and researchers with new algorithms to predict patient condition and compliance together with a knowledge-graph platform that can be used to perform complex analytics to extend these algorithms and develop new ones. Ultimately, this dissertation work aims to improve the remote treatment and management of the ever increasing number of patients and conditions.
CHAPTER 2

Review of the Literature

This work relies on advancements in RHMS technology as well EHR mining for clinical care, both of which have been the subject of active research in the past decade. For better comprehension, the relevant research of each will be studied independently.

2.1 Remote Health Monitoring

The history of Remote Health monitoring, otherwise known as telemonitoring, is extensive. One of the first recorded telemonitoring solutions was created in 1905 by the creator of Electro Cardiogram (ECG), Einthoven [13]. This solution involved the transmission of only one signal: the ECG of a patient. Perhaps the first complete telemonitoring solutions were the ones developed by NASA to monitor the vitals of astronauts [69]. Since then, remote health monitoring solutions have evolved significantly. Most recently, the term “RHMS” has been primarily used to describe clinical informational systems that monitor individuals outside of traditional healthcare environments. Monitoring can be done in a variety of locales, such as an individual’s domicile, a nursing home or a workplace during activities of daily living (ADLs).

Today the use of RHMS in healthcare has grown exponentially as more and more healthcare providers opt to complement traditional care with RHMS. One driver behind this rapid adoption is the hospital readmission reduction program which levies financial punishment on hospitals with high readmission rates. Statistics show that nearly 20% of insured patients are readmitted to hospitals within 30 days after discharge, incurring
approximately $17 billion in charges in 2009 ([46, 19]). To address this, a variety of RHMS studies have shown positive outcomes; Antonicelli et al. [9] demonstrated both reduced mortality as well as reduced readmission rates for congestive heart failure patients in RHMS. Morguet et al. [63] report a 50% reduction in hospital admissions (38 versus 77/100 patient years, $P = 0.034$) and 54% reduction in hospital length of stay for CHF patients. Several studies reported that compared to usual care, patients with diabetes and hypertension that receive RHMS resulted in improved glucose control, cholesterol, and blood pressure [91, 100, 1]. For brevity, the reader is referred to the excellent reviews of [67, 60]. A meta-analysis of RHMS found that patients with heart failure that receive RHMS has a 42% reduction in hospitalizations [107], and another study concluded that RHMS can prevent up to 627,000 heart failure-related hospital readmissions each year [113]. Moreover, comparative studies performed on 17,025 patients enrolled in the Veterans Affairs (VA) home monitoring program in 2006 and 2007 show a 25% reduction in bed days of care, 20% reduction in readmissions, and a satisfaction score of 86% [32]. Another VA study on COPD patients using RHMS reported that 71.5% of patients had a reduction in the number of emergency department visits [4].

In the last decade, early successes coupled with advances in sensors have evolved RHMS systems from very basic (phone interaction, written reports) to advanced end-to-end systems [54]. A few representative solutions are presented to demonstrate the evolution of RHMS systems. An early study by Chaudhry et al. [29] required participants to make daily phone calls to an automated telemonitoring system for a period of 6 months. Each call played a prerecorded voice message that consisted of a series of questions about symptoms and weight for which the participants had to provide answers using the keypad on the phone. The responses were then downloaded from the telemonitoring system to an Internet website for daily review by clinicians. Another heart failure study conducted by Soran [94, 95] included an electronic scale and an individualized symptom response system connected to a computer database via a stan-
standard phone line. Patients were instructed to weigh themselves and answer a series of questions daily. Nurses reviewed the transmitted data on a daily basis and immediately contacted patients whenever the data fell out of a healthy range. These and similar RHMS systems are commonly referred to as conventional or threshold-based RHMS. That is, thresholds based on medical expertise are put in place to alert clinicians when physiological data deviate from the set thresholds.

As later studies and systems evolved [20, 54], they employed more sophisticated data collection methodologies as well as data analysis algorithms. In contrast with conventional RHMS, these analytics-based RHMS employ machine learning algorithms to predict the risk of a medical adverse event. It can be concluded from the evidence [57] that analytics-based RHMS work better than threshold-based ones and can help further reduce treatment costs.

Recent research efforts have identified preexisting factors that influence a patient’s adherence [5] and outcome [6]. Additional work [74] has shown that taking into account early warnings leading up to an adverse event can help both identify heterogeneous patient cohorts as well as improve prediction. These and similar achievements are motivators for a new generation of RHMS, one that would be based on population specific characteristics to improve the effectiveness of remote monitoring. However, several challenges must be addressed before this can be achieved:

- Analytics approaches require a substantial effort to collect initial data, spanning months or even years depending on the studied chronic condition.

- Selecting in advance which physiological data to collect as well as the collection frequency is based on guesswork since there is limited data to support the decision.

- More importantly, previous RHMS research literature has focused exclusively on a single class of conditions for a single population. As such, there is limited evidence that these approaches can be generalized to larger, more diverse
populations.

This work aims to address these challenges by relying on richer data sources, especially Electronic Health Records. While there has been limited application of EHR based insights on RHMS systems, there is extensive work on utilizing EHR to improve clinical care.

2.2 Electronic Health Records

An EHR is an electronic version of a patient’s medical history, that is maintained by the provider over time. Such records contain a variety of relevant information, including demographics, patient history, medications, allergies and test results among others. EHR were introduced to address several challenges of patient care:

- Patients with chronic conditions get treated in multiple locations over multiple encounters. Before EHR, sharing a complete patient history between providers was a big challenge. Tests were often duplicated.

- Hand-written or typed reports were prone to errors and ambiguity. This often lead to medical errors.

- Patients had no easy access to their own test results and medical history, which limited their ability to make informed decisions.

- Finally, before EHR, retrospective research studies were really hard to carry out as data had to be manually processed, which took a significant amount of time.

Most countries around the world have either incentivized or legislated the use of EHR in patient care to overcome these issues. This in turn has led to an increasing adoption of EHR and a renewed interest in exploring EHR for improving healthcare. Jensen et al.[47] discussed the current status and challenges in mining EHRs. A review of recent approaches is provided by [61]. Rao et al.[76] described insights from
mining EHR to improve cardiac care. Yao et al. [115] examined EHR mining in the context of drug efficiency / side-effect analysis. Bauer-Mehren et al. [15] created a network from unstructured EHR to examine the efficiency of certain treatments and identify patient cohorts. Roque et al. [80] use EHR to discover disease correlations and map them to biological frameworks. Lasko et al [56] employ unsupervised feature extraction on EHR to infer computational phenotypic patterns from population-scale data. Besides EHR, there exists extensive literature in structuring and exploring other types of medical information, which will be covered in the next section.

2.3 Structured Knowledge Representation in Medicine

A variety of projects have been devoted to modeling medical knowledge. SNOMED CT (Systematized Nomenclature of Medicine–Clinical Terms) is a comprehensive clinical terminology owned, maintained, and distributed by the International Health Terminology Standards Development Organization (IHTSDO), a not-for-profit association in Denmark. The Unified Medical Language System (UMLS) integrates and distributes key terminology, classification and coding standards, and associated resources to promote creation of more effective and inter-operable biomedical information systems and services, including electronic health records. Medical Subject Headings (MeSH) is the National Library of Medicine’s controlled vocabulary thesaurus. It consists of sets of terms in a hierarchical structure that permits searching at various levels of specificity. RxNorm provides normalized names for clinical drugs and links its names to many of the drug vocabularies commonly used in pharmacy management and drug interaction software. The International Classification of Diseases (ICD) is the standard diagnostic tool for epidemiology, health management and clinical purposes. The most recent of these projects, Disease Ontology (DO) aims to provide an open source ontology for the integration of biomedical data that is associated with human diseases. DO will have a formally correct (in the ontology sense), semantically computable structure.
Terms in DO are well defined, using standard references. These terms are linked to well-established, well-adopted terminologies that contain disease and disease-related concepts such as SNOMED, ICD-9 and ICD-10, MeSH, and UMLS.

Conceptual graphs have been used widely in the past in various domains to model entities. Some of the earliest works in this area include [97] and [81]. A thorough review of the evolution of the conceptual graphs found at [8]. Vicknair et al. [105] provide a good comparison between SQL and graph databases. In medical literature, Campbell et al. [24] and Volot et al. [106] utilized graphs to develop a semantic representation of medical terms, conditions, symptoms and their interactions. Baud et al. [14] developed an innovative natural language processing system using conceptual graphs applied to the domain of digestive surgery discharge summaries. Bell et al. [16] built a semantic network for radiologic records. Ellis et al. [37] Provides a good overview of these early efforts in the medical domain.

The initial interest in graphs as a knowledge representation began to subside around the year 2000 perhaps due to limitations in technology. In the past few years there has been a re-emergence of interest in storing and managing graph data. In academia and research, we see many new attempts at providing a database model for large graph data, particularly social graphs and the Web graph. An increasing number of commercial applications are looking towards graph databases for their dynamic schema and ease of use in storing more complex data. Freebase [108] was conceived as a crowd-sourced knowledge graph to structure human knowledge. Its popularity instigated several related projects such as the Google Knowledge Graph or the Facebook Open Graph. In the recent years, there have also been several initiatives for making health data more openly available such as openFDA. Researchers and institutions alike are also at looking at Big Data solutions [65] to model and explore the newly available wealth of information.
CHAPTER 3

Building Severity of Condition Prediction Models from
Electronic Health Records for Remote Health Monitoring Patients

3.1 Background

3.1.1 Electronic Health Records

We used electronic health records (EHR) from the Ronald Reagan UCLA Medical Center between 2005 and 2009. We first identified CHF hospitalizations for adult patients using the same International Classification of Diseases version 9 (ICD-9) codes used in the Centers for Medicare & Medicaid Services (CMS) 30-day readmission measure. A total of 1179 admissions were extracted with primary diagnosis a CHF related ICD-9 code. These records correspond to 913 unique patients out of which 169 had more than one CHF related admission.

3.1.2 Remote Health Monitoring System (RHMS) Clinical Trial

We conducted a RHMS-based study called BEAT-HF, one of the largest randomized controlled trials of telemonitoring in patients with heart failure [20]. BEAT-HF targeted the remote monitoring of 1500 patients 50 years of age or older with heart failure problems, 700 of which used a RHMS and the remaining received usual care. The primary purpose of the BEAT-HF study was to compare the effect of implementing wireless re-
mote monitoring combined with structured telephone monitoring versus conventional care on variation in hospital readmission. The study population included patients with a wide range of demographic and socioeconomic characteristics.

This study was conducted in collaboration with the UCLA Department of Medicine, UC-Davis, UCSF, UCI, UCSD, and Cedar Sinai Hospital. Heart failure patients who were hospitalized at any of the six participating medical centers were recruited in either the control or intervention arm through a randomized trial process. Data collection for this study started in November 2011 and ended December 2013. In this study, 700 patients with heart failure are given wireless devices (such as a wireless weight scale and wireless blood pressure cuff) to measure weight, systolic and diastolic blood pressure and heart rate on a daily basis. These devices also collected self-reported questionnaires regarding heart failure symptoms. Then, the collected data was transmitted via an information gateway such as a patient’s smartphone. The gateway then transmitted the data to a secure remote database via a phone-line, an Internet connection at home, or through cellular networks. Patients measured their weight and blood pressure and replied to questionnaires on a daily basis.

3.2 Methodology

Previous work [93] has shown that first day vitals can be used to predict the severity of a CHF patient’s condition by estimating the length of stay (LOS) in the hospital. We developed an EHR-based risk model that can be used to predict severity of condition. In this work, we evaluated the effectiveness of our EHR-based risk model on patients enrolled in the BEAT-HF study. We also extended this methodology to a similar insight for the last day of hospital stay. Since patients are released when their condition is deemed stable, their vitals during the last day should contain some information about how soon they get readmitted [35].
3.2.1 EHR Data Processing

For each admission record we extracted seven features: min, max, range, average, median, standard deviation and variance, computed from the patient’s heart rate, systolic/diastolic blood pressure, and weight during the first day of hospital stay. Furthermore, 19 categorical features are calculated based on the patient’s age, gender, race, ethnicity, ZIP code, type of external care, insurance coverage, perceived severity of condition, perceived mortality and the first ten comorbidities reported.

We rank these features based on their correlation with a patient’s length of stay in the hospital. We employ correlation-based feature subset selection (CFS) [41]. CFS is a heuristic algorithm that identifies feature subsets which contain features highly correlated with the class, but not with each other. The heuristic can be formalized as follows:

\[
\text{Merit}_S = \frac{k \overline{r_{cf}}}{\sqrt{k + k(k - 1) \overline{r_{ff}}}}
\]

where merit is the “merit” of a specific subset \( S \) with \( k \) features, \( \overline{r_{cf}} \) is the mean feature correlation (for features in \( S \)) and \( \overline{r_{ff}} \) the mean feature intercorrelation. Evaluating the available features against LOS using equation (3.1), we conclude that the most prominent categorical and numerical features are age, gender, ethnicity, minimum systolic pressure and heart rate range for the first day (See Table 3.1).

3.2.2 Predicting Length of Stay in EHR

A reliable indicator of a patient’s severity of condition is the Length of Stay (LOS) in the hospital. As multiple previous studies have indicated [28, 59, 79] contextual information such as age, gender and comorbidities can affect the LOS of a patient along with the actual condition. To predict the LOS of a patient, we use the most correlated categorical features as well as the most correlated vitals that correspond to the first day. Using generalized linear regression, we combine the effects of a patient’s vitals with the effects of his contextual information to predict LOS. Table 3.2 presents the average
Table 3.1: Attributes as ranked by the correlation-based feature subset selection

<table>
<thead>
<tr>
<th></th>
<th>Attribute</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>age</td>
</tr>
<tr>
<td>2</td>
<td>gender</td>
</tr>
<tr>
<td>3</td>
<td>ethnicity</td>
</tr>
<tr>
<td>4</td>
<td>min. systolic pressure first day</td>
</tr>
<tr>
<td>5</td>
<td>range of heart rate first day</td>
</tr>
<tr>
<td>6</td>
<td>std systolic pressure first day</td>
</tr>
<tr>
<td>7</td>
<td>std diastolic pressure first day</td>
</tr>
<tr>
<td>8</td>
<td>perceived severity</td>
</tr>
<tr>
<td>9</td>
<td>perceived mortality</td>
</tr>
<tr>
<td>10</td>
<td>insurance</td>
</tr>
<tr>
<td>11</td>
<td>type of outside care</td>
</tr>
<tr>
<td>12</td>
<td>2nd co-morbidity</td>
</tr>
<tr>
<td>13</td>
<td>4th co-morbidity</td>
</tr>
<tr>
<td>14</td>
<td>5th co-morbidity</td>
</tr>
<tr>
<td>15</td>
<td>8th co-morbidity</td>
</tr>
<tr>
<td>LOS Considered</td>
<td>RMSE (days)</td>
</tr>
<tr>
<td>---------------</td>
<td>-------------</td>
</tr>
<tr>
<td>≤ 15 days</td>
<td>3.2788</td>
</tr>
<tr>
<td>≤ 30 days</td>
<td>4.9527</td>
</tr>
<tr>
<td>≤ 60 days</td>
<td>8.2770</td>
</tr>
</tbody>
</table>

Table 3.2: Average root mean square error of generalized linear regression with 10-fold cross validation.

Root mean square error (RMSE) (equation (3.2)) using 10-fold cross validation and including only patients that stayed less than 15 days, 30 days, and 60 days respectively. It can be seen that the RMSE increases as we try to predict longer stays. This could be due to the lack of data for these longer stays, since the vast majority of patients stay less than 15 days. It could also be due to the fact that for longer stays, we need intermediate information from a patient’s hospital stay.

\[
RMSE = \sqrt{\frac{\sum_{i=1}^{N} (predicted_i - observed_i)^2}{N}}
\]  

(3.2)

3.2.3 EHR-based Risk Model for First Readmission

In the previous section we showed that it is possible to predict a patient’s length of stay based on contextual information and first day vitals. This prediction model can be useful when evaluating a patient on a RHMS. It provides nursing staff with an estimate of how long that patient would stay in the hospital if he/she were to be admitted that day. To simplify the decision process for the support staff, we design a clustering algorithm that groups patients in terms of their individual risk based on the top two vital features (minimum systolic pressure, heart rate range for the first day). In this study, we define LOS plus 3 times ICU stay as a "risk factor" [93] (Eq. (3.3)). This weighting scheme is based on the observation that the daily cost of ICU is approximately 3 times greater than that of normal hospital stay [53, 83]. In other words, our risk factor is analogous
to the projected cost of hospitalization.

\[
\text{risk factor} = \text{LOS} + 3 \times \text{ICU} \tag{3.3}
\]

We cluster the admission records based on hierarchical clustering [50]. Through experimentation, we found that using Chebyshev’s distance function [27] and complete linkage for the cost function [34] results in the most stable clusters. Figure 3.1 demonstrates the results of the hierarchical clustering using 8 clusters. The number of clusters was chosen empirically as it appears to be more stable with the number of subjects considered. We rank and color the clusters based on the average risk from low to high. Red corresponds to the highest and blue/green to the lowest risk.

As it can be seen in Figure 3.1, there are some issues with our risk factor as the rankings do not follow medical expertise. For example, people with really high systolic blood pressure appear to have shorter duration of stay at the hospital than people with normal levels. The main reason for these discrepancies is that there is a strong age bias on how long a patient stays in the hospital as previous studies have shown [28]. For example, Figure 3.2 shows that CHF patients between the ages of 55 and 65 years old for example end up staying in the hospital longer on average than older or younger patients.

To amend these issues and obtain a more objective risk grouping regardless of age/gender, we normalize the risk factor of each patient by multiplying it with a value inversely proportional to the average risk factor of his age and gender group (See Eq. (3.4)).

\[
\text{NormalizedRisk}_i = \frac{\text{Risk}_i}{\sum_{x \in C(i)} \frac{(\text{Risk}_x)}{\|C(i)\|}} \tag{3.4}
\]

where \(\text{Risk}_i\) is the risk of individual \(x\) according to Eq (3.3), \(C\) is the clustering function according to age and gender. After risk normalization, and with the same clustering methodology, we obtain a risk ranking that is closely aligned with medical
Figure 3.1: Hierarchical Clustering of patients according to their minimum systolic pressure, heart rate range during the first day of hospitalization. Colors represent average risk per group, with blue/green and red representing the lowest and highest risk respectively.
Figure 3.2: Age versus Length of Stay for CHF patients. It can be seen that there is an age bias on length of hospitalization. The LOS values are scaled between 0 and 1.
Figure 3.3: Hierarchical Clustering of patients with the risk factor normalized with regards to Age/Gender.

The above clustering suggests that it is possible to rank a patient’s severity of condition based on systolic pressure and heart rate measurements. This information can be useful in the context of a RHMS for clinicians. Threshold alerts such as “systolic pressure above 130 millimeters of mercury” end up overwhelming the clinicians as they tend to accumulate very fast with a growing number of patients. Our clustering methodology provides a much more detailed status for each patient as it allows ranking severity and making more informed decisions. In addition, such a scheme could be employed to suggest short term health improvements for the patients.
3.2.4 Applying First Admission Risk Model in BEAT- Clinical Trial Data

We demonstrate the effectiveness of the proposed risk modeling technique on data gathered on the BEAT-HF study of 700 heart failure patients. In that study heart rate, systolic/diastolic blood pressure, and weight data was collected daily from a Remote Health Monitoring System [54] over 12 months. We extracted the same seven features as in the EHR data over seven-day windows: min, max, range, average, median, standard deviation and variance, computed from the patients’ heart rate, systolic/diastolic blood pressure, and weight. We used a seven-day window on the BEAT-HF data to ensure that the patients had successfully transmitted at least one of each measurement in the seven days. We excluded any sample points for which there were no RHMS physiological data. Patients were instructed to take those measurements at the same time of the day before any meals or significant activity. This way, variability between measurements is primarily attributed to actual health variations.

For each of the seven-day data points we assign a readmission boolean value. This value denotes whether a patient was readmitted in a window of days following the seven-day measurement window. We examine readmissions in the subsequent 3, 10 and 30 days. For each seven-day BEAT-HF window, patients were assigned to the nearest cluster computed from the EHR risk model in Figure 3.3 to generate the predicted first readmission risk. To measure how well our prediction model works we calculate the percentage of actual 3-day, 10-day and 30-day BEAT-HF readmissions per cluster and color each cluster according to the overall rank to generate the observed risk. We then calculate the root mean square error between the predicted first readmission risk and the observed risk for the 3-day, 10-day and 30-day readmissions. The results are shown in Section 3.3.1.
3.2.5 EHR-based Risk Model for Readmission

Once we are able to provide risk models that predict a hospital admission from physiological data we can prioritize which patients need more care and attention. However, once patients enter the hospital we need a risk model that attempts to predict those that are readmitted. There are a few risk models that have been designed based on last day vitals [35]. From our EHR data we are able to calculate the days between two subsequent admissions. We explore this information to develop risk models similar to the prior risk prediction model, but based on days to readmission. We define the readmission risk to be inversely proportional to the number of days it took for the patient to be readmitted:

\[ \text{risk factor} = \frac{1}{\text{days} - \text{to} - \text{readmission}} \quad (3.5) \]

This risk measure was chosen as it is in the range (0,1] and reflects the likelihood of getting readmitted soon. In developing the risk model we only take into account the patients that were readmitted at least once, since otherwise we cannot determine the \( \text{days} - \text{to} - \text{readmission} \). To develop the long term risk model we used a similar clustering approach as with the first readmission measure. Based on the new risk factor, we rank the features according to their correlation with a patient’s number of days to readmission. Using correlation-based feature subset selection [41], we find that the top ranked categorical features are gender, race, and ethnicity and the top ranked last day vitals are the mean weight and mean systolic blood pressure. As in the previous section we normalize the risk against those three top ranked categorical features and cluster subjects based on mean weight and mean systolic blood pressure. Figure 3.4 provides the normalized risk model.
Figure 3.4: Readmission risk clusters from EHR data normalized by gender, race and ethnicity.
3.2.6 Applying Readmission Risk Model in BEAT-HF data

We applied the EHR readmission risk models to the BEAT-HF data, where we assigned the BEAT-HF test samples to the nearest cluster, computed from the readmission EHR risk model. We then show the actual percentage of a BEAT-HF 3-day, 10-day and 30-day readmission and color each cluster according to the overall rank to generate the observed risk. A green color signifies low readmission percentages and red patients with the highest readmission risk. We then calculate the root mean square error between the predicted readmission risk and the observed risk for the 3-day, 10-day and 30-day windows. The results are provided in Section 3.3.2.

3.3 Results

3.3.1 Predicting First Readmission in BEAT-HF

In Figure 3.5 we see the predicted first readmission risk based on the EHR short term risk model. To visualize the predictive power of the risk model, we calculate the percentage of readmissions that occurred after 3-days for each cluster and color each cluster according to its overall rank to generate a 3-day observed risk, the result is shown in Figure 3.6. We then do the same for the 10-day and 30-day readmissions, and the results are shown in Figure 3.7 and Figure 3.8, respectively.

To evaluate the accuracy of the BEAT-HF predicted risk, we calculate the root mean square error between the predicted risk and the 3-day, 10-day and 30-day readmission. The results are illustrated in Figure 3.9. It can be seen that the predicted risk achieves excellent RMSE scores for all predictions, with a 0.866 RMSE score for 3-day readmission predictions, a 2.12 RMSE for 10 day predictions and a 1.94 RMSE for 30-day predictions.

As it can be seen in figures 3.5-3.9, the EHR based risk clustering provides an accurate prediction of readmission risk for patients being monitored remotely. While
Figure 3.5: Predicted Risk

Figure 3.6: 3-day Readmission Rate

Figure 3.7: 10-day Readmission Rate

Figure 3.8: 30-day Readmission Rate
The collection frequency is different in the case of the RHMS and hospital data, the features in both cases are calculated over 24-hour intervals.

### 3.3.2 Predicting Readmission in BEAT-HF

In Figure 3.10 we see the predicted readmission risk based on the EHR long term risk model. Similar to the short term risk model, we calculate the percentage of readmissions that occurred after 3-days, 10-days and 30-days and color each cluster according to its overall rank, the results are illustrated in Figures 3.11-3.13.

The root mean square error between the predicted risk and the 3-day, 10-day and 30-day readmission are also shown in Figure 3.14. It can be seen that the predicted risk for readmissions achieves excellent RMSE scores close to 1.0 for all predictions, with a 1.0 RMSE score for 3-day readmission predictions, a 1.29 RMSE for both 10-day and 30-day readmissions.
Figure 3.10: Predicted Risk
Figure 3.11: 3-day Readmission Rate
Figure 3.12: 10-day Readmission Rate
Figure 3.13: 10-day Readmission Rate
3.4 Conclusions

Using statistical models constructed from electronic health records we have shown that it is possible to predict the length of stay of a CHF patient with high accuracy. We have also presented a short term risk model for modeling patient groups in terms of severity of condition based solely on hospital records. We have tested this risk model on our RHMS Beat-HF dataset. We show the effectiveness of the risk model to predict readmissions in BEAT-HF. We then develop a long term risk model that predicts a readmission based on electronic health records. Again we show the efficacy of this model to predict a readmission in the BEAT-HF data set. These models can be used to analyze daily information collected from an RHMS and allow a nurse / doctor to prioritize his/her intervention. The risk factor information could also be used to provide personalized advice to the patient on how to improve their condition in the short and long term as to avoid readmission. In the future we plan to further validate our approach.
by testing its predictive abilities on other RHMS datasets on congestive heart failure. We also plan to extend our approach to different chronic conditions such as diabetes and liver disease.
CHAPTER 4

Multiple Model Analytics for Adverse Event Prediction in Remote Health Monitoring Systems

4.1 Background

Given the diversity of patient populations and the continuous nature of patient monitoring, a single static model is inefficient for accurately predicting medical adverse events. Consequently, an effective model is more likely to be the one that includes multiple prediction models in a hierarchical architecture. Our results show that clustering patients based on their baseline information, and performing per-cluster prediction achieves significant improvement in the accuracy of prediction compared to applying a single universal prediction model. Figure 4.1 demonstrates the block diagram of the proposed multiple prediction modeling. In this study, we applied the K-Means clustering algorithm to split the data samples into 8 clusters in a two-dimensional feature space based on baseline weight and systolic blood pressure. Figure 4.2 shows a sample dataset clustered in this domain.

4.2 Methodology

A subset of data collected through our remote health monitoring system involving 600 patients with Heart Failure (HF) was used for this analysis. The dataset contains physiological measurements including weight, systolic and diastolic blood pressure, and heart rate collected on a daily basis from 600 patients with heart failure. Our goal was to clus-
In this paper, we tried both wrapper and filter methods; the former utilizes a classifier (or any other learning method of interest) to evaluate feature subsets in an iterative manner according to their predictive power. A new feature selection method not only decreases the computational complexity of the system by reducing the dimension of each feature with the goal of finding features that provide the maximum amount of useful information for prediction. Thus, selecting a feature subset includes a set of accurate prediction models rather than one single universal predictor for the entire dataset.

As shown in Figure 1, after clustering the dataset into 8 clusters, these features will be used in the classification of subsets of features evaluated on a testing dataset to assess the relative usefulness of subsets of features. Applying an effective feature selection algorithm not only decreases the computational complexity of feature space splitting but also increases the performance of the classifier by removing redundancy, but also increases the complexity of the system by reducing the dimension of features in a given feature set. Applying an effective feature selection algorithm in the filter feature selection method.

Figure 2 illustrates the steps involved in the process of selecting a feature subset. In the first stage, the proposed clustering algorithm is applied to the dataset to split the data samples into different clusters based on their baseline weight and systolic blood pressure. Figure 3 shows the diagram of the steps involved in the process of selecting a feature subset...
Our results show that clustering patients based on their baseline information, and performing prediction achieves significant improvement in prediction rather than using a single model to perform the prediction for the entire dataset. To address this issue, we propose a method that includes a set of accurate prediction models rather than one single universal predictor for the entire dataset.

As shown in Figure 1, after clustering the data samples, we applied multiple prediction modeling. Each patient is assigned to the cluster that contains the closest mean value based on their baseline weight and systolic blood pressure. Consequently, the physiological data measurements and vital signs may differ from patient to patient. Thus, in similar health risk conditions, the symptoms (and signs) may differ from patient to patient.

In this paper, we tried both wrapper and filter methods to select the most prominent features for each cluster and subsequently apply group prediction model.

Figure 4.2: Clustering the data samples based on weight and blood pressure.

Figure 1 demonstrates the block diagram of the steps involved in the proposed framework. The main goal of feature selection is to eliminate redundant or irrelevant features that provide the maximum amount of useful information for prediction. Thus, specific metrics to quantify the relevance and/or redundancy of each feature with the goal of finding the smallest subset of features that provides the maximum amount of useful information for prediction. Thus, our results show that clustering patients based on their baseline information, and performing prediction based on accurate prediction models that includes a set of accurate prediction models rather than one single universal predictor for the entire dataset.
ter patients and perform a per-cluster adverse event prediction analysis to compare the accuracy of a single prediction model with that of a multiple model. After feature extraction, we split the dataset into 8 clusters using a K-Means clustering algorithm, and then applied individual feature selection and classification on each cluster. The clustering was performed on a two dimensional feature space including weight and blood pressure measurements. We used the random forest classifier [21] and 10-fold cross validation over the data clusters and also the entire dataset to evaluate the performance.

4.3 Results & Conclusions

Figure 4.3 shows the receiver operation characteristic (ROC) curves on a true positive rate (TPR) vs. false positive rate (FPR) plane. In this figure, the red dashed curve represents the ROC of a single model prediction trained on the entire dataset. Other curves demonstrate the classification results for each cluster using cluster-specific feature sets and individual per-cluster classifiers. As shown in Figure 4.3, clustering and then per-cluster feature selection and classification is an effective approach to multiple modelling, and can significantly improve the performance and accuracy of the adverse event prediction.
Figure 4.3: ROC curves for single model and cluster-specific multiple model.
CHAPTER 5

Effects of Coaching on Adherence in Remote Health Monitoring Systems

Among all of the major organizations, including the World Health Organization, the Centers for Disease Control and the Pew report the focus on disease prevention is critical. Given the rapid advances in technology it has become clear that there is a critical role for remote health monitoring systems (RHMS) in the prevention of chronic disease. The number of interdisciplinary clinical trials has increased over the past few years. In preventive medicine RHMS are designed to reinforce patient education, monitor and insure participant adherence, and provide a means of communicating information to the clinician. In this work we examined data collected from a smartphone-based RHMS intervention for young Black women at risk for cardiovascular disease. The goal of the intervention was to improve cardiovascular risk factors through classroom-instituted self-management education on risk factor reduction in an attempt to change behavior over time. To augment the classroom education we created a means of remediation and social support through wireless monitoring and coaching creating a sense of “connectedness”. This chapter reports the effects of RHMS automated messages, clinician text messages and clinician phone calls in improving adherence to the study protocol. Typically participants with low adherence maintain low adherence, and participants with a high level of adherence maintain a high level of adherence. We examined the effects of messaging on adherence from one week to another and show that prior week adherence is a valuable predictor of the following week’s adherence level. Our findings show that RHMS messages are successful in motivating the group of participants that
are inconsistently adhering to the study regimen. We also show that it is possible to predict weekly participant adherence with high accuracy.

5.1 Introduction

Despite efforts in health care and prevention, Cardiovascular Disease (CVD) remains the leading cause of death in the United States. Statistics from the American Heart Association show that more than 2,150 Americans die each day of CVD [64]. In 2009 CVD was responsible for 31.3% of the total deaths in the U.S. This translates to an average of 1 death every 40 seconds. A similar situation exists in Europe as well, where recent statistics in the European Heart Journal [68] attribute almost 4.1 million deaths per year or 46% of all deaths to CVD.

The reasons for the high prevalence of CVD can be traced to the increased proportion of the population that have one or more risk factors for CVD at a very young age. An estimated 31 million US adults exhibit cholesterol levels above 240mg/dL. Additionally, 32.6% of US adults above 20 years of age have with hypertension. Finally, 10% of US adults are affected by Diabetes Mellitus and an alarming 35.3% have pre-diabetes: a condition characterized by abnormal fasting glucose levels. There is a high propensity of hypertension and diabetes in the Black community and a higher rate of morbidity and mortality from CVD.

The Women’s Heart Health study presented in this work focuses on Black women, a particularly vulnerable part of the population. In fact, CVD related deaths were 32.1% higher for Black females than white women (248.6 deaths per 100,000 vs 188.1). A remote health monitoring system (RHMS) Wanda-CVD, was deployed to monitor the effects of education and social support from automated and clinician messages and coaching on reducing risk factors for CVD. In this work, we focus on the intervention group and evaluate the relationship between automated/clinician coaching and distal effects of adherence to the study protocol from one week to another. We further design
a framework to predict weekly participant adherence to optimize clinician intervention by informing them about participant future adherence. We hope to further validate our systems predictive power and the proximal effect of our treatment through micro-randomization of messages in future trials [58].

This chapter is organized as follows. Section 2 discusses background and related works. Section 3 discusses the Women’s Heart Health Study and the deployment of automated and clinician messages in Wanda-CVD. Section 4 describes the methodology used in analyzing the collected data. Section 5 provides the results and discussion. Finally, we conclude and discuss future work in Section 6.

5.2 Background and Related Works

While CVD is the number one cause of death globally, research has shown us that lifestyle changes are the key to reducing risk factors for CVD. Increased physical activity is among the top three factors for CVD risk reduction along with eliminating exposure to tobacco and maintaining healthy eating habits [33]. Class 1 recommendation guidelines for stroke prevention focus on diet and exercise especially in participants at high risk for stroke. Studies have shown that black women have a particularly high risk of stroke [22]. Intervention trials that focus on behavior change have initially been successful however have not stood the tests of time. This is based on the premise that a mobile phone application that focuses on coaching and motivation, along with self-monitoring and an opportunity to link social support might improve the user’s self-monitoring experience and adherence to a physical activity program [7].

To address this, a variety of RHMS studies have shown positive outcomes. Antonicelli et al.[9] demonstrated both reduced mortality as well as reduced readmission rates for congestive heart failure patients in RHMS. Morguet et al.[63] report a 50% reduction in hospital admissions (38 versus 77/100 patient years, P = 0.034) and 54% reduction in hospital length of stay for congestive heart failure patients. Several
studies reported that compared to usual care, patients with diabetes and hypertension that receive RHMS resulted in improved glucose control, cholesterol, and blood pressure [91, 100, 1, 67, 60]. Two meta-analyses [107, 52] of RHMS found that patients with heart failure that receive RHMS have a 7% and 42% reduction in hospitalizations respectively. Moreover, comparative studies performed on 17,025 patients enrolled in the Veterans Affairs (VA) home monitoring program in 2006 and 2007 show a 25% reduction in bed days of care, 20% reduction in readmissions, and a satisfaction score of 86% [32].

Mobile health approaches are also found to be useful in several applications including child obesity prevention [102]. Spring et al. show that remote coaching supported by mobile technology and financial incentives significantly changed behavior, increasing daily fruit/vegetable intake and decreasing sedentary leisure and saturated fat [98]. In our study we would like to show the impacts of remote coaching on mobile technology and report correlations to adherence of the study protocol.

5.3 Study & RHMS design

5.3.1 Wanda-CVD RHMS

To carry out the study, we relied on an improved version of a previously developed RHM system named Wanda [54]. We expanded the original RHMS to incorporate an increased sense of “connectedness” between participants and clinicians by introducing Wireless Coaching. Wireless Coaching in Wanda comes in the form of automated messages (push messages) reminding the participant to commit to certain actions related to the goal of the study and clinician messages (either through text messages of phone calls).

The complete Wanda-CVD system is outlined in Figure 5.1. Data from participants are collected on an Android-based smartphone application. Using embedded sensors,
Bluetooth and Wi-Fi/Cellular network technology, the smartphone application can be programmed to connect to many stand-alone patient monitoring systems. The application also receives and displays Wireless Coaching push messages. This information is transmitted securely to a backend server, where it is stored and machine learning algorithms process the data to identify patterns and learn patient models. The final component is a web-based portal that provides clinicians with a visual cue and summary of what is happening with each participant, alerting them when a matter requires their attention.

Figure 5.1: Wanda-CVD architecture diagram
5.3.2 Women’s Heart Health study

The Women’s Heart Health Study is a UCLA IRB-approved study of 90 young black women aged 25-45 years that have a minimum of two risk factors for CVD. Randomization occurred at the beginning of the study and the participants were randomized by church upon entry into the study. Wanda-CVD was deployed to analyze the effects and lifestyle changes that result from social support via wireless coaching. In addition to the data collected from Wanda-CVD during the study, the participants also completed baseline screening of cholesterol levels, blood pressure, BMI, demographic and psychosocial questionnaires and received educational classes. Each of the intervention group participants received nutrition and lifestyle education, along with a Bluetooth blood pressure monitor and a smartphone with Wanda-CVD pre-loaded. Clinicians instructed the participants to wear the smartphone around their waist to detect their level of physical activity throughout the day. The participants were monitored for a period of 6 months and were requested to respond to daily and weekly questionnaires. Through the smartphone, participants received 12 daily questions spread out over 6 days (weekdays and Saturday). They also received 12 weekly questions (answered on a single day, typically a Sunday). Finally, they were asked to measure their systolic and diastolic blood pressure twice in one sitting per week. The intervention group also received wireless coaching messages, and in this effort we attempt to study the weekly distal effects of messaging on study protocol adherence.

5.3.3 Automated Wireless Coaching Messages

There are four categories of automated wireless coaching messages: inspirational, blood pressure, daily questionnaire and weekly questionnaire related messages. One out of seventeen inspirational messages is pushed to a participant every other day as a means of reinforcing the education received during baseline. The system also analyzes the participants’ adherence to the study protocol and accordingly pushes targeted coach-
ing messages designed to improve adherence. If the user forgets to take a measurement on a particular day the system would send them a corresponding daily questionnaire, weekly questionnaire or blood pressure reminder the following day. Tables 5.1 and 5.2 provide a list of the automated messages participants in the intervention group received over the course of the study.

5.3.4 Clinician Messages

In addition to the automated coaching provided by Wanda-CVD, a clinician specialized in cardiovascular disease and a kinesiologist met weekly to review the progress of each participants’ blood pressure, physical activity level, and adherence to the daily and weekly questionnaires. Following the weekly review, clinicians selectively contacted participants by phone or text messages and provided personalized education to encourage them to adhere to the study protocol (Clinician Adherence), or discuss their BP measurement (Clinician B.P.), or provide a means of positive reinforcement if all is going well (Clinician Positive Reinforcement). Table 5.3 displays the three different categories of clinician messages.

5.4 Methodologies

The objective of this work is to study the effects of clinician-participant coaching in improving participant adherence. In this section we first define how we quantify adherence and how we process the messages. Then we discuss the data analysis techniques used to generate the results in the following section.

5.4.1 Quantifying Adherence

For physical activity, we consider the number of days a participant was active above a certain threshold ($ \geq 1000 $ points, which is equivalent to 21 minutes of moderate (4
Table 5.1: Automated Messages - Inspirational & Blood Pressure Reminders

<table>
<thead>
<tr>
<th>Inspirational Messages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breathe in through your nose and become aware of your breathing. Breathe easy.</td>
</tr>
<tr>
<td>Don’t allow anyone to pollute the air your breath.</td>
</tr>
<tr>
<td>Exercise increases your energy level - take a light walk around the neighborhood.</td>
</tr>
<tr>
<td>Reduce your stress by laughing or just relaxing - tension and anxiety is lowered.</td>
</tr>
<tr>
<td>Try a new fruit or vegetable today.</td>
</tr>
<tr>
<td>Take a walk with a friend today - light walking is considered a light-leveled activity.</td>
</tr>
<tr>
<td>Go outside and enjoy the fresh air for a few moments.</td>
</tr>
<tr>
<td>Even a small amount of physical activity is better than none.</td>
</tr>
<tr>
<td>Small changes over time = :)</td>
</tr>
<tr>
<td>Deeply relax all your muscles, beginning at your feet and going all the way to your face</td>
</tr>
<tr>
<td>Increasing your physical activity level strengthens the heart and the lungs.</td>
</tr>
<tr>
<td>Take some time for yourself - relieve stress and take a nice walk.</td>
</tr>
<tr>
<td>Keep going, you can do it - give yourself a pep talk.</td>
</tr>
<tr>
<td>Every day you increase your activity you help your body burn fat.</td>
</tr>
<tr>
<td>Be active, be a role model for your children/family!</td>
</tr>
<tr>
<td>Put a little pep in your step!</td>
</tr>
<tr>
<td>Take the steps today and reduce your risk for heart disease.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Blood Pressure Reminders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monitoring your blood pressure helps you to maintain heart healthy habits.</td>
</tr>
<tr>
<td>High blood pressure is a risk factor for heart disease.</td>
</tr>
<tr>
<td>Remember to take your blood pressure today.</td>
</tr>
<tr>
<td>Have you forgotten to take your blood pressure today?</td>
</tr>
</tbody>
</table>
### Table 5.2: Automated Messages - Questionnaire Reminders

<table>
<thead>
<tr>
<th>Daily Questionnaire Reminders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-care is the task of taking care of your health every day.</td>
</tr>
<tr>
<td>Self-awareness starts with you. Remember to take the daily questionnaires.</td>
</tr>
<tr>
<td>Have you been completing the daily questionnaires? Healthy living starts with education and learning.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Weekly Questionnaire Reminders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-care that focuses on heart health is a powerful way to decrease your risk for heart disease.</td>
</tr>
<tr>
<td>Keep yourself healthy = keeping your families healthier</td>
</tr>
<tr>
<td>People who take care of themselves are better equipped to take care of others. Remember to complete your questions.</td>
</tr>
</tbody>
</table>

### Table 5.3: Clinician Messages

<table>
<thead>
<tr>
<th>Clinician Adherence:</th>
<th>A text or call was made to participants that did not fully adhere to the study requirements the week before.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinician B.P:</td>
<td>Clinicians made a text or call to participants if there was a deviation from typical participant blood pressure values.</td>
</tr>
<tr>
<td>Clinician Positive Reinforcement:</td>
<td>A text or call was made to participants if they were adhering and had good levels of physical activity.</td>
</tr>
</tbody>
</table>
We define an active day based on the 2008 Physical Activity Guidelines for Americans, which sets the minimum requirement of physical activity to 150 minutes, which is about 21 minutes daily. As a result, we define an active day to be at-least 21 minutes of moderate activity. Since participants often missed a few days a week wearing the activity device daily, we could not measure accurately whether the participant truly performed 150 minutes of activity throughout the week, as a result we considered the number of active days in a week as a measure of adherence. The daily questionnaires acted as prompts, reminders and reinforcement of their education. Since participants could forget a day or two to answer the daily questions, we considered participants to be adherent in this category if they completed at-least 50% of the questions, which includes 3 out of the 6 daily questionnaires a week. Participants adhere in blood pressure if we receive at-least one blood pressure measurement a week. The weekly questionnaire was completed as a single unit once a week, as a result they were considered adherent if they completed all the questions in the weekly questionnaire.

5.4.2 Data Analysis

From the final dataset of the intervention group we removed 16 participants (8 completed baseline but did not return to the study, and 8 did not have enough reliable data for prediction). Overall, we end up with a dataset of 925 valid weekly datapoints totaling 25 weeks from 37 participants. We only use 25 instead of the 26 weeks, because we do not have a priori information to predict adherence in week one.

We first attempt to determine whether each type of automated or clinician message correlates with adherence or non-adherence to study protocols on a weekly basis. We then attempt to study, using machine learning algorithms, how well the data can help predict adherence in four categories: active days, daily questionnaire, weekly questionnaire, and blood pressure measurement.
5.4.2.1 Analyzing Correlation of Messages

Firstly, we would like to explore how clinician and automated messages correlate with the following week’s adherence. We provide statistical tests of independence for each of the seven intervention types defined in Table 1 (4 automated messages and 3 clinician messages) and the adherence outcome of the four categories: daily questionnaire, weekly questionnaire, blood pressure measurements, and active days. In each category we split the weeks for each participant into those that adhered and those that did not adhere. We then compare the participants that received a clinician or automated message with those that do not, and analyze whether they remain adherent or not the following week. We do this by testing the hypothesis that the two distributions have equal means using a Welch t-test. The Welch test will give us an idea of whether interventions had a positive or negative impact on participant adherence, especially given unequal sample sizes (which is evident in our study). Due to the limited number of participants in this study, participants that did not adhere automatically received automated reminders, however in a larger future trial we will attempt to study the effects of micro-randomization of messages among participants.

5.4.2.2 Predicting Weekly Participant Adherence

We explore whether it is possible to predict weekly adherence using collected data from the previous week as well as baseline contextual information. Such a system is useful in optimizing clinician intervention and minimizing the associated costs by allowing clinicians to target the participants that most probably will not adhere and as a result need more support. We train predictive models for Active Days, Blood Pressure, Daily and Weekly Questionnaires. Our goal is to categorize patients into those that will adhere and those that will not for each category. We use the full dataset (925 datapoints) and combine weekly adherence features (seven categories of messages) with baseline contextual features to predict adherence levels the following week.
The predictive pipeline consists of feature selection using the information gain criterion [51] and four classification algorithms (Figure 5.2). We compare classification models using four well-known algorithms: C4.5, Random Forests, Bayesian Networks and Logistic Regression. For each of these classifiers, we select commonly used training parameters (default parameters in Weka) as fine-tuning the classifiers is beyond the scope of this work.

5.5 Results and Discussion

5.5.1 Analyzing Correlation of Messages

The results of the two-tailed t-tests (Welch) are shown in Table 5.4. We only show the statistically significant ($p < 0.01$) results that had a positive impact. The 95% percent confidence interval shows the difference in next week adherence mean values between the participants that received the intervention (With) and those that did not (Without). The difference will be negative (showing a positive effect) when those that did not receive the intervention have a lower mean than those that did.

Overall, clinician positive reinforcement texts and calls resulted in the most significant effect across all categories of adherence. For the Active Days category, those that did not adhere and received positive reinforcement had between 0.24 and 3.26 more active days the following week than those that did not receive positive reinforcement ($p = 0.0268$). Clinician positive reinforcement was also important in the Active Days category among those that adhered ($p = 0.00002$) in helping them maintain or improve their adherence level. Finally, positive reinforcement resulted in positive effects on blood pressure both in the non-adherent and adherent groups, and for daily and weekly questionnaires it helped among those that were not adherent.

For physical activity adherence, we further break down the participants into low, medium and high activity adherence levels ((0,2), [2,5], (5,7] days). In this case, we
Table 5.4: Statistically significant interventions that result in different levels of adherence. Results are generated using two-tailed t-tests (Welch)

<table>
<thead>
<tr>
<th></th>
<th>t</th>
<th>p-value</th>
<th>95% C.I.</th>
<th>Without</th>
<th>With</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Active Days [did not adhere]</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clin. Pos. Reinf.</td>
<td>-2.5402</td>
<td>0.0268</td>
<td>(-3.2646,-0.2409)</td>
<td>389</td>
<td>12</td>
</tr>
<tr>
<td><strong>Active Days [adhered]</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daily Reminder</td>
<td>-1.9512</td>
<td>0.0538</td>
<td>(-1.1320,0.0093)</td>
<td>79</td>
<td>445</td>
</tr>
<tr>
<td>Inspirational</td>
<td>-3.3430</td>
<td>0.0081</td>
<td>(-3.5810,-0.7023)</td>
<td>10</td>
<td>514</td>
</tr>
<tr>
<td>Clinician Adh.</td>
<td>1.7689</td>
<td>0.0777</td>
<td>(-0.0398,0.7530)</td>
<td>344</td>
<td>180</td>
</tr>
<tr>
<td>Clin. Pos. Reinf.</td>
<td>-4.5109</td>
<td>0.0002</td>
<td>(-1.6784,-0.6529)</td>
<td>453</td>
<td>71</td>
</tr>
<tr>
<td><strong>Daily Questionnaires [did not adhere]</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clin. Pos. Reinf.</td>
<td>-4.9274</td>
<td>0.0003</td>
<td>(-7.7752,-3.0250)</td>
<td>308</td>
<td>13</td>
</tr>
<tr>
<td><strong>Weekly Questionnaires [did not adhere]</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inspirational</td>
<td>-14.0631</td>
<td>0</td>
<td>(-5.0586,-3.8171)</td>
<td>3</td>
<td>338</td>
</tr>
<tr>
<td>Weekly Reminder</td>
<td>-14.1000</td>
<td>0</td>
<td>(-5.1023,-3.8529)</td>
<td>6</td>
<td>335</td>
</tr>
<tr>
<td>Clin. Pos. Reinf.</td>
<td>-1.8277</td>
<td>0.0873</td>
<td>(-6.3437,0.4836)</td>
<td>326</td>
<td>15</td>
</tr>
<tr>
<td><strong>Blood Pressure [did not adhere]</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clin. Pos. Reinf.</td>
<td>-3.5535</td>
<td>0.0024</td>
<td>(-0.6586,-0.1685)</td>
<td>366</td>
<td>17</td>
</tr>
<tr>
<td><strong>Blood Pressure [adhered]</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clin. Pos. Reinf.</td>
<td>-2.0323</td>
<td>0.0450</td>
<td>(-0.2096,-0.0024)</td>
<td>476</td>
<td>66</td>
</tr>
</tbody>
</table>
Table 5.5: Clinician Positive Reinforcement and Inconsistent Adherence

<table>
<thead>
<tr>
<th>Did not adhere</th>
<th>t</th>
<th>p-value</th>
<th>95% C.I.</th>
<th>Without</th>
<th>With</th>
</tr>
</thead>
<tbody>
<tr>
<td>the previous</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(&lt;2 active days)</td>
<td>-1.7230</td>
<td>0.1346</td>
<td>(-3.8388,0.6569)</td>
<td>300</td>
<td>7</td>
</tr>
<tr>
<td>Adhered Inconsistently</td>
<td>-2.7541</td>
<td>0.0124</td>
<td>(-2.2123,-0.3036)</td>
<td>280</td>
<td>18</td>
</tr>
<tr>
<td>(2-5 active days)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fully Adhered</td>
<td>-0.4444</td>
<td>0.6579</td>
<td>(-0.7589,0.4816)</td>
<td>173</td>
<td>53</td>
</tr>
<tr>
<td>(&gt;5 active days)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

can see that the intervention impact for really low and really high adherence participants is inconclusive. However, clinician positive reinforcement is strongly correlated ($p = 0.0124$) with increased activity adherence for inconsistently adhering participants (those that exhibited anywhere between 2-5 active days) during the following week.

5.5.2 Predicting Weekly Participant Adherence

For each algorithm we used a Leave One Out Cross Validation (LOOCV) evaluation methodology. In this case, LOOCV refers to all the data points collected from the same participant. Accuracy, Area Under the Curve (AUC) and standard deviation of error across LOOCV for each classification task and each algorithm are compared in Figures 5.5.2, 5.5.2. Figures 5.3, 5.4 show the receiver operating characteristic (ROC) curve for each classifier.

As seen in Figures 5.5.2, 5.5.2 and 5.3, 5.4, it is possible to predict participant adherence from one week to the next with an accuracy of 82.5%, 72.4%, 83.4%, 76.8% for Active Days, Blood Pressure, Daily and Weekly Questionnaires respectively.

The most predictive features selected from our pipeline are shown in Table 5.6. Since evaluation was done using LOOCV we report the features that were selected in
Figure 5.2: Predictive pipeline in Wanda-CVD to estimate participant adherence the following week.

the majority of the folds. Interestingly, the contextual feature that is most predictive is a participant’s initial BMI and Lipid profiles. From the weekly data, the number of active days and number of daily questions answered are the most predictive features for a participant’s adherence during the following week.

5.5.3 Discussion

In the Women’s Heart Health Study based on a sample of young Black women (ages 25-45) in the urban Los Angeles setting staying connected via a smart phone application had a positive effect on adherence to an educational program and behavior change over a 6-month period. Overall, clinician positive reinforcement text messages and calls resulted in the most significant effect across all categories of adherence (BP, Daily Questionnaires, Weekly Questionnaires and Active Days). Recent studies have shown that case management and additional support given to patients by nurses, or other professional team members can impact patient’s health [112]. In our group of women education and setting attainable individualized goals was of utmost priority given the presence of two or more CVD risk factors for CVD and/or stroke. We set out to explore the use of technology as an adjunct to previous findings on lifestyle and behavioral
Figure 5.3: Receiver operating characteristic curves for the Activity and Blood Pressure prediction tasks.
Figure 5.4: Receiver operating characteristic curves for the Daily and Weekly Questionnaire prediction tasks.
<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Active Days</strong></td>
<td>Physical activity during the previous week that the participant adhered</td>
</tr>
<tr>
<td>Daily Questions</td>
<td>Number of daily questions answered the previous week (1-12)</td>
</tr>
<tr>
<td>BMI</td>
<td>Body Mass Index of the participant at the beginning of the study</td>
</tr>
<tr>
<td>HDL</td>
<td>Lipid profile of the participant at the beginning of the study</td>
</tr>
<tr>
<td><strong>Blood Pressure</strong></td>
<td></td>
</tr>
<tr>
<td>Daily Questions</td>
<td>Number of daily questions answered</td>
</tr>
<tr>
<td>Perceived Support</td>
<td>Perception of help and support received from others (baseline)</td>
</tr>
<tr>
<td>WaistC</td>
<td>Waist Circumferance of the at the beginning of the study</td>
</tr>
<tr>
<td>HDL</td>
<td>Lipid profile of the participant at the beginning of the study</td>
</tr>
<tr>
<td><strong>Daily Questionnaires</strong></td>
<td></td>
</tr>
<tr>
<td>Daily Questions</td>
<td>Number of daily questions answered</td>
</tr>
<tr>
<td>Perceived Support</td>
<td>Perception of help and support received from others</td>
</tr>
<tr>
<td>HDL / LDL</td>
<td>Lipid profile of the participant at the beginning of the study</td>
</tr>
<tr>
<td>BMI</td>
<td>Body Mass Index of the participant the beginning of the study</td>
</tr>
<tr>
<td><strong>Weekly Questionnaires</strong></td>
<td></td>
</tr>
<tr>
<td>Daily Questions</td>
<td>Number of daily questions answered</td>
</tr>
<tr>
<td>Weekly Questions</td>
<td>Number of weekly questions answered</td>
</tr>
<tr>
<td>Blood Pressure</td>
<td>Did the participant measure their blood pressure the previous week</td>
</tr>
<tr>
<td>Adherence</td>
<td></td>
</tr>
<tr>
<td>HDL</td>
<td>Lipid profile of the participant the previous week (1-12)</td>
</tr>
</tbody>
</table>
Figure 5.5: Accuracy, Area under the Curve and LOOCV standard deviation of error for the Activity and Blood Pressure prediction tasks and each of the classifiers
Figure 5.6: Accuracy, Area under the Curve and LOOCV standard deviation of error for the Daily and Weekly Questionnaire prediction tasks and each of the classifiers.
interventions that have not been sustainable. Instituting individualized health coaching, maintaining the connection and providing targeted support to meet these goals was successful in participants who were well motivated and consistent throughout the program. Of note, our findings show that RHMS messages were successful in motivating the group of participants that were inconsistently adhering to the study regimen one week, and in response to the communication improved their adherence or maintained positive adherence the following week. Additionally, results indicate that adherence during the previous week is the best predictor of adherence for the upcoming week.

5.6 Conclusion and Future Work

In this era of M-Health, the Women’s Heart Health Study has provided new knowledge and useful insights into the effectiveness of wireless coaching. In our study of urban young Black women, the connection provided a relationship utilizing emotional support and motivation which appears to be very important among this group of women in making lifestyle and behavioral changes. In addition, in this group of women with CVD risk factors, clinician positive reinforcement resulted in significant improvements in participant adherence for participants that were not adhering well to the protocol. It further encourages participants who adhere to keep their levels of adherence increased. When exploring together the effects of interventions, we found that participants in the middle of the adherence spectrum benefited most from it. Finally, we presented a framework for predicting weekly participant adherence based on weekly and contextual features. Such a framework has the potential to optimize clinician intervention by informing them about participant future adherence. In the future we hope to further validate our pilot study findings in a larger sample in diverse settings applying a micro-randomized approach to health coaching. Interventions like these that take little time and resources may be a major factor in increasing prevention efforts and decreasing the incidence of chronic disease in minority populations.
CHAPTER 6

A Knowledge-Graph Platform for Collecting and Exploring Health Data

6.1 Background

Health related data is inherently multi-modal and highly interconnected. Traditional methods for storing this information such as SQL databases or key-value data stores conceal that structure. Subjects and their data form a densely connected network that can be analyzed to extract additional information such as patient groupings and hidden effectors. The graph offers a much richer picture of the network. Furthermore, graph modeling removes the need for knowing all the kinds of data we may want to eventually store, as by design new types of nodes can be added without a need to account for the previous structure. Finally, a graph structure makes it straightforward to integrate data from different domains, allowing, for example, data from nutrition databases, drug databases etc. to be organically integrated with patient vitals.

6.2 Methodology

We developed a system and a method for dynamically organizing health knowledge in such a way that researchers, algorithms and users can explore dynamic medical knowledge, learn from it, and add to it as needed. The system has three components: a knowledge base, a backend system, and knowledge expansion learning algorithms. The platform also includes a set of application programming interfaces (APIs) that allow these
components to work together. The knowledge base is both statically and dynamically populated. Data from active sources is collected and integrated into the knowledge base. Knowledge from valid medical sources co-exist with the dynamic data, and links between them are automatically created to improve the inference power of the graph. Medical knowledge may be in the form of semantic association networks, documents or other media, or it may be previous research findings. Each piece of knowledge is associated with various types of meta-knowledge about what the knowledge is for, what form it is in, and so on. The information can also consist of data extracted from other media sources, such as scientific documents, books, journals, Web pages, film, video, audio files, and course notes.

The initial content of the health knowledge graph comprises of Electronic Health record data, Remote Health Monitoring data and disease semantic networks. It further includes meta-data about these ontologies such as disease co-occurrence frequencies. The health knowledge graph allows building a universal model for individuals’ health status and creates organic relationships across these statuses. It can further improve the prediction of behavioral and health outcomes via historical data as well as real-time monitoring. The graph platform can also give automated guidance in order to assist them in reaching their health goals.

6.3 Architecture

In the first version of our knowledge graph, we model patients, hospital/RHMS measurements and diseases. Our knowledge graph by construction is very flexible and additional concepts can be added and linked to the existing graph. A schematic of the current architecture can be seen in Figure 6.1
Figure 6.1: Health Knowledge Graph Architecture
6.3.1 Patients

A core concept of every health application, patients in the knowledge graph are modeled as individual nodes with properties such as date of birth, gender, race etc. No patient names are stored, however, to minimize privacy concerns. Instead, each subject is assigned a unique id when first inserted into the system.

6.3.2 Patient measurements

For the rest of the document, we refer to all time indexed patient information as measurements. This definition covers both actual measurements captured by a sensor as well as other types of typed dependent information such as physician/nurse notes, diagnoses, self-reported symptoms, questionnaires etc. (Figure 6.2)

Our graph design is aimed at efficiency of retrieval. For this, every patient is directly linked with his relevant vitals as they are entered the system. For every type of measurement, e.g. Heart Rate, Blood Pressure, a new linked list is created. The
edges from the patient node to the first measurement of each type are labeled according to the type of the measurement. Each subsequent measurement is linked with an edge labeled “next”. This convention allows fast retrieval of patient measurements in chronological order since we only have to traverse a single linked list. By comparison, a similar operation on a traditional database would require filtering based on the type of measurement and patient ID followed by an ORDER BY operation, all of which can be computationally expensive when dealing with large databases.

However, this proposed scheme, is plagued by increased insertion times since each such operation requires retrieving the tail of the respective linked list. We circumvent that limitation by dynamically populating a look up table indexed by patient id and measurement type. While this does increase space requirements for the graph storage, the overall space overhead is not significant. In fact, it is analogous to \( \text{number of patients} \times \text{types of measurements} \) which is relatively small compared to the size of the graph. Lookup operations are constant time, and after each insertion we update the lookup table with the index of the last inserted node. Deletions are also efficient since by default, property graphs edges can be bidirectional.

6.3.3 Diseases

Diseases and diagnoses are very important concepts for classifying patients. There is no universal way to model the thousands of possible disease concepts. The US health system diagnoses codes diagnoses using the ICD-9 system, although in 2015 will be replaced with ICD-10. The latter is already used in health systems around the world. One of our targets with this work is to correlate those diagnoses with specific outcomes. To achieve that, we need to be able to identify similar conditions specified in different terminology. We also want to define how similar two diagnoses are. We achieve that by expanding the disease subgraph to include the following:

- Disease ontologies: The Disease Ontology has been developed as a standardized
ontology for human diseases with the purpose of providing the biomedical community with consistent, reusable and sustainable descriptions of human disease terms, phenotype characteristics and related medical vocabulary disease concepts through collaborative efforts of researchers at Northwestern University, Center for Genetic Medicine and the University of Maryland School of Medicine, Institute for Genome Sciences.

- Disease co-occurrence frequencies: Empirically two distinct diseases/diagnoses are related based on how frequently they are co-diagnosed on the same patient. We model this statistical similarity by inserting a weighted undirected edge connecting two diseases. The weight of the edge is calculated from the Jaccard score:

$$ S(A, B) = \frac{P(A \cap B)}{P(A \cup B)} $$

Currently the disease co-occurrence frequencies are calculated using the Healthcare Cost and Utilization Project (HCUP).
6.4 Applications

6.4.1 Patient Similarity

Patient similarity assessment is an important task in the context of patient cohort identification for clinical decision support applications and critical for meaningful reuse of health data. The goal is to derive clinically meaningful metrics to measure the similarity and correlation between patients. These metrics will be used to create relationships between patient health statuses and patients themselves. Like social networks, patients and health statuses are connected through their contextual characteristics and clinical indicators and form groups. These connections are implicit rather than explicit. For example, patient comorbidities can provide insight into those similarities especially when used in conjunction with other information such as disease co-occurrences. Behavioral patterns emerging from a group of patients can be extended to those not yet experienced those patterns. With the right similarity measures as a basis, many applications can be enabled:

1. Automatic treatment suggestions and guidance based on cohort history: Cohort-tailored suggestions can be implemented using a variant of collaborative filtering (CF). CF is a method of making automatic predictions (filtering) about the interests of a user by collecting preferences or taste information from similar users. It been applied successfully to a variety of applications such as online shopping, automatic movie/book/music suggestion and more.

2. Cohort-specific risk analysis: Most RHMS today analyze risk and determine the need for clinician intervention based on a windowed evaluation of the most recent vital signs. With an established similarity measure, we can use similar patient history to determine what are the expected values and variances and better analyze and predict risk and individual outcomes.

3. Another domain in which patient similarity could be applied is tailored health
communication. Identifying individual groupings with a high risk of developing a primary or secondary condition enables the implementation of an effective preventive care program.

4. Underlying risk factor discovery. In the process of managing the health condition of an individual, data collection focuses on observable processes such as historical data, contextual information, vitals, activity indicators, symptoms and questionnaire data. This information usually has deeper information embedded. For example, information about meals also contains information about the nutrients and dietary profile of each meal. Similarly, information about medication intake contains information about side effects and active ingredients as well as compliance toward treatment. These extra layers of information can be utilized to analyze the importance of these underlying factors. Furthermore, using algorithms from graph analysis such as centrality analysis, we can analyze the influence of each of the identified factors.

6.4.2 Visualization

Our knowledge graph allows systematic collection of health related data. It also allows new insights by providing different ways to visualize the wealth of collected information. We describe the ways that our platform visualizes data for clinicians and researchers.

1. Complex graph queries: Using graph traversal queries, very complicated database queries can easily be expressed.

   e.g.: *return all patients that have at least 3 common comorbidities, age difference of less that 5 years and blood pressure average of 130 mmHg last week.*

Our platform provides a query interface to submit such queries (Figure 6.4). It supports submitting the queries using different graph query languages such as Gremlin and Cypher. The results are returned as data tables.
2. Interactive graph exploration: The data from the graph queries can be also displayed in graph form. The graph visualization allows interactive exploration of the data by clicking on a node. The system responds to such clicks by displaying all the adjacent nodes and respective relationships.
CHAPTER 7

A Flexible Data-Driven Comorbidity Feature Extraction Framework

Disease and symptom diagnostic codes are a valuable resource for classifying and predicting patient outcomes. In this work, we propose a novel methodology for utilizing disease diagnostic information in a predictive machine learning framework. Our methodology relies on a novel, clustering-based feature extraction framework using disease diagnostic information. To reduce the data dimensionality, we identify disease clusters using co-occurrence statistics. We optimize the number of generated clusters in the training set and then utilize these clusters as features to predict patient severity of condition and patient readmission risk. We build our clustering and feature extraction algorithm using the 2012 National Inpatient Sample (NIS), Healthcare Cost and Utilization Project (HCUP) which contains 7 million hospital discharge records and ICD-9-CM codes. The proposed framework is tested on Ronald Reagan UCLA Medical Center Electronic Health Records (EHR) from 3041 Congestive Heart Failure (CHF) patients and the UCI 130-US diabetes dataset that includes admissions from 69,980 diabetic patients. We compare our cluster-based feature set with the commonly used comorbidity frameworks including Charlson’s index, Elixhauser’s comorbidities and their variations. The proposed approach was shown to have significant gains between 10.7% to 22.1% in predictive accuracy for CHF severity of condition prediction and 4.65 % to 5.75 % in diabetes readmission prediction.
Disease and symptom diagnostic codes are a valuable resource for classifying and predicting clinical outcomes. To provide a common framework for reporting, billing and research, the medical community has adopted the International Statistical Classification of Diseases (ICD) standard. ICD encodes diseases, symptoms, complaints as well as injuries and causes of accidents. ICD was designed with a hierarchical structure to map conditions to corresponding categories based on similarity, and is periodically revised to incorporate new medical findings.

The ICD-9-CM (Clinical Modification - Version 9) in use in the US health system contains more than 15,000 different disease and symptom diagnostic codes. This number is even higher in the newer ICD-10 coding scheme. Considering each of the diagnostic codes as a separate feature in a predictive framework can be computationally costly. It is also complicated due to the high variability in disease occurrence frequencies.

To allow doctors to quantify risk of patient mortality and other outcomes, several comorbidity schemes have been proposed. The two most commonly used comorbidity frameworks are Charlson’s score and Elixhauser’s comorbidities and their variations. They were designed to improve prediction of specific outcomes such as patient mortality and hospital charges/length of stay. Both of these schemes along with their variations have been extensively studied, but focus only on specific conditions and outcomes.

7.1.1 Comorbidity Frameworks

7.1.1.1 Charlson Index

Charlson’s comorbidity index was first developed in 1987 [28] with the goal of providing a scoring system for patient mortality risk. It is used to predict ten-year mortality
Figure 7.1: Comorbidity co-occurrence network. The nodes are scaled and colored according to their degree in the network. Such networks can explain how certain disease groups are more likely than others. They can also be used to analyze similarities between patients, diseases and different treatments.

for patients based on a range of comorbid conditions such as AIDS, Cancer and Heart Failure. Twenty two conditions are considered in the Charlson Index and are weighted based on their severity with a score of 1, 2, 3, or 6. Charlson’s index is nowadays most commonly calculated on the basis of the Quan revision of Deyo’s ICD-9 mapping [75].

7.1.1.2 Elixhauser Comorbidities

Elixhauser’s comorbidity [36] measure was developed based on administrative data from the State of California. It takes into consideration a list of 30 comorbidities based on the ICD-9-CM codes. Since the comorbidities affect length of stay, hospital charges, and mortality differently, a unified index or score was not developed. The 30 comorbidities are also calculated on the basis of Quan’s revision [75]. Another commonly used revision of Elixhauser’s comorbidities was presented from the Agency for Healthcare Research and Quality (Elixhauser-AHRQ) [75]. Van Walraven et al. [104] presented a
unified score for Elixhauser’s comorbidities for predicting hospital death.

For a comparison of these measures and their variations in the context of predicting inpatient death, in-hospital adverse events, and readmission risk, we refer the reader to the review papers of Southern et al. [96], Farley et al.[38] and Sharabiani et al.[90].

In addition to these schemes, with the introduction and proliferation of Electronic Health Records (EHR) several methods have been proposed to identify risk factors by mining administrative data.

7.1.2 EHR mining methods

EHR data mining has been explored systematically to discover disease related information such as correlations, drug efficiency and other findings. Roque et al. [80] describe a framework to discover disease correlations through co-occurrence and map them to biological frameworks. Bauer-Mehren et al [15] created a network from unstructured EHR to examine the efficiency of certain treatments and identify patient cohorts. Yao et al. [115] examined EHR mining in the context of drug efficiency/ side-effect analysis. Baneyx et al. [12] built an ontology of pulmonary diseases from EHR using Natural Language Processing (NLP). In a prior effort [93], we described a framework for modeling the severity of condition for hospitalized congestive heart failure (CHF) patients. Other notable efforts include Chen et al [30], Rindflesch et al [78] and Cao et al [26, 25] that focus on disease meta-information mining. Jensen et al. [47] discussed the current status and challenges in mining EHRs. A review of recent approaches is provided by [61].

To enable the effective utilization of disease and symptom information in a general classification scheme we propose a flexible, data-driven feature extraction scheme from ICD-9-CM diagnostic codes that can easily adapt to different classification tasks. We examine the efficiency of our scheme in the context of quantifying and predicting CHF patient risk. More specifically, this research makes the following contributions:
• Designing and developing a flexible, data-driven approach for feature extraction from disease diagnostic information.

• Developing classification schemes for categorizing patients into high and low-risk cohorts.

Section 7.2 presents the different steps of our methodology and defines six methods for modeling severity of condition for CHF patients. Section 7.3 demonstrates the classification accuracy gains of the proposed methodology versus the commonly used Charlson Index, Elixhauser’s Comorbidity and AHRQ schemes and section 7.4 provides a summary of our effort.

A preliminary version of this work has been reported in [92].

7.2 Methodology

7.2.1 Feature Extraction from ICD-9-CM codes

Our objective was to utilize all the information provided by recorded ICD-9-CM diagnostic codes while ensuring that the resulting dimensionality of the disease features is low enough to be handled efficiently by the classification system. To reduce the dimensionality of diagnostic codes, we first identified disease groups with high frequency of co-occurrence. These co-occurrence frequencies are modeled through the following Jaccard score:

$$S(A, B) = \frac{P(A \cap B)}{P(A \cup B)}$$  \hspace{1cm} (7.1)

This score will be 0 when two codes never appear together (independent) and 1 when they always appear together. The co-occurrence frequencies along with the ICD-9-CM codes form a densely connected network with weighted edges (Figure 7.1). Such networks can explain how certain disease groups are more likely than others. They can also be used to analyze similarities between patients, diseases and different treatments.
Using the calculated Jaccard score, we computed the distance matrix of the ICD-9-CM codes with the following distance function:

$$D(A, B) = 1 - S(A, B)$$  \hspace{1cm} (7.2)

Finally, we clustered the ICD-9-CM codes according to their distance using hierarchical clustering based on the minimum variance method [110]. This clustering methodology ensures that the resulting clusters are as disjoint as possible. The clusters are converted into binary features by assigning a “1” if the patient presents a code within that cluster and “0” otherwise. The first 3 disease diagnoses are also encoded as categorical features based on the cluster number they belong.

From the hierarchical clustering, a dendrogram is generated. The dendrogram is a tree diagram that demonstrates the arrangement of the clusters. Depending on the height at which the dendrogram is cut, a different number of disease clusters are generated. Using a large number of such features tends to improve classification accuracy but above a certain number of features over-fitting becomes a concern (Figure 7.2). We initialize the search at the maximum height of the dendrogram and we reduce the height (thus increasing the number of features) at every step as long as the classification accuracy on the validation dataset increases. At each step we generate a new set of disease features and calculate the resulting classification accuracy on the validation dataset.

### 7.2.2 Data

We obtained EHR from the Ronald Reagan UCLA Medical Center between 2005 and 2009. The dataset consists of patients admitted primarily for CHF and related complications. The dataset includes patient demographics (gender, age, race), diagnostic information encoded in ICD-9-CM and hospitalization specific information including blood test results. It contains data from 4406 admissions from 3041 patients over 4 years. This dataset is the motivating factor behind this work as clinicians are increasingly interested in identifying risk factors from EHR data which can provide useful
Figure 7.2: Classification Accuracy vs number of disease clusters. The classification accuracy increases with the number of generated features until over-fitting phenomena appear.
insights into how patients should be monitored outside the hospital. The dataset contains multivariate signals from 31 sources, although not all sources are reported for all patients. These signals have different lengths depending on the length of stay of the patients. However, all patients have heart rate and blood pressure signals reported which are measured at an interval of 15 minutes for the duration of their hospital stay.

In addition to UCLA EHR data, we also used the UCI 130-US dataset, which includes admission data from 130 US hospitals from 1999 to 2008. The UCI 130-US dataset [101] consists of 100,000 patient admissions and contains 55 features: ICD-9-CM diagnoses, patient demographics, length of stay (LOS), readmissions and more. This dataset is intended to predict readmission probability using the recorded features.

Finally, we used administrative data from the 2012 HCUP NIS sample. The NIS sample contains over 7 million discharge records with ICD-9-CM codes. The NIS dataset includes ICD-9-CM diagnostic codes, length of stay, procedures, age and other demographic information.

We used the HCUP data to compute empirical co-occurrences between ICD-9-CM codes. The Ronald Reagan dataset as well as the UCI diabetes dataset were used as applications to validate our frameworks. Using separate datasets minimizes the risk of over-fitting our framework to a specific dataset.

7.2.3 CHF Patient Risk Classification

7.2.3.1 Data Preprocessing

From the Ronald Reagan Medical Center dataset, heart rate, systolic & diastolic blood pressure and weight during the first day of hospitalization were captured. From these signals, representative statistical features are extracted for each patient to be used as predictive features. The following features were computed during the first day of hospitalization: mean, max, min, range, and standard deviation. In our predictions, we also introduce administrative and demographic data. After cleaning and processing the
dataset, a total of 3048 valid records were identified from 1948 unique patients.

### 7.2.3.2 Outcomes

Our objective with the Ronald Reagan dataset is to obtain a prediction of long term risk and severity of condition for patients after they leave the hospital and to stratify patients into high and low risk. In RHMS for heart failure, it is important to monitor when a patient exceeds certain heart rate and blood pressure thresholds (Table 7.1). To validate our comorbidity framework, we employ it to predict the approximate volume of threshold based alerts a patient will generate. This prediction is important, as with conventional approaches in RHMS, the volume of alerts overwhelm clinicians [57] and leads to increased reduce treatment costs. An accurate prediction can lead to more efficient targeting of clinician interventions.

Alerts were extracted from physiological signals from every single admission record on a daily basis. We excluded from these calculations data from the first day of admission as that data is used as baseline features. The alerts are also ranked as medium or high priority depending on the degree of deviation from the defined thresholds.

![Figure 7.3: Predictive pipeline that utilizes the proposed comorbidity framework.](image)

To convert these alerts to long term outcomes, we computed the average daily value
Table 7.1: Daily Threshold Alerts

<table>
<thead>
<tr>
<th>Alert</th>
<th>Description</th>
<th>Priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I_1$</td>
<td>Heart Rate $\geq 120$ bpm</td>
<td>High</td>
</tr>
<tr>
<td>$I_2$</td>
<td>Heart Rate $\leq 50$ bpm</td>
<td>High</td>
</tr>
<tr>
<td>$I_3$</td>
<td>$100 \leq$ Heart Rate $&lt; 120$ bpm</td>
<td>Medium</td>
</tr>
<tr>
<td>$I_4$</td>
<td>Systolic BP $&lt; 80$ mmHg</td>
<td>Medium</td>
</tr>
<tr>
<td>$I_5$</td>
<td>Systolic BP $&gt; 160$ mmHg</td>
<td>Medium</td>
</tr>
</tbody>
</table>

per alert type for each admission and binarized them into “0” if the patient has 0 average alerts per day or “1” if greater than 0. These outcomes distinguish between low risk individuals (i.e no daily deviation from thresholds) and high risk, for each of the outcomes. This is particularly helpful in managing CHF patients as clinicians can use this information to prioritizing contact and treatment of patients in RHMS. The features and outcome statistics (average daily values for the outcomes) in the final dataset are summarized in Table 7.2.

7.2.4 Diabetic Patient Readmission Prediction

7.2.4.1 Data Preprocessing

From the original UCI diabetes dataset, we removed three features due to high missing value rate. Those features correspond to patient weight (97% missing), payer code (39.56% missing) and the medical specialty of the admitting physician (49.1% miss-
Table 7.2: Feature and outcome statistics in the Ronald Reagan CHF dataset

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Mean</th>
<th>St. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>65.327</td>
<td>17.066</td>
<td>18</td>
<td>107</td>
</tr>
<tr>
<td>mean heart rate</td>
<td>82.337</td>
<td>16.443</td>
<td>38.800</td>
<td>160.400</td>
</tr>
<tr>
<td>mean systolic blood pressure</td>
<td>122.469</td>
<td>23.059</td>
<td>48.000</td>
<td>246.000</td>
</tr>
<tr>
<td>mean diastolic blood pressure</td>
<td>67.178</td>
<td>12.414</td>
<td>19.500</td>
<td>137.250</td>
</tr>
<tr>
<td>mean weight</td>
<td>79.378</td>
<td>23.731</td>
<td>0.450</td>
<td>245.940</td>
</tr>
<tr>
<td>maximum heart rate</td>
<td>88.933</td>
<td>19.846</td>
<td>42</td>
<td>240</td>
</tr>
<tr>
<td>maximum systolic blood pressure</td>
<td>133.818</td>
<td>26.551</td>
<td>54</td>
<td>287</td>
</tr>
<tr>
<td>maximum diastolic blood pressure</td>
<td>74.601</td>
<td>15.077</td>
<td>21</td>
<td>180</td>
</tr>
<tr>
<td>maximum weight</td>
<td>79.685</td>
<td>23.900</td>
<td>0.450</td>
<td>245.940</td>
</tr>
<tr>
<td>minimum heart rate</td>
<td>76.122</td>
<td>16.294</td>
<td>27</td>
<td>153</td>
</tr>
<tr>
<td>minimum systolic blood pressure</td>
<td>111.410</td>
<td>24.279</td>
<td>11</td>
<td>246</td>
</tr>
<tr>
<td>minimum diastolic blood pressure</td>
<td>60.404</td>
<td>13.357</td>
<td>0.440</td>
<td>127.000</td>
</tr>
<tr>
<td>minimum weight</td>
<td>79.062</td>
<td>23.915</td>
<td>0.000</td>
<td>245.940</td>
</tr>
<tr>
<td>range heart rate</td>
<td>12.811</td>
<td>15.372</td>
<td>0</td>
<td>149</td>
</tr>
<tr>
<td>range systolic blood pressure</td>
<td>22.408</td>
<td>22.248</td>
<td>0</td>
<td>230</td>
</tr>
<tr>
<td>range diastolic blood pressure</td>
<td>14.197</td>
<td>13.731</td>
<td>0.000</td>
<td>128.000</td>
</tr>
<tr>
<td>range weight</td>
<td>0.624</td>
<td>5.659</td>
<td>0.000</td>
<td>106.060</td>
</tr>
<tr>
<td>gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>female:</td>
<td>1275</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>male:</td>
<td>1773</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$I_1$</td>
<td>0.075</td>
<td>0.172</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$I_2$</td>
<td>0.046</td>
<td>0.155</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$I_3$</td>
<td>0.263</td>
<td>0.329</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$I_4$</td>
<td>0.080</td>
<td>0.187</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$I_5$</td>
<td>0.129</td>
<td>0.265</td>
<td>0.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>
ing). We further remove subsequent admissions from the same patient to avoid overfitting to frequently admitted patients. We also remove admissions where the patient expires or is discharged to a hospice [101]. Finally, incomplete cases are removed leading to a final number of 66851 admissions. The reader is referred to the work of Strack et al. [101] for a detailed analysis of the dataset including the various features present.

7.2.4.2 Outcomes

In the UCI diabetes dataset, readmission outcomes are encoded as one of three values: less than 30 days, more than 30 days after discharge or no readmission. We converted the readmission values into two outcomes:

- Any readmission (R1): Was the patient readmitted (encoded as “1”) or not (encoded as “0”).

- Early readmission (R2): Was the patient readmitted in less than 30 days (encoded as “1”) or not (encoded as “0”).

7.3 Results

To validate our methodology, we compare it against the most commonly used comorbidity frameworks in the context of predicting the outcomes of sections 7.2.3.2, 7.2.4.2. The comorbidity frameworks we used as a comparison are the following:

- Charlson Index (original) [28]: Calculated on the basis of Deyo’s ICD-9 mapping [75].

- Elixhauser’s Comorbidity [36]

- Elixhauser’s Comorbidity (Quan’s Mapping) [75]

- AHRQ [75]
To compute these mappings, we used the icd9 package provided for the R statistical language [111]. We randomly split our datasets into 60% training, 20% validation and 20% testing. The validation set is used to adjust the number of generated clusters during training. During the training stage, for each of the binary outcome variables, we extract the most correlated demographic, administrative and physiological features. For each of the compared methods and for each of the subproblems we train Sequential Minimal Optimization (SMO) classifiers [73] and calculate the predictive accuracy in the test dataset. Figure 7.3 shows the steps of the predictive framework.

Tables 7.3, 7.4 summarize the classification accuracy achieved by our framework versus the examined comorbidity schemes. Our scheme outperforms existing comorbidity frameworks significantly in CHF outcome prediction from a minimum of 10.7% in $I_2$ to a maximum of 22.1% in $I_3$. Our scheme improves the readmission prediction for the diabetic patients to a lesser extent (4.65% - 5.75%). This can be explained by the fact that readmission prediction might be less dependent on co-morbidities and more on other factors such as insurance or age.

Figures 7.4, 7.5 and 7.6 show the Receiver Operating Characteristic (ROC) curve for the test set in CHF prediction. We only show the performance of our framework for the number of clusters that maximize the validation accuracy. The ROC curve demonstrates the trade-off between sensitivity and specificity. In each type of outcome,
Table 7.4: UCI dataset readmission prediction

<table>
<thead>
<tr>
<th></th>
<th>Proposed</th>
<th>Charlson</th>
<th>Elixhauser</th>
<th>Elixhauser (Quan)</th>
<th>AHRQ</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>R1</strong></td>
<td>57.21%</td>
<td>51.46%</td>
<td>50.83%</td>
<td>50.87%</td>
<td>50.88%</td>
</tr>
<tr>
<td><strong>R2</strong></td>
<td>55.82%</td>
<td>51.13%</td>
<td>50.00%</td>
<td>51.17%</td>
<td>49.81%</td>
</tr>
</tbody>
</table>

it can be clearly seen that our methodology provides an improved curve over all other co-morbidity frameworks.

Similarly, Figure 7.7 shows that our scheme outperforms traditional frameworks in the diabetes readmission prediction task.

Figures 7.8, 7.10, ?? and 7.11 demonstrate how the classification accuracy changes with a varying number of clusters. The accuracy value for each of the other comorbidity frameworks is displayed with a single line across all clusters for comparison. It can be seen that the improvement is not monotonic and on occasion it fluctuates significantly. Due to this fact, when applying our framework, it is important to examine the validation accuracy against different numbers of generated clusters.

7.3.1 Cluster Analysis

The number and type of clusters generated in each type of outcome and dataset vary significantly. To provide the reader with insight into the generated disease features a random height of 1.5 was chosen to cut the cluster dendrogram. This produces a total of 13 clusters. Table 7.5 describes the main disease types found in each such cluster. The contents were populated for each cluster by mining frequent words in the ICD-9-CM descriptions. One of those clusters ($D_1$) is significantly larger than the rest. It contains ICD-9-CM codes that are very sparsely mentioned in the HCUP dataset. The $D_2$ cluster contains conditions that are known to co-exist in chronic heart failure such as kidney and respiratory failure. $D_3$ is also interesting, with the co-occurrence of both CHF and
Figure 7.4: Receiver Operating Characteristic curves for the $I_1$-$I_2$ CHF outcomes
Figure 7.5: Receiver Operating Characteristic curves for the $I_3$-$I_4$ CHF outcomes
Figure 7.6: Receiver Operating Characteristic curves for the $I_5$ CHF outcome

Tobacco use (which is a self-reported ICD-9-CM code).

7.3.2 Limitations

Our framework outperforms existing co-morbidity frameworks in each of the examined prediction tasks. We have relied on a specific dataset (HCUP 2012) to extract co-occurrence information that was collected in the US health system. Further testing and different datasets might be needed when applying this scheme in regions that are outside the United States.

7.4 Conclusions

We have presented a novel data-driven framework to extract predictive features from disease and symptom diagnostic codes. Comparing it existing comorbidity frameworks,
Figure 7.7: Receiver Operating Characteristic curves for the UCI dataset readmission tasks
Figure 7.8: Classification Accuracy vs # disease clusters for the $I_1$-$I_2$ CHF outcomes. The accuracy value for each of the other comorbidity frameworks is displayed with a single line across all clusters for comparison.

Our proposed method can be used in predicting any type of patient outcomes such as adverse events (e.g. readmissions) and severity of condition. The number of comorbidity features extracted can be automatically adjusted to maximize accuracy in the particular task. We have successfully applied this framework to predict a CHF patient’s severity of condition and showed significant gains compared with the commonly used Charlson Index and Elixhauser’s comorbidities. We further explored gains achieved in predicting the probability of readmission in diabetic patients and our methodology was shown to
Figure 7.9: Classification Accuracy vs # disease clusters for the $I_3$-$I_4$ CHF outcomes significantly outperform the current comorbidity frameworks. This approach can be easily extended to other chronic diseases by targeting different risk metrics as outcome variables.
Figure 7.10: Classification Accuracy vs # disease clusters for the \( I_5 \) CHF outcome

7.5 Abbreviations

Figure 7.11: Classification Accuracy vs # disease clusters for the UCI dataset readmission tasks
<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_1$</td>
<td>Remaining Codes*, mention of malignant neoplasms</td>
</tr>
<tr>
<td>$D_2$</td>
<td>Heart failure, Valve disorders, Acute kidney failure, Respiratory failure</td>
</tr>
<tr>
<td>$D_3$</td>
<td>Congestive Heart Failure, Coronary Disease, Tobacco Use</td>
</tr>
<tr>
<td>$D_4$</td>
<td>Cerebrovascular disease, Malignant neoplasm, Substance Abuse</td>
</tr>
<tr>
<td>$D_5$</td>
<td>Diabetes II or unspecified, Atherosclerosis, Peripheral circulatory disorders</td>
</tr>
<tr>
<td>$D_6$</td>
<td>Diabetes II, nephropathy, retinopathy, neurological complications</td>
</tr>
<tr>
<td>$D_7$</td>
<td>Diabetes II, Liver disease, Cirrhosis, Digestive Complications</td>
</tr>
<tr>
<td>$D_8$</td>
<td>Chronic Kidney Disease , Renal Disease (End Stage)</td>
</tr>
<tr>
<td>$D_9$</td>
<td>Neoplasms of Hematopoietic cells , Leukemia</td>
</tr>
<tr>
<td>$D_{10}$</td>
<td>Diabetes I (juvenile type), related complications</td>
</tr>
<tr>
<td>$D_{11}$</td>
<td>Pressure Ulcer</td>
</tr>
<tr>
<td>$D_{12}$</td>
<td>Sepsis, Septic Shock</td>
</tr>
<tr>
<td>$D_{13}$</td>
<td>Dementia, Cerebral Atherosclerosis</td>
</tr>
</tbody>
</table>
CHAPTER 8

A Big-Data platform for Medical Knowledge Extraction from Electronic Health Records: Automatic Assignment of ICD-9 Codes

In this research effort, we present a big data platform for knowledge categorization in Electronic Health Records and examine its application to automatic assignment of ICD-9 codes. Our platform relies on reusable, adaptable components that can perform knowledge extraction at a large scale. For the ICD-9 automatic assignment, we build and validate our approach using data from the MIMIC II Clinical Database that contains over 20,000 discharge summaries. We show that our platform can achieve state of the art performance in this dataset and that the classification results improve with more data. Overall, in the first level of the ICD-9 hierarchy our algorithm achieves an average precision of 79.7% for an average recall of 70.2%.

8.1 Introduction

Due to the increasing adoption of Electronic Health Records (EHR) as well as the proliferation of Remote Health Monitoring Systems (RHMS), the volume of collected patient data has increased significantly. The data is not only increasing in volume but in variety as well. Advances in sensing technology have brought on new wearable devices and related data. Additionally, advances in software engineering have equally led to novel data sources and meta-data collection. This ever increasing wealth of
data presents new challenges in analyzing and extracting useful insights from it. To enable the effective utilization of this data, we propose a Big-Data platform capable of handling this that can easily adapt to different classification and mining tasks. We examine the efficiency of our platform in the context of automatic assignment of disease diagnostic codes.

### 8.1.1 Modeling & Mining Medical Knowledge

Structurizing medical knowledge has been the focus of many research efforts over the past decades. A variety of projects have been devoted to modeling medical knowledge. SNOMED CT (Systematized Nomenclature of Medicine–Clinical Terms) is a comprehensive clinical terminology maintained, and distributed by the International Health Terminology Standards Development Organization (IHTSDO), a not-for-profit association in Denmark. The Unified Medical Language System (UMLS) integrates and distributes key terminology, classification and coding standards, and associated resources to promote creation of more effective and inter-operable biomedical information systems and services, including electronic health records. Medical Subject Headings (MeSH) is the National Library of Medicine’s controlled vocabulary thesaurus.
Recently, Electronic Health Record (EHR) data mining has been the focus of data mining researchers to extract disease related information including correlations, drug efficiency and patient groupings. Roque et al. [80] describe a framework to discover disease correlations using co-occurrence frequencies and associate them with biological frameworks. In previous works we used EHR to model the severity of condition for hospitalized congestive heart failure (CHF) patients [93] and to extract dynamic features from comorbidity data [92]. Other notable efforts include Chen et al [30], Rindflesch et al [78] and Cao et al [26, 25] that focus on disease meta-information mining. We refer the reader to [61] for a review of recent approaches and to Jensen et al. [47] for a discussion of current status and challenges.

8.1.2 Disease Diagnostic Codes

As discussed in chapter 7, disease diagnostic codes are a valuable resource for classifying and predicting patient outcomes, generally through comorbidity measures. Since ICD codes are assigned primarily for billing purposes, they may not accurately represent the information contained in the medical data, making data mining less accurate. In addition, manual coding can be noisy and is prone to human error [70]. Previous efforts in automatic assignment have focused on assigning ICD codes based only on medical notes [71], [55] and manual rules.

Our methodology is focused on developing a big-data informational system to automatically assign disease diagnostic codes to discharge summaries. More specifically, this research makes the following contributions:

- Designing and developing an adaptable, big-data knowledge extraction platform from medical data.
- Developing a hierarchical classification scheme for automatically assigning ICD-9 codes using both text data from discharge summaries as well as demographic and administrative information.
This chapter is organized as follows: Section 8.2 presents the platform architecture and the different steps of our methodology and evaluation. Section 8.3 demonstrates the classification performance for automatic ICD-9 code assignment. Section 8.4 provides a summary of our effort.

8.2 Methodology

8.2.1 Architecture

Our platform relies on highly parallel, distributed processes to efficiently process the volume of available data. An overview of the platform can be seen in Figure 8.1. Data, as well as features and prediction results are stored in a distributed database engine. Feature extraction pipelines can be scheduled to run at predefined intervals and transform the raw data into input features for classifiers. To maximize reusability of those features, the feature extraction pipelines are independent of machine learning pipelines and the processed features are stored in the distributed database.

The machine learning pipelines are split into training and prediction. Training consists of retrieving processed input features for a given subset of the data (training set) and then training a machine learning algorithm on it. The trained model is stored to be used by the prediction pipelines. Our platform is designed to enable training of any supervised or unsupervised algorithm. Furthermore, since prediction results are stored in the Distributed Storage Engine, the pipelines are able to use the predictions as features for hierarchical classification applications. Finally, prediction pipelines are run at predefined intervals to produce predictions for the entire dataset.

8.2.2 Data

To validate our platform and ICD-9 assignment methodology we utilize the publicly available MIMIC II Clinical database [86]. The database contains over 20,000 de-
anonymized records from patients in the Intensive Care Unit (ICU) between 2001 and 2007. Patient demographics, nursing progress notes, administrative data and physiological signal recordings are reported. Lab tests, medications and other events are contained as well. Furthermore, each record reports the associated ICD-9-CM codes given during discharge.

The MIMIC-II dataset represents a large and diverse dataset. The following top level ICD categories are the most populous: diseases of the circulatory system (39.1%), trauma (10.2%), diseases of the digestive system (9.7%), pulmonary diseases (9.0%), infectious diseases (7.0%) and neoplasms (6.8%).

Figure 8.2: Hierarchical machine learning pipeline
8.2.3 Machine Learning Pipeline for ICD-9 Assignment

To be able to simultaneously consider text input with demographic and administrative features, we devised a hierarchical classifier. The overall classifier can be seen in Figure 8.2. The first level of classifiers (child classifiers) predict based on features from medical notes. More specifically, we compute the term frequency-inverse document frequency [11] and classify each instance of notes using Logistic Regression. The output of these classifiers, which represents the probability of assigning a certain ICD-9 code, is transformed and used as a feature for the parent classifier. We computed standard deviation, maximum, minimum, median, and mean of the probabilities for each subject’s notes set. In addition, the number of notes correctly classified and the number of notes per subject are used as features.

The second level of the hierarchical classifier analyzes the features from the first level together with the following demographic and administrative features: gender, date of birth, date of death, total intensive care unit days of subject, total lab events of subject, marital status, ethnicity, religion, overall payor group, admission type and admission source. To prevent the parent classifier from over-fitting to the first level classifiers, we use separate subsets of the training set for the parent and for the child classifiers. For the parent classifier, we present results from the commonly used Random Forest Classifier [21].

8.2.4 Evaluation

From the original Mimic II dataset, we removed patients with multiple hospital admissions. 20,000 patient records were used for training and testing. The first 16,000 records were used as training data and the remaining 4000 as test set. The training set was split further into first level training set (first 14,000 records) and second level training set (remaining 2000 records).

The ICD-9 scheme is a hierarchical coding scheme. Each code belongs to a group
of diseases and that group belongs to an even bigger group of conditions for a total of 5 levels. To assign ICD-9 values, we first classify an instance in the medical records as belonging to one of the top level groupings in the ICD-9 hierarchy. We further classify those belonging in that grouping into which. We train separate classifiers (as described in Section 8.2.3) for each level of the hierarchy and for each code. These classifiers are one versus all, where the label TRUE represents records which have a primary diagnosis in the examined grouping and FALSE otherwise.

8.3 Results

Figure 8.3: Precision, recall, f-measure and classification accuracy for the most populated ICD-9 top level categories.

Figure 8.3 presents the achieved precision, recall, accuracy and F-measure for each of the ICD-9 top level categories. Some of them are omitted due to small number (less than a hundred) of positive instances. Our methodology achieves high precision and recall for each of these categories.

Of note is that the classification precision and recall is highly dependent on the number of available positive instances in each ICD-9 class. For example, the top level category V01-V99 in our dataset has 708 positive training instances whereas 460-519
Table 8.1: Classification Results For Diseases Of The Circulatory System

<table>
<thead>
<tr>
<th>Class</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>390-459</td>
<td>93.11%</td>
<td>88.22%</td>
<td>87.95%</td>
<td>88.08%</td>
</tr>
<tr>
<td>440-449</td>
<td>95.46%</td>
<td>23.50%</td>
<td>72.88%</td>
<td>35.54%</td>
</tr>
<tr>
<td>393-398</td>
<td>98.90%</td>
<td>61.54%</td>
<td>19.51%</td>
<td>29.63%</td>
</tr>
<tr>
<td>430-438</td>
<td>90.11%</td>
<td>34.60%</td>
<td>93.01%</td>
<td>50.44%</td>
</tr>
<tr>
<td>410-414</td>
<td>96.63%</td>
<td>83.93%</td>
<td>93.74%</td>
<td>88.56%</td>
</tr>
<tr>
<td>420-429</td>
<td>85.26%</td>
<td>22.36%</td>
<td>65.80%</td>
<td>33.38%</td>
</tr>
<tr>
<td>415-417</td>
<td>99.51%</td>
<td>53.57%</td>
<td>78.95%</td>
<td>63.83%</td>
</tr>
<tr>
<td>396</td>
<td>91.86%</td>
<td>5.23%</td>
<td>65.22%</td>
<td>9.68%</td>
</tr>
<tr>
<td>414</td>
<td>64.06%</td>
<td>13.98%</td>
<td>82.20%</td>
<td>23.89%</td>
</tr>
<tr>
<td>410</td>
<td>56.35%</td>
<td>11.55%</td>
<td>78.10%</td>
<td>20.12%</td>
</tr>
<tr>
<td>428</td>
<td>90.61%</td>
<td>11.11%</td>
<td>66.67%</td>
<td>19.05%</td>
</tr>
<tr>
<td>424</td>
<td>89.36%</td>
<td>18.24%</td>
<td>96.43%</td>
<td>30.68%</td>
</tr>
<tr>
<td>427</td>
<td>82.61%</td>
<td>2.96%</td>
<td>69.23%</td>
<td>5.68%</td>
</tr>
<tr>
<td>434</td>
<td>88.80%</td>
<td>7.52%</td>
<td>88.57%</td>
<td>13.87%</td>
</tr>
<tr>
<td>432</td>
<td>98.02%</td>
<td>10.00%</td>
<td>58.33%</td>
<td>17.07%</td>
</tr>
<tr>
<td>433</td>
<td>94.85%</td>
<td>4.57%</td>
<td>44.44%</td>
<td>8.29%</td>
</tr>
<tr>
<td>430</td>
<td>98.87%</td>
<td>46.34%</td>
<td>52.78%</td>
<td>49.35%</td>
</tr>
<tr>
<td>431</td>
<td>96.95%</td>
<td>34.02%</td>
<td>44.59%</td>
<td>38.60%</td>
</tr>
<tr>
<td>441</td>
<td>44.52%</td>
<td>1.80%</td>
<td>97.22%</td>
<td>3.54%</td>
</tr>
<tr>
<td>440</td>
<td>87.82%</td>
<td>3.70%</td>
<td>84.21%</td>
<td>7.10%</td>
</tr>
</tbody>
</table>
contains only 138 such instances. Our hierarchical classifier is able to achieve 98.2% precision, 99.1% recall in the former and 68.6%, 47.5% respectively in the latter. This is promising as it signifies that our scheme can learn from an ever increasing volume of data.

Detailed results from the ICD-9 subset for diseases of the circulatory system are presented in Table 8.1. Our classification scheme's precision degrades in the lower levels. The reason is that in those levels there is a significant overlap between different codes.

![Graph](image)

**Figure 8.4:** Precision recall curve for each level of the classifier

Figure 8.4 displays the Precision-Recall curves for each level of the hierarchical classifier (Section 8.2.3) for the ICD-9 top level category 390-459. It can be clearly seen that the parent classifier has improved precision and recall from the classifier that
8.4 Conclusions

We have presented a big-data platform capable of adapting on a variety of medical knowledge extraction applications. We have successfully implemented on our platform a classification scheme for automatic assignment of ICD-9 codes to Electronic Health Records data. Our classification scheme achieves state of the art performance on the MIMIC-II dataset. Furthermore, it is able to take advantage of an increasing volume of data to improve the precision and recall of each ICD-9 class. In the future, we plan to extend our platform to other medical knowledge extraction applications.
CHAPTER 9

Building Continuous Arterial Blood Pressure Prediction Models Using Recurrent Networks

This paper presents a methodology for developing highly-accurate, continuous Arterial Blood Pressure (ABP) models using only Photoplethysmography (PPG). In contrast to prior approaches, we develop a system that exhibits dynamic temporal behavior which leads to increased accuracy in modeling ABP. We validate our approach using data from patients in the intensive care unit (ICU). We show that it is possible to build highly accurate, continuous blood pressure models using only finger work pulse oximeters. Our methodology achieves accurate systolic blood pressure estimation with a root mean square error of $2.58 \pm 1.23$ across the patient sample used. Furthermore, the continuous ABP signal is estimated with a root mean square error of $6.042 \pm 3.26$ and correlation coefficient of $0.95 \pm 0.045$. Our method enables designing robust Remote Health Monitoring Systems (RHMS) for Heart Failure patients without requiring traditional blood pressure monitors.

9.1 Introduction

Cardiovascular Disease (CVD) has been the leading cause of death in the United States over the last few decades. Statistics from the American Heart Association demonstrate that over 2,150 Americans die each day due to CVD [64]. In fact, in 2009 CVD was responsible for 31.3% of the total deaths in the U.S. CVD is also prevalent in Europe, and statistics in the European Heart Journal [68] report almost 4.1 million deaths per year.
due to CVD. Among all types of heart failure, blood pressure and heart rate changes are among the most significant non-invasive indicators of a change in a patient’s condition [1].

Remote Health Monitoring Systems (RHMS) have shown great promise in managing and preventing CVD. RHMS are clinical information systems designed to monitor individuals outside of traditional healthcare environments. Data from sensors are collected (usually by the patient himself) and transmitted wirelessly to a remote server for analysis. Clinicians and automated algorithms process the data to identify abnormal patterns. A schematic of such a system can be seen in Figure 9.1. RHMS systems, however, rely on robust collection of heart rate and blood pressure signals to produce high quality risk predictions. In recent years, continuous heart rate monitoring has been made significantly easier with wireless pulse oximeters where the subject only has to wear a small device on his/her fingertip. Blood pressure monitoring is less convenient due to the complexity of cuff-based monitoring devices. For RHMS scenarios, collecting reliable blood pressure measurements is associated with several major challenges:

- **Patient Condition**: Older patients, or those with severe disabilities may have difficulty properly collecting blood pressure data. They often forget or are too weak to capture the data themselves.

- **Patient Compliance**: Continuous blood pressure monitoring is challenging outside of a patient’s domicile as blood pressure devices are large and visible. Patients often are not motivated or even embarrassed to comply with RHMS data collection protocols.

- **Lack of training**: Finally, even when such data is captured, poor cuff placement and other measuring errors can lead to noisy data. This is a less of an issue with pulse oximeters as they are easier to place and wear.

In this paper, we focus our efforts on developing a methodology that will benefit
Figure 9.1: Standard architecture of an RHMS. Data collected from the patient is transmitted to a cloud server. The server analyzes and presents the data to clinicians and provides feedback to patients.
patients in RHMS by facilitating blood pressure collection. To achieve this, we propose a novel methodology to train accurate, continuous models for inferring blood pressure continuously through pulse oximeters using deep recurrent neural networks. Our work relies on the dynamic temporal behavior of the recurrent networks to learn long and short term characteristics of patient’s Photoplethysmography (PPG) data. As a result our system is able to produce high quality approximations of arterial blood pressure signals using input from a finger pulse oximeter.

This paper is organized as follows. Section 2 discusses background and related works. Section 3 describes our methodology. Section 4 provides the results and limitations. Finally, we provide concluding remarks in Section 5.

9.2 Background & Related Works

9.2.1 Cuffless Blood Pressure monitoring

Cuffless blood pressure monitoring has received a lot of attention in the last decade due to its potential to facilitate blood pressure data collection. This is rendered possible by the fact that blood pressure is highly related with pulse transit time, i.e. the time it takes the heart beat pulse to propagate from the heart to the peripheral arteries [72]. Most such methods rely on electrocardiograph (ECG) and PPG data to capture the pulse transit time. Such methods include those of Kachuee et al. [48] and Wang et al. [109]. Their results are promising, but ECG data collection is not without challenges as electrodes have to be attached to several parts of the patient’s body. Other researchers attempted to model systolic and diastolic blood pressure using only PPG data. Ruiz-Rodriguez et al. [82] used deep networks without temporal behavior and reported promising results albeit with high variance in accuracy. Samria et al. [88] do not report the actual prediction error but rather correlation with measurements. Our proposed system explicitly models the time dependency in the PPG signal which leads to increased accuracy in modeling arterial blood pressure. Furthermore, previous method-
ologies are limited to predicting only specific metrics of the blood pressure waveform (systolic, diastolic) while our proposed system can reliably output a continuous ABP signal.

9.2.2 Remote Health Monitoring Systems

In the last decade, early successes coupled with advances in sensors have evolved Heart Failure RHMS systems from very basic forms (phone interaction, written reports) to advanced end-to-end systems [54]. Among successful applications of RHMS, the system designed by Antonicelli et al. [9] both reduced mortality as well as readmission rates for congestive heart failure patients in RHMS. Another RHMS study carried out by Morguet et al.[63] concluded that a 50% reduction in hospital admissions (38 versus 77/100 patient years, P = 0.034) and a 54% reduction in hospital length of stay is achievable for congestive heart failure patients with RHMS. A few representative solutions are presented to demonstrate the evolution of RHMS systems. An early study by Chaudhry et al. [29] required participants to make daily phone calls to an automated telemonitoring system for a period of 6 months. Each call played a prerecorded voice message that consisted of a series of questions about symptoms and weight for which the participants had to provide answers using the keypad on the phone. The responses were then downloaded from the telemonitoring system to an Internet website for daily review by clinicians. Another heart failure study conducted by Soran [94, 95] included an electronic scale and an individualized symptom response system connected to a computer database via a standard phone line. Patients were instructed to weigh themselves and answer a series of questions daily. Nurses reviewed the transmitted data on a daily basis and immediately contacted patients whenever the data fell out of a healthy range.

As later studies and systems evolved [20, 54], they employed more sophisticated data collection methodologies as well as data analysis algorithms. In contrast with conventional RHMS, these analytics-based RHMS employ machine learning algorithms to predict the risk of an adverse medical event. It can be concluded from the evidence [57]
that analytics-based RHMS work better than threshold-based ones and can help further reduce treatment costs.

Regardless of the algorithms used in a RHMS, accuracy is limited by the data quality. Achieving high-quality blood pressure measurements will enable more reliable monitoring of CVD patients outside of the hospital.

9.3 Methodology

9.3.1 Data

To validate our approach, we collected data from the MIMIC database [62]. The database was collected from patients in the intensive care unit (ICU). As a result, the dataset contains highly varied blood pressure measurements as the patients are undergoing treatment and receiving drugs. As such, it is a good benchmark for our algorithm’s accuracy. We examined signals from 200 patients in the dataset. After excluding short signals and signals with unacceptable blood pressure values (due to being collected in the ICU), we trained and validated regression models for 42 patients. The extracted data contain two signals per patient. The PPG from the fingertip and invasive arterial blood pressure (mmHg). Both signals are recorded at 125 Hz frequency.

9.3.2 Long-Short Term Memory Neural Networks

Deep (or multilayer) neural networks research originates in the 1980s with the seminal papers of Hornik et al. [45] as well as that of Hinton et al. [42]. A neural network is a model loosely based on biological neural networks represented by a set of “neurons” connected with edges with numerical weights. While artificial neural networks initially received a lot of attention from researchers, there were technological and algorithmic limitations that prohibited training very deep networks.

The research field was revived around 2006 with the work of Hinton et al. [43]. In
Figure 9.2: A depiction of the training and test sets used
the recent years, advancements in distributed and parallel computing and a new set of training algorithms led to a proliferation of new approaches and applications with state of the art performance in many prediction and classification tasks. For a more thorough understanding of deep neural networks we refer the reader to the book of Bengio et al. [18] and the review papers of Arel et al. [10] and Bengio et al. [17].

Our work relates closely to a variation of deep neural networks, called recurrent neural networks. In this class of artificial neural networks, the connections between the “neurons” form a directed cycle. This enables them to learn patterns in time-dependent sequences. In fact, they have been applied successfully in handwriting [40] and speech recognition [87] among others. The variation of recurrent networks we utilized are the Long Short-Term Memory (LSTM) models [44]. These models have the ability to learn long term dependencies without the issues that affect traditional recurrent networks. The complete network we used consists of an input layer of nodes, an LSTM and a fully connected output layer with one output node. The complete network is shown in Figure 9.3.

For each patient, we trained a single recurrent neural network. Performing per patient training allows achieving high quality results, as the blood pressure signal depends on the specific properties of the peripheral system of each patient. In an RHMS, calibration does not pose an issue as it can be done when the patient enrolls in the RHMS (usually done at the hospital). Alternatively, a non-invasive, continuous blood pressure signal can be used. It has been shown to correlate well [89] with invasive arterial blood pressure measurements.

The slow drift component in each signal (PPG and ABP) was removed by subtracting the result of a linear, least-squares fit from the data. Subsequently, each patient record was split into training, validation and test parts using an 80% - 10% - 10% non-overlapping split across time (Figure 9.2). For each of these sets, we extracted overlapping signal windows from the PPG signal. As output, we used the value of the ABP signal that corresponds to the last datapoint in each window.
Figure 9.3: The input of the network is the processed, windowed PPG signal and the output is the ABP value at the current time step.
9.4 Results

9.4.1 Arterial Blood Pressure Estimation

The network was optimized using the RMSprop algorithm [103] and mean square error as its objective. To prevent over-fitting, Dropout [99] was used during training. With this technique, network nodes and their connections are randomly dropped to minimize learned feature dependencies (co-adaptation). From the estimated arterial blood pressure signal, we compute the systolic and diastolic blood pressure values as the local maxima and minima respectively. Over successive epochs, we update the trained model if the current systolic blood pressure validation error is less than the current best (Algorithm 1).

```
while maxepochs > current epoch do
    network optimization over the MSE of the ABP
    compute validation MSE of SBP
    if MSE of SBP < current best then
        update model
    else
        continue
    end
end
```

**Algorithm 1**: Training algorithm

The full signal MSE or the diastolic blood pressure MSE can also be used as targets depending on the application. Training is stopped after a predetermined number of epochs. Figures 9.4 and 9.5 demonstrate the evolution of the validation and test mean square error during the training of the network. The validation error decreases epoch after epoch up to a certain limit and fluctuates thereafter.

Detailed regression results for each of the examined patients are presented in Tables 9.1, 9.2. The results presented are for the continuous arterial blood pressure signal.
Figure 9.4: Systolic and diastolic blood pressure validation mean square error over training epochs (single patient)
Figure 9.5: Systolic, diastolic and arterial blood pressure test mean square error over successive model updates (single patient)
Overall, we were able to achieve high quality predicted blood pressure signals with an RMSE of $6.042 \pm 3.26$. The largest reconstruction errors are observed in the big slope regions of the blood pressure signal (Figure 9.9). These regions are usually less important than local maxima and minima, which correspond to the systolic and diastolic blood pressure.

Finally, Figures 9.8 and 9.9 show the output of the network on the test dataset in the beginning and the end of training respectively. It can be seen that the predicted signal closely follows the actual arterial blood pressure signal.

### 9.4.2 Comparison with Linear Regression

To evaluate the learning ability of our deep learning network, we compared it against linear regression. As shown in Figures 9.6 and 9.7, both systolic and diastolic root mean square error is significantly larger in linear regression than our proposed methodology. Some extreme values for the error of linear regression are omitted from the graphs for readability.

### 9.4.3 Limitations

The methodology and results we presented correspond to data collected from ICU patients. Tuning the network and the learning parameters may be necessary to generalize to patients outside the ICU. In addition, the dataset used did not contain age or other contextual information that could improve output accuracy. Finally, optimizing the window size, the number of hidden nodes and the network architecture was beyond the scope of this work. Our experiments indicate that further gains can be made by exploring those parameters as well.
Table 9.1: Regression Results

<table>
<thead>
<tr>
<th>patient</th>
<th>root mean square error</th>
<th>correlation</th>
<th>error $\sigma$</th>
<th>avg. error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SBP</td>
<td>DBP</td>
<td>ABP</td>
<td>ABP</td>
</tr>
<tr>
<td>1</td>
<td>1.753</td>
<td>2.806</td>
<td>7.140</td>
<td>0.951</td>
</tr>
<tr>
<td>2</td>
<td>3.945</td>
<td>1.865</td>
<td>6.671</td>
<td>0.953</td>
</tr>
<tr>
<td>3</td>
<td>2.607</td>
<td>2.218</td>
<td>5.217</td>
<td>0.964</td>
</tr>
<tr>
<td>4</td>
<td>1.324</td>
<td>1.605</td>
<td>3.708</td>
<td>0.985</td>
</tr>
<tr>
<td>5</td>
<td>2.242</td>
<td>2.908</td>
<td>5.089</td>
<td>0.969</td>
</tr>
<tr>
<td>6</td>
<td>2.584</td>
<td>1.465</td>
<td>4.294</td>
<td>0.971</td>
</tr>
<tr>
<td>7</td>
<td>2.022</td>
<td>0.988</td>
<td>3.665</td>
<td>0.979</td>
</tr>
<tr>
<td>8</td>
<td>4.071</td>
<td>1.891</td>
<td>3.429</td>
<td>0.984</td>
</tr>
<tr>
<td>9</td>
<td>3.392</td>
<td>1.629</td>
<td>3.521</td>
<td>0.984</td>
</tr>
<tr>
<td>10</td>
<td>1.989</td>
<td>1.360</td>
<td>5.432</td>
<td>0.970</td>
</tr>
<tr>
<td>11</td>
<td>3.493</td>
<td>1.528</td>
<td>4.216</td>
<td>0.985</td>
</tr>
<tr>
<td>12</td>
<td>2.736</td>
<td>5.986</td>
<td>5.856</td>
<td>0.909</td>
</tr>
<tr>
<td>13</td>
<td>3.009</td>
<td>3.251</td>
<td>7.174</td>
<td>0.944</td>
</tr>
<tr>
<td>14</td>
<td>6.187</td>
<td>2.422</td>
<td>7.847</td>
<td>0.930</td>
</tr>
<tr>
<td>15</td>
<td>3.151</td>
<td>1.773</td>
<td>22.850</td>
<td>0.791</td>
</tr>
<tr>
<td>16</td>
<td>1.250</td>
<td>1.663</td>
<td>5.903</td>
<td>0.973</td>
</tr>
<tr>
<td>17</td>
<td>1.041</td>
<td>0.925</td>
<td>3.491</td>
<td>0.991</td>
</tr>
<tr>
<td>18</td>
<td>1.554</td>
<td>3.214</td>
<td>9.435</td>
<td>0.926</td>
</tr>
<tr>
<td>19</td>
<td>0.892</td>
<td>0.990</td>
<td>9.536</td>
<td>0.849</td>
</tr>
<tr>
<td>20</td>
<td>1.095</td>
<td>1.792</td>
<td>9.236</td>
<td>0.816</td>
</tr>
<tr>
<td>patient</td>
<td>root mean square error</td>
<td>correlation</td>
<td>error σ</td>
<td>avg. error</td>
</tr>
<tr>
<td>---------</td>
<td>------------------------</td>
<td>-------------</td>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>SBP</td>
<td>DBP</td>
<td>ABP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>2.453</td>
<td>0.655</td>
<td>3.735</td>
<td>3.508</td>
</tr>
<tr>
<td>22</td>
<td>3.097</td>
<td>2.177</td>
<td>5.380</td>
<td>5.069</td>
</tr>
<tr>
<td>23</td>
<td>4.394</td>
<td>3.340</td>
<td>5.134</td>
<td>5.054</td>
</tr>
<tr>
<td>24</td>
<td>1.859</td>
<td>1.087</td>
<td>4.175</td>
<td>4.142</td>
</tr>
<tr>
<td>25</td>
<td>1.752</td>
<td>1.107</td>
<td>5.100</td>
<td>4.867</td>
</tr>
<tr>
<td>26</td>
<td>1.857</td>
<td>1.740</td>
<td>4.185</td>
<td>4.145</td>
</tr>
<tr>
<td>27</td>
<td>5.835</td>
<td>0.960</td>
<td>6.476</td>
<td>6.463</td>
</tr>
<tr>
<td>28</td>
<td>2.113</td>
<td>0.489</td>
<td>7.170</td>
<td>7.170</td>
</tr>
<tr>
<td>29</td>
<td>4.640</td>
<td>2.164</td>
<td>7.793</td>
<td>7.753</td>
</tr>
<tr>
<td>30</td>
<td>1.344</td>
<td>1.798</td>
<td>6.374</td>
<td>6.235</td>
</tr>
<tr>
<td>31</td>
<td>2.150</td>
<td>3.828</td>
<td>6.854</td>
<td>6.800</td>
</tr>
<tr>
<td>32</td>
<td>2.409</td>
<td>3.666</td>
<td>5.455</td>
<td>5.210</td>
</tr>
<tr>
<td>33</td>
<td>3.914</td>
<td>1.775</td>
<td>4.430</td>
<td>4.425</td>
</tr>
<tr>
<td>34</td>
<td>2.317</td>
<td>2.714</td>
<td>9.070</td>
<td>8.870</td>
</tr>
<tr>
<td>35</td>
<td>2.710</td>
<td>3.284</td>
<td>6.192</td>
<td>6.181</td>
</tr>
<tr>
<td>36</td>
<td>2.930</td>
<td>2.065</td>
<td>9.105</td>
<td>9.102</td>
</tr>
<tr>
<td>37</td>
<td>1.076</td>
<td>1.441</td>
<td>4.187</td>
<td>4.179</td>
</tr>
<tr>
<td>38</td>
<td>0.784</td>
<td>0.658</td>
<td>3.764</td>
<td>3.760</td>
</tr>
<tr>
<td>39</td>
<td>1.396</td>
<td>1.578</td>
<td>4.748</td>
<td>4.648</td>
</tr>
<tr>
<td>40</td>
<td>2.300</td>
<td>1.027</td>
<td>2.993</td>
<td>2.991</td>
</tr>
<tr>
<td>41</td>
<td>4.120</td>
<td>1.740</td>
<td>2.448</td>
<td>2.448</td>
</tr>
<tr>
<td>42</td>
<td>2.376</td>
<td>1.458</td>
<td>5.268</td>
<td>5.240</td>
</tr>
</tbody>
</table>
Figure 9.6: Systolic Blood Pressure RMSE in linear regression vs our proposed methodology
Figure 9.7: Diastolic Blood Pressure RMSE in linear regression vs our proposed methodology
Figure 9.8: Arterial blood pressure signal prediction (mean subtracted) at the beginning of the training (subject 7)
Figure 9.9: Arterial blood pressure signal prediction (mean subtracted) at the end of the training (subject 7)
9.5 Conclusions

We have presented a methodology for developing highly-accurate continuous models for inferring arterial blood pressure from finger photoplethysmography. The presented methodology enables continuous blood pressure monitoring for patients in RHMS without the inconvenience of a blood pressure cuff or the ECG sensor. Such a system has the potential to increase patient compliance and provide more accurate risk classification and prediction. In turn, this can lead to better management of heart failure.
CHAPTER 10

Summary and Contributions

The research presented in this dissertation has provided new predictive methodologies and tools for a more systematic approach in building RHMS. Each part of this work was validated using actual patient data collected in RHMS and hospitals. The proposed algorithms and frameworks were compared with existing methods and baselines and shown to outperform them. The evaluation metrics used included Receiver Operating Characteristic (ROC) curves and the Root Mean Square Error (RMSE), both of which are commonly used metrics in Machine Learning, Signal Processing and Medicine. In fact, ROC curves have been extensively used in evaluating diagnostic tests as they offer an overview of the trade off between the True Positive Rate and False Positive rate for the evaluated algorithm. They have also been used in epidemiology, radiology and other medical domains.

More specifically, the contributions of this dissertation can be categorized in the following two groupings:

- Predictive Methodologies for Remote Health Monitoring
- Tools and Frameworks for Mining Health data

10.1 Predictive Methodologies for Remote Health Monitoring

New predictive algorithms and analytics frameworks were developed that rely on various informational sources. Chapter 3 presented a severity—of—condition prediction framework that allows clinicians and support staff prioritize their interventions. This
can potentially enable scaling RHMS to larger populations without the ballooning costs associated with employing nursing and other supporting staff. This approach was validated on data from patients with CHF but can be adapted to a variety of conditions and scenarios.

Furthermore, chapter 4 extends the patient grouping methodologies of chapter 3 to train independent adverse event prediction models for each severity group. We demonstrate this to be a very effective approach that outperforms traditional single-model approaches.

Finally, the Women’s Heart Health Study allowed new predictive techniques to provide insight into patient compliance and success in RHMS. This is a very important task as many participants fail to adhere to the designed protocol of remote health monitoring which ultimately leads to increased need for clinician intervention. By providing accurate predictions on which patient is going to adhere or not on weekly basis, as described in chapter 5, the compliance of patients can be increased by targeting correctly the individuals with appropriate reminders.

Overall, chapters 3—5 introduce a set of predictive methodologies intended to improve quality of care for the patients enrolled in RHMS. More targeted clinician interventions can lead to lower costs for scaling out these systems, allowing RHMS to reach larger and more diverse populations. This research motivates a new type of information systems, the “Remote Health Prediction Systems”. On top of allowing clinicians to monitor and treat individuals, this class of systems relies on powerful predictive analytics to partially automate patient management and treatment.

10.2 Tools and Frameworks for Mining Health data

The first chapters of this dissertation introduced frameworks for specific prediction tasks, such as severity of condition and patient adherence. The rest of this work focused on developing a prototype RHMS platform and associated methodologies to enable
more systematic approaches in mining and utilizing data from RHMS and EHR.

Chapter 6 presented a prototype platform to collect, model and explore health data. The system relies on the expressiveness of property graphs to represent the data as an heterogeneous network of information. The proposed platform enables clinicians and researchers to derive clinically meaningful metrics to measure the similarity and correlation between patients. Furthermore, the knowledge—graph platform allows powerful data exploration & visualization.

Extending the knowledge—graph of chapter 6, chapter 7 presented a new framework that relies on the disease—disease subgraph to extract meaningful predictive features from comorbidities. The comorbidity framework presented was shown to outperform commonly used approaches including Charlson’s and Elixhauser’s comorbidity frameworks. This new methodology is also more flexible and can be easily adapted to different prediction tasks.

To improve feature extraction from comorbidities and to enable large—scale exploration of medical data, chapter 8 introduced a big—data medical knowledge extraction platform. This platform, which relies on Open Source, Distributed Computation components, can be adapted easily to different prediction and modeling tasks. Additionally, a specific application was designed on top of the platform to automatically assign disease codes to medical notes. This can potentially improve the features extracted with the framework of chapter 7.

Finally, chapter 9 extended the RHMS platform of chapter 6 with the ability to infer a patient vital signal that is costly or inconvenient to measure from an easily observable one. This methodology was validated by exploring its ability to infer arterial blood pressure from the photoplethysmography (PPG) signal. This contribution could have a significant impact on how data are collected in RHMS. A patient could potentially only measure blood pressure when needed instead of doing so at predefined intervals.

**Conclusion:** Combined, these two sets of advancements provide clinicians and
researchers with a set of tools to perform complex health analytics and improve remote health monitoring. Patient quality of care as well as quality of life could be significantly improved by employing these methodologies. In fact, this research enables clinicians to more effectively target patients and ensure high compliance. This effectiveness of targeting interventions could also lower costs for scaling out these systems. In turn this would allow extending the health benefits of RHMS to larger and more diverse populations.


125


[52] C. Klersy, A. De Silvestri, G. Gabutti, F. Regoli, and A. Auricchio. A meta-
analysis of remote monitoring of heart failure patients. *Journal of the American


[54] M. Lan, L. Samy, N. Alshurafa, M.-K. Suh, H. Ghasemzadeh, A. Macabasco-
O’Connell, and M. Sarrafzadeh. Wanda: An end-to-end remote health mon-
itoring and analytics system for heart failure patients. In *Proceedings of the

[55] L. S. Larkey and W. B. Croft. Automatic assignment of icd9 codes to discharge

[56] T. A. Lasko, J. C. Denny, and M. A. Levy. Computational phenotype discov-
ery using unsupervised feature learning over noisy, sparse, and irregular clinical

[57] S. I. Lee, H. Ghasemzadeh, B. Mortazavi, M. Lan, N. Alshurafa, M. Ong, and
M. Sarrafzadeh. Remote patient monitoring: what impact can data analytics
ACM, 2013.

[58] P. Liao, P. Klasnja, A. Tewari, and S. A. Murphy. Micro-randomized trials in

[59] J. Librero, S. Peiró, and R. Ordiñana. Chronic comorbidity and outcomes of
hospital care: length of stay, mortality, and readmission at 30 and 365 days.

[60] A. Lymberis. Smart wearables for remote health monitoring, from prevention
to rehabilitation: current r&d, future challenges. In *Information Technology


[68] M. Nichols, N. Townsend, P. Scarborough, and M. Rayner. Cardiovascular dis-


