May 2021

Two Essays on Leveraging Analytics to Improve Healthcare

Deepika Gopukumar
University of Wisconsin-Milwaukee

TWO ESSAYS ON LEVERAGING ANALYTICS

TO IMPROVE HEALTHCARE

by

Deepika Gopukumar

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in Management Science

at

The University of Wisconsin-Milwaukee

May 2021

ABSTRACT

TWO ESSAYS ON LEVERAGING ANALYTICS TO IMPROVE HEALTHCARE

by

Deepika Gopukumar

The University of Wisconsin-Milwaukee, 2021
Under the Supervision of Professor Huimin Zhao

The healthcare cost has continued to increase over the past few years despite various policies, efforts, and initiatives taken by the government. It is still projected to grow over the next few years by the Centers for Medicare and Medicaid Services (CMS). Readmissions have been a major contributor to the increase in costs and have always been a contributing factor. To get a perspective, considering the fact that at least 9% of individuals who had COVID-19 were likely to get readmitted shortly, according to a study by the Centers for Disease Control and Prevention (CDC) COVID-19 response team, along with their high estimated treatment cost, the problem of high healthcare costs will continue to grow. The implementation of the American Recovery and Reinvestment Act of 2009 has led to massive increase in digital health data facilitating various studies to utilize analytics to improve healthcare. The goal of the two essays in this dissertation is to address the identified research gaps in the literature in readmission analytics.

In Essay 1, I deploy the term readmission in two different ways and then focus on building and identifying predictive models that are suitable for costs billed by hospitals for the identified readmission categories. By using a data-driven approach, my initial analysis revealed that 21%

of readmitted individuals (regardless of the number of days to readmission) alone contributed to 48% of the healthcare cost. Apart from that, my analysis revealed that the readmission cost (for the identified readmission categories in this study) varied from the previous admission cost at both individual and aggregated levels. Deep learning-based models performed the best for all scenarios.

In Essay 2, I focus on creating a multitask learning-based joint model for predicting different outcomes related to readmissions, namely, likelihood, cost, and length of stay. I then evaluate the performance of the joint model and analyze its usefulness. Analysis was done for the identified top three categories of readmission belonging to the same major diagnostic groups from Essay1. Results showed that the joint model performed slightly better than the single-task baseline model for specific scenarios. The joint model was also beneficial in determining predictors that were consistently important to predict all the outcomes related to readmissions regardless of not giving us a universally best model.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my professors and mentors Dr. Huimin Zhao and Dr. Abhijeet Ghoshal (University of Illinois at Urbana-Champaign), for their incredible encouragement and support. I would also like to thank them for patiently guiding me throughout my dissertation with valuable feedback and being available for me always irrespective of their busy schedule. I have learnt a ton from them. I learnt not only about conducting research from them, but they have played a key role as an inspiration for me to pursue a career in academia. I would also like to thank Dr. Sanjoy Ghose for guiding, motivating, and supporting me throughout from the beginning of my Doctoral studies till now, and providing me his valuable feedback on my dissertation. He has also inspired me a lot to pursue a career in academia. I would also like to thank Dr. Yang Wang for all his ideas, support, and valuable feedback on my dissertation. I would also like to express my genuine gratitude to Mr. and Mrs. Lubar for their generous financial support. I would also like to thank Dawn, Andrea, James, and Keith for helping me during this journey. I would also like to express my gratitude to all other faculty and fellow Doctoral students at Lubar School of Business for their wonderful support. I am grateful to Agency for Healthcare Research and Quality for providing their National Readmission Database for my research.

I would also like to express my deepest gratitude to my husband Ranjit Nair for all his undeterred solid support throughout this journey irrespective of his busy schedule. I highly appreciate him for the way he took care of our baby Aditi while I was working on my dissertation so that I can focus uninterrupted. I would also like to thank my dad Dr. Gopukumar Sukumaran for being a well-wisher during this journey and my late grandfather for his blessings. Finally, I would like to thank my sweet little angel Aditi for being so understanding, forgiving, and always trying to cheer me up whenever she senses stress in me.

CHAPTER 1

Introduction

As per the National Health Expenditure report by the Centers for Medicare and Medicaid Services (CMS), the healthcare cost has continued to increase during this decade, and projections show such increase in costs to continue over the next few years as well. This projection was even without including the pandemic due to its uncertain nature. A part of this rising healthcare costs is due to the frequent readmission of patients. A recent study by COVID-19 response team of Centers for Disease Control and Prevention (CDC) revealed that at least 9 percent of the COVID-19 patients were expected to get readmitted shortly after discharge with a 1.6 percent having chance of getting readmitted more than once. Readmissions are already expensive. To get a perspective, considering that the average cost of COVID-19 treatment with complication has been estimated to be over $20,000 during the hospital stay which indicates that this problem of high costs due to readmissions will continue to grow. Fortunately, the implementation of the American Recovery and Reinvestment Act of 2009 has led to a rapid adoption of electronic health records by physicians and hospitals, and a massive increase in digital data about patients and treatments. This voluminous data has facilitated several studies focusing on predicting early readmissions, frequency of readmissions, and timing of readmissions. There have also been studies focusing on predicting health care costs utilizing heuristic, regression, classification trees, and clustering algorithms. On the policy side, measures have also been introduced by the Government to reduce excess readmissions. The Hospital Readmission Reduction Program (HRRP) was established by the CMS under the Affordable Care Act beginning October 2012 to impose penalty on hospitals with all-cause excess readmissions within 30-days for several chronic conditions and specific procedures. However, studies show that although the HRRP might have decreased readmissions significantly, it might not have helped in decreasing the health care cost substantially. On the contrary, studies indicate that the HRRP, in some cases, might have even unintentionally led to an increase in mortality incidents. Apart from this, the HRRP might have prevented readmission within 30 days only for the patients to get readmitted at later stages (that is, after more than the 30 days window has passed), which would have later contributed to the increase in healthcare costs. Based on the results from systematic literature review, a few research gaps were identified.

This dissertation advances the applications and techniques used for costs billed by hospitals for readmission-based analytics by developing predictive models. It consists of two essays. The first essay identifies the predictive model most suitable for predicting readmission costs by applying a variety of predictive techniques ranging from simple to sophisticated methods. The second essay proposes using a multitask learning-based simple interpretable model for jointly predicting tasks related to readmissions, namely, likelihood, cost, and length of stay.

Essay 1 - A Machine Learning Approach on Costs Billed by Hospitals for Readmissions

There are many studies focusing on predicting early readmissions or the length of stay but very limited studies on costs billed by hospitals for readmissions. In this essay, I primarily focus on using a variety of predictive methods to predict the cost billed by hospitals for readmissions based on the two defined categories of readmission to identify the models that are best suitable. Apart from that, I also analyze the following questions:

- Is there a variation in percentage of individuals contributing to healthcare costs?
- Did the number of readmissions vary based on the two defined categories of readmission?
- Did the readmission cost vary significantly compared to the previous admission cost based on the two defined categories of readmission?

Initially, I deploy the term readmission in two different ways and consider them as two different categories for my analysis. By using a data-driven approach, my analysis revealed that most of the cost were contributed by individuals who got readmitted without considering the number of days to readmission for both the identified categories of readmission. My analysis also showed that the number of readmissions varied based on the two defined categories of readmission. Apart from this, my analysis revealed that the cost associated with readmissions varied from the previous admission cost at both individual and aggregated levels. A variety of predictive methods, including state-of-the-art, were applied for predicting costs billed by hospitals for readmission of individuals belonging to the two identified categories of readmission mentioned above. The results revealed that deep learning models based on multilayer perceptron performed the best for both the identified categories of readmission.

Essay 2 - A Multitask Learning Approach for Heterogenous Tasks With Specific Sub-population

Many studies have focused on building single-task models in the context of readmissions, healthcare costs, and the length of stay. Even though the studies have found impact of the length of stay on readmissions and costs but none of these studies have tried to learn these outcomes using a joint model. There might be shared predictors that are useful in predicting all three outcomes. Multitask learning for homogenous task (classification) has been previously used in Information Systems (IS) research. However, to the best of my knowledge, there are no studies that have reported to apply multitask learning-based method for heterogenous tasks with specific sub-population. Initially, I focus on how to create a multitask learning-based joint model for heterogenous tasks with specific sub-population. I then identify the scenarios where the joint model was beneficial, and in what way the joint model was beneficial. For identifying the scenarios where joint model was beneficial, I compare the performance of the joint model with the single-task learning-based baseline model for the identified top three categories of readmission belonging to the same major diagnostic groups from Essay 1. I also identify the joint predictors that are helpful for all the outcomes i.e., readmissions, readmission costs, and the readmitted length of stay. Results showed that the joint model performed slightly better compared to the single-task learning-based baseline model for specific scenarios. Results also indicated that even though the joint model performed only slightly better in terms of predictive power in comparison to the baseline model, it was beneficial in determining predictors that was consistently important to predict all the outcomes specific to all three identified top categories of readmission belonging to the same major diagnostic category regardless of it not giving us a best model for readmissions.

CHAPTER 2

ESSAY 1 - A Machine Learning Approach on Costs Billed by Hospitals for Readmissions

2.1. Introduction

Electronic health record (EHR) is now widely used by hospitals. Data shows that the adoption of EHR by physician offices has almost doubled since 2008 (Jamoom and Yang 2016). At least partly, the reason is the mandate by the Government related to using EHR for both payers and providers, which was implemented as part of the American Recovery and Reinvestment Act of 2009 to treat patients administered under government insurance (Atherton 2011). As per HealthIT.gov website created by the Office of the National Coordinator for Health Information Technology (ONC), EHR data consists of patient-related data, such as administrative data (such as billing, payments), health-related data (such as diagnosis, procedures, comorbidities), and hospital-related details of the patient getting treatment, etc. Research shows that electronic health records contribute toward improving the overall health care quality and in reducing adverse drug reaction events (Campanella et al. 2016; Plantier et al. 2017).

Even with the implementation of technological innovations like EHR and various reforms for funding healthcare initiatives, the healthcare cost has continued to increase. As per the recent National Health Expenditure fact sheet by the Centers for Medicare and Medicaid Services (CMS), the National Health Expenditure has grown 4.6% in the year 2018 and attributed for 17.7% of Gross Domestic Product. The United States of America (USA) spends over USD 10,000 per resident, which is considerably high compared to other countries included in Organization for Economic Co-operation and Development where the average cost is only USD 4,000 per person after adjusting for purchasing power (OECD 2019).

Among various possibilities, the shorter length of stay, which is a measure of hospital performance (also referred to as hospital efficiency), decreases costs per discharge (OECD 2020). Even a single day reduction in the length of stay had projected to save USD 680 among patients with community-acquired pneumonia (Fine et al. 2000). The length of stay during a single in-patient hospital visit started to decrease after Medicare implemented payment of a fixed amount for a particular diagnosis of a patient, which created hidden financial incentives for hospitals by releasing patients sooner (Frakt 2016). The insurance status of a person has also shown to influence the length of stay with uninsured having a shorter length of stay (Mainous et al. 2011). However, a shorter length of stay has been found to be positively correlated with frequent readmissions. For example, a hospital visit related to heart failure having a shorter length of stay has shown to increase in readmissions related to cardiovascular and heart failure conditions (Sud et al. 2017). As such, people with chronic conditions tend to get readmitted frequently (Mudge et al. 2011). The hospital cost associated with 30-day all-cause readmissions was about $41.3 billion for the year 2011 (Hines et al. 2014). Even pre-pandemic recent news reported that the hospital readmission costs $26 billion annually for Medicare alone (Wilson 2019). During this pandemic, as per a recent study by COVID-19 response team of Centers for Disease Control and Prevention (CDC), at least 9 percent of the Covid-19 patients were expected to get readmitted shortly after discharge with a 1.6 percent having chance of getting readmitted more than once based on the data from Premier Healthcare Database, and readmission was found to be common among patients who combated COVID-19 (Kuehn 2020; LaPointe 2020). As per Kaiser Health News, the average cost of COVID-19 treatment with complication during a hospital stay was estimated to be over $20,000 (Byrne 2020). This means that there could be further increase in readmission costs, which in turn would contribute to rising healthcare costs. Predicting readmission costs could help hospitals in determining optimal length of stay, as both the shorter and longer length of stay could lead to higher healthcare costs. For example, if a hospital determines that the cost associated with readmission is greater than just extending the length of stay for a patient, they could increase the length of stay of the patient, and vice versa. This might help in avoiding unnecessary readmissions due to early discharges. According to CMS, a readmission is defined as an admission to a hospital within 30 days of discharge from the same or another hospital irrespective of the cause of readmission (Commission 2007). Readmission policies vary from one nation to another nation (Kristensen et al. 2015). So, in this study, I deploy the term readmission in two different ways, namely, Readmission with Same Major Diagnostic Category (RSDC) and All-cause Readmission Category (RADC). RSDC is defined as an admission to a hospital within 30 days of discharge from the same or another hospital with the cause of readmission being the same, and RADC is defined as an admission to a hospital within 30 days of readmission irrespective of the cause of readmission. In this context, the term "cause of readmission" is considered to be readmissions based on major diagnostic category (MDC). Both planned and unplanned readmissions are considered.

Application of machine learning algorithms have previously shown to accurately predict healthcare costs and the prior cost has shown to be helpful in predicting the future cost (Bertsimas et al. 2008). Apart from this, predictive analytics has proven beneficial in other areas of healthcare like prediction of early readmissions, risk analysis, and preventive care (Bates et al. 2014; Lin et al. 2017). Based on descriptive statistics by the Agency for Healthcare Research and Quality (AHRQ), costs associated with readmissions were found to be higher for two-third of the common diagnosis for the year 2016 (Kommers 2019). As there are limited studies on predicting readmission costs compared to predicting readmissions, it is not known what kind of models will be better suitable for predicting readmission costs. So, in this essay, I focus on building predictive models for readmission costs for the identified categories of readmission (i.e., both RSDC and RADC).

Before directly focusing on building predictive models for both RSDC and RADC, I tried to do a few additional analyses. Firstly, for understanding the contribution of readmissions in terms of rising healthcare costs based on my definitions of readmission i.e., RSDC and RADC, I checked the variation in percentage of individuals contributing to healthcare costs. Next, I analyze whether readmissions varied for MDCs based on RSDC and RADC. This was done to see if redefining readmission with respect to MDC made any difference since readmission policies vary across countries. Then, I determined whether the readmission cost varied significantly compared to the previous admission cost for both RSDC and RADC. Finally, I strive to build predictive models for predicting readmission costs billed by hospitals for both RSDC and RADC.

The rest of the paper is organized as follows. In the next section, the existing literature in related areas are reviewed and the identified research gaps are explained. Then, the models used are explained and the results are reported using various performance measures. Finally, the implications, contributions, limitations, and proposed directions for future research are discussed.

2.2. Literature Review

Excess readmissions contribute to increasing healthcare costs. In an effort to reduce excess readmissions, the Hospital Readmission Reduction Program (HRRP) was established by the CMS under the Affordable Care Act beginning October 2012 to impose penalty on hospitals with all-cause excess readmissions within 30-days for chronic conditions, namely, Acute Myocardial Infarction (AMI), Heart Failure (HF), Pneumonia, Chronic Obstructive Pulmonary Disease

(COPD), Elective Primary Total Hip Arthroplasty and/or Total knee Arthroplasty (THA/TKA), and Coronary Artery Bypass Graft (CABG) surgery (McIlvennan et al. 2015). The HRRP decreased payment to hospitals with excess readmissions covered under Medicare insurance. There have been mixed views after the HRRP was implemented as after implementation, the number of readmissions decreased (Wasfy et al. 2017). However, a few studies indicate that the HRRP had led to unintended increases in mortalities, which might be plausibly causing a decrease in readmissions (Gupta et al. 2018; Wadhera et al. 2018).

A previous study whose objective was only to analyze the cost variation and not build any predictive models specific to readmission costs indicated that the cost variation between readmitted and non-readmitted individuals without considering Medicare hospitalization stay and readmission was about 60% higher in the case of readmitted individuals (USD 56,856) compared to the non-readmitted individuals (USD 35,465) in 2000-2011 (Zheng et al. 2019). Furthermore, the conditions of readmitted individuals were found to be worse based on Charlson Comorbidity Index (CCI), making them more likely to get readmitted after the 30-day period to avoid penalty, hence questioning the studies of reduction in readmissions if they actually helped in reducing healthcare costs (Zheng et al. 2019).

Extant studies in predictive analytics on readmissions and healthcare costs could be broadly classified into different categories, namely, predicting all-cause 30-day readmissions, predicting readmissions specific to a population, predicting time to readmissions, and predicting general healthcare costs (Appendix A). A literature review on the models for predicting readmissions showed that most of the developed models performed poorly (Kansagara et al. 2011). Readmissions were also found to be a strong predictor of mortality within a year in cancer patients after colectomy (Greenblatt et al. 2010).

There has been an increase in studies to predict readmissions to reduce excess costs associated with them. Recent studies on predicting readmissions have focused mainly on deep learning methods, such as artificial neural networks and convolution neural networks, as they were helpful in capturing the plausible non-linear dependencies among independent variables in electronic health records (Jamei et al. 2017; Wang et al. 2018). Apart from these, several studies used multiple logistic regression, support vector machine, neural networks, tree-based methods, etc., to predict readmissions early on and most of them performed better than the traditional standard tools used by hospitals like LACE index or HOSPITAL scores (Cui et al. 2018; Schoonover et al. 2014; Shadmi et al. 2015; Sushmita et al. 2016; Wang et al. 2018; Xiao et al. 2018; Yu et al. 2015). Bayesian analysis was found useful in calculating the probability of the future condition of a patient, which includes identifying whether the patient would be readmitted or not (Cai et al. 2016).

With the implementation of the HRRP, several studies nowadays have focused on creating and using models like beta geometric Erlang-2, naïve Bayes, multivariate logistic regression, etc. for identifying readmissions specific to a chronic condition or procedure mainly related to heart conditions (Bardhan et al. 2015; Shameer et al. 2017; Tabata et al. 2014). Predicting readmissions have also been focused on specific populations like the elderly population, pediatric populations, etc., where machine learning techniques have proven to be better compared to the standard tools like LACE index, and tree-based lasso techniques have shown to provide better interpretability (Cotter et al. 2012; Radovanovic et al. 2015). Multivariate Cox proportional hazard model was used in identifying time to readmission in patients with repeated hospitalizations due to psychosis (Schmutte et al. 2010).

As there are limited studies on predicting readmission costs, the literature review was also done on studies related to predictive modeling of general healthcare costs. Prior studies on healthcare costs showed that the previous cost was helpful in the prediction of the future healthcare cost (Bertsimas et al. 2008; Sushmita et al. 2015). Healthcare costs could be accurately predicted using data mining techniques (Bertsimas et al. 2008). Linear regression models built using factors, such as age, gender, and count measures like number of diagnosis, were helpful in predicting general health expenditure (Farley et al. 2006).

Based on the above literature review, a few research gaps were found. Most of the studies have a narrow contextual focus, such as identifying early readmissions or readmissions for a particular disease, single hospital, or diseases that are part of the HRRP, or time until readmission. Most of these studies also considered readmission based on the standard definition provided by the CMS, which considered readmission irrespective of the diagnostic category, i.e., all-cause. Only a few of the studies extended the number of days to 90 and 120 for readmission or built predictive models focusing on readmission costs. Almost all the studies focused on building models using a single hospital or a single region (California, Washington, and North Texas).

Only one study tried to predict costs associated with readmissions, but it tried only a few methods, such as linear regression and tree-based models, for building their predictive models, and the study was based on a small-scale dataset from a specific region (Sushmita et al. 2016). So, based on a single study that used a small-scale dataset from only one specific region, conclusions cannot be drawn on the models that would be best suitable for predicting readmission costs. Also, they did not use any sophisticated models, such as deep learning-based models, which has shown to improve predictive accuracy and have been found more suitable for healthcare-related data (Hammoudeh et al. 2018; Piccialli et al. 2021; Wang et al. 2018). Apart from that, their study did not analyze a rich dataset about patients from different institutions or regions, thereby, decreasing the scope for external validity of the model for predicting readmission costs.

Even though nowadays the focus is mainly on building simpler interpretable models as part of explainable artificial intelligence (XAI), restricting models to use glass box-based models will cause limitations in scenarios, especially where simpler methods might not capture the intricacies within the data and perform badly. Prior studies have shown that the machine learning algorithms in healthcare population-based research suffer from dataset drift as the input data is created from non-stationary units and they have issues to be generalized for a newer population as there could be differences in administrative practices (Kelly et al. 2019). Deep learning-based models have shown to work well with large datasets as more data would help it learn effective representations of the outcome variable.

Due to the above-mentioned reasons, the analysis for this essay was done on a nation-wide readmission dataset as it could help to generalize the results by taking into account the problem of dataset drift and varying administrative practices across different institutions or regions. In addition, given that electronic health data are susceptible to having the number of features to exceed the number of observations or having correlated features, voluminous, incomplete, and imbalanced, a need for using modern predictive methods that could handle such inherent problems in the data arises (Jovanovic et al. 2016).

2.3. Methodology

2.3.1. Dataset and its Description

Nationwide Readmission Database (NRD) by AHRQ was used for this essay (Databases 2013). The dataset includes individuals from the entire United States of America who were admitted for the year 2013. The total number of records in the dataset was 14,325,172. It includes both with and without repeat hospital visits. Each admission record consists of demographic (gender, age, median household income, etc.), clinical information (diagnosis, procedure used, etc.), comorbidities (hypertension, diabetes, depression, etc.), hospital details (bed size, teaching or non-teaching hospital, etc.), severity details (All Patient Refined Diagnostic Related Groups i.e., APR DRG for severity of illness, risk of mortality, etc.), cost-related and administrative-related data (length of stay, cost billed by hospitals, etc.). Variables used in this essay along with their descriptive statistics and description are given in the appendix (Appendix B). 576,701 and 1,091,580 individuals were identified for each of the identified readmission categories, i.e., RSDC and RADC, respectively.

The dataset has close to 285 mutually exclusive categories of ICD-9 codes for grouping diagnosis and procedures related to patients that could be used for adjustment of risks. Prior studies have shown that aggregated higher-level grouping of diseases were sufficient in providing better results compared to going to a specific disease at the lowest level of hierarchy in case of pediatric readmissions (Radovanovic et al. 2015). [1]

Before going into the next section, the terms previous admission cost, sum of previous admission costs, and average of previous admission costs are defined as these terms differ with respect to RSDC and RADC. The previous admission cost for RSDC is defined as the cost billed by the hospital for only previous last admission having same MDC. The previous admission cost for RADC is defined as the cost billed by the hospital for only previous admission irrespective of the MDCs. The rehospitalization cost or readmission cost or total charge for both RSDC and RADC is defined as the cost billed by the hospital associated with one readmission visit with the

1. As per the CMS, Diagnostic Related Groups (DRG) are grouped under Major Diagnostic Categories (MDC) formed focusing on a specific medical specialty and are mutually exclusive to make them clinically consistent. They are built based on principal diagnosis codes (ICD9-Codes in this dataset). In this context, MDC codes are considered, which are grouped at the higher level rather than going to the specific DRG payment codes.

readmission criteria based on the definitions of RSDC and RADC, respectively. The sum of previous admission costs for RSDC is defined as the total cost billed by the hospital for all the previous admissions having same MDC. The sum of previous admission costs for RADC is defined as the total cost billed by the hospital for all the previous admissions irrespective of the MDCs. The average of previous admission costs for RSDC is defined as the average cost billed by the hospital for all the previous admissions having same MDC. The average of previous admission costs for RADC is defined as the average cost billed by the hospital for all the previous admissions irrespective of the MDCs.

2.3.2. Models Used and Their Description

Before going to the results, a brief overview about the learning algorithms used for this study and the rationale behind selecting these specific models for analysis is given. Initially, a few experiments were run by splitting the data randomly (80% for training and 20% for testing) several times and the following were experimented:

- Having the baseline model that gives the predicted value of the readmission cost for all individuals as the overall average cost of previous admissions for all individuals with respect to RSDC and RADC.
- Having the baseline model that gives the predicted value of the readmission cost for all individuals as the overall average cost of readmissions for all individuals with respect to RSDC and RADC.
- Having the baseline model that gives the predicted value of the readmission cost for all individuals as the ratio of average of readmission costs to the average of the previous admission costs and then multiply it with the previous admission cost with respect to RSDC and RADC.

- Running the model using average of the previous admission cost with respect to RSDC and RADC as an independent variable.
- Running the model using sum of the previous admission cost with respect to RSDC and RADC as an independent variable.
- Running the model by just using the last previous admission cost with respect to RSDC and RADC as an independent variable.

Based on the above experiments, it was found that linear regression performed better than baseline models for both RSDC and RADC. Similarly, it was also found that using average of the previous admission cost with respect to RSDC and RADC as an independent variable performed slightly better compared to using sum of the previous admission cost and only the last previous admission cost with respect to RSDC and RADC as independent variables. So, for the actual analysis, I considered the below models for both RSDC and RADC with the average of previous admission costs as one of the independent variables (All the numeric independent variables were scaled except for the average of previous admission costs for which I applied log transformation. Log transformation was also applied to the readmission cost, i.e., dependent variable):

- Linear Regression (Baseline Model)

  Rationale: It is a simple and most widely used method in majority of the studies for modeling general healthcare costs (Farley et al. 2006; Leigh et al. 2005; Sushmita et al. 2015).
- Lasso Regression, Elastic Net Regression, and Ridge Regression

  Rationale: In the machine learning literature, regularization techniques are suggested to prevent overfitting and multicollinearity by placing a constraint on the loss function. It could either add the penalty as a sum of absolute value of coefficients (*L1* penalty) in case of Lasso or as a sum of squared value of coefficients in case of Ridge regression (*L2* penalty). Lasso gives us sparse solutions by shrinking the estimates to zero whereas ridge regression shrinks the estimates near zero. As explained in the literature review section, electronic health data are complex in nature with issues of multicollinearity, etc., so I considered applying regularization techniques for predicting readmission costs. Another advantage of regularization technique is that the created models are interpretable. Elastic Net regression takes the advantage of both Lasso and Ridge regression by linearly combining both *L1* and *L2* penalties.

- Extreme Gradient Boosting (XGBoost)

  Rationale: It is very popular and currently the most efficient gradient-boosted trees algorithm. Its performance for capturing both linear and non-linear relationships has made it one of the widely used algorithms. Prior studies have also shown tree-based models to be beneficial (Sushmita et al. 2016). So, I included this tree-based model for predicting readmission costs to take care of any non-linearity that exists.

- Deep learning model using Multilayer Perceptron (MLP)

  Rationale: As seen in the literature review section previously, deep learning-based models are found to be beneficial with regard to electronic health related data. As the dataset is tabular, I used deep learning model using MLP instead of convolution neural network (CNN). I did not use recurrent neural network (RNN) as majority of the records in the dataset for both RADC or RSDC had only one or two readmissions. MLP with shallow neural nets will have either one or two hidden layers whereas deep neural nets use three or more hidden layers. MLP hyper-parameter tuning was performed based on methods used in prior studies; the values used in this application are shown in Table 1. Even though having the number of hidden layers as one, two or three performed better in comparison to linear regression, I increased the number of hidden layers to 4 to get fine-tuned low error values and fewer number of epochs with consistent error values for majority of the epochs.

Table 1 Deep Learning MLP Configuration Details

Configuration	**Value**
Number of hidden layers	4
Number of neurons in first hidden layer	80
Number of neurons in second hidden layer	60
Number of neurons in third hidden layer	50
Number of neurons in fourth hidden layer	20
Activation functions used in hidden layers	ReLU
Final activation function	Linear
Batch type	Minibatch
Mini-batch (weights get updated after each mini-batch)	30
Momentum	0.9
Learning rate	0.0001
Number of epochs (1 epoch = 1 forward pass + 1 backward pass)	200

2.3.3. Performance Measures Used

Eight verification statistics were used to measure the performance of the methods used. I define n as the total number of observations, i.e., patients, y_i as the actual value of readmission costs incurred by patients, and $\hat{y}_i$ as the predicted values of readmission costs. Then, the measures are provided as follows:

<u>*Mean Absolute Percent Error (MAPE)*</u>: MAPE measures the error size in terms of percentage.

$$MAPE = \sum_{i=1}^{n} \left(\left(\frac{|\hat{y}_i - y_i|}{|y_i|} \right) \right) * 100$$

Root Mean Squared Error (RMSE): RMSE gives the standard deviation of the residual, which is the difference between actual and predicted values.

$$RMSE = \sqrt{\left(\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}\right)}$$

Mean Absolute Error (MAE): MAE gives the average value of errors for a given set of predictions.

$$MAE = \frac{\sum_{i=1}^{n}|\hat{y}_i - y_i|}{n}$$

Relative Absolute Error (RAE): RAE compares the total absolute error of the model to the total absolute error of the simplest model.

$$RAE = \frac{\sum_{i=1}^{n}|\hat{y}_i - y_i|}{\sum_{i=1}^{n}|mean(y_i) - y_i|}$$

Root Relative Squared Error (RRSE): RRSE gives the relative comparison of what it would have been if a naïve model was used.

$$RRSE = \frac{\sqrt{\sum_{i=1}^{n}|\hat{y}_i - y_i|^2}}{\sqrt{(\sum_{i=1}^{n}|mean(y_i) - y_i|^2)}}$$

Normalized Root Mean Square Error (NRMSE1): NRMSE1 is used to compare models with different scales.

$$NRMSE1 = \frac{RMSE}{\max(y_i) - \min(y_i)}$$

Normalized Root Mean Square Error (NRMSE2): NRMSE2 is used to compare models with different scales.

$$NRSME2 = \frac{RMSE}{\sum_{i=1}^{n} y_i}$$

Mean Absolute Deviation (MAD): MAD describes how the values are spread away from the mean.

$$MAD = \frac{MAE}{\sum_{i=1}^{n} y_i}$$

The lower the MAPE, RMSE, MAE, RAE, RRSE, NRMSE1, NRMSE2, and MAD, the better is the fit of the model.

2.4. Results

Initially, I analyzed the distribution of costs (in percentage) contributed by individuals (in percentage) by giving different criteria for readmissions within RADC and RSDC. From Table 2 row 2 (at least one readmission without any condition on the number of days) (excluding column headings), it is clear that 48% of healthcare costs came from 21% of individuals who got readmitted without considering the number of days to readmission for both RADC and RSDC. Similarly, from row 3 (excluding column headings) of Table 2, it is also clear that even if the number of days to readmissions is considered as 30 days, 11% of the individuals contributed towards 31% of the cost for RADC and 6% of the individuals contributed towards 17% of the cost for RSDC.

Table 2 Distribution of Individuals and Their Contribution Towards Healthcare Cost

Distribution Category	**RADC**		**RSDC**	
	% of individuals	**% of cost**	**% of individuals**	**% of cost**
Without any readmissions	79	52	79	52
At least one readmission without any condition on number of days	21	48	21	48
At least one readmission and having number of days less than or equal to 30 days	11	31	6	17
At least two readmissions and having at least one of the records with number of days less than or equal to 30 days	6	22	3	13
At least two readmissions and having at least two of the records with number of days less than or equal to 30 days	0.43	3.39	0.17	1.19
At least three readmissions and having at least one of the records with number of days less than or equal to 30 days	3	14	2	10
At least three readmissions and having at least three of the records with number of days less than or equal to 30 days	0.05	0.42	0.02	0.17

My further analysis on the dataset showed that the cost associated with readmissions varied from the initial admission cost for most of the diagnosis (Figure 1). The cost of readmissions was found to be higher compared to the previous admission cost for 49% of individuals for RADC category. Among that, the cost associated with readmissions was found to be thrice compared to the previous admission cost for nearly 15% of individuals. Similarly, the cost of readmission was found to be higher compared to the previous admission cost for 53% of individuals for RSDC category. Among that, the cost associated with readmissions was found to be thrice compared to the previous admission cost for nearly 16% of individuals. I also observed that the number of readmissions decreased for RSDC in comparison with RADC when the number of days was considered as a criterion for readmissions.

Next, I identified the major diagnostic categories having the highest number of readmissions for both RADC and RSDC. The two groups are quite similar in terms of MDCs having the highest number of readmissions. The categories with the highest number of readmissions for both RSDC and RADC are shown in Table 3 in descending order.

Table 3 MDCs Having the Highest Number of Readmissions

RSDC	**RADC**
Diseases and Disorders of the Circulatory System	Diseases and Disorders of the Circulatory System
Diseases and Disorders of the Respiratory System	Diseases and Disorders of the Respiratory System
Diseases and Disorders of the Digestive System	Diseases and Disorders of the Digestive System
Infectious and Parasitic DDs (systemic or unspecified sites)	Pregnancy, Childbirth and Puerperium
Diseases and Disorders of the Kidney and Urinary Tract	Mental Diseases and Disorders
Diseases and Disorders of the Nervous System	Diseases and Disorders of the Nervous System

Next, I analyzed if the average readmitted cost for each of the MDCs for both RSDC and RADC varied from the previous admission cost (Figure 1(a) and 1(b)). In case of RSDC, the average readmitted cost was higher compared to the average of previous admitted cost for 80% of the MDCs, whereas it was only 52% of the MDCs for RADC.

Figure 1 Comparison of Average Previous Admission Costs and Average Readmission Costs

Figure 1(a) RSDC

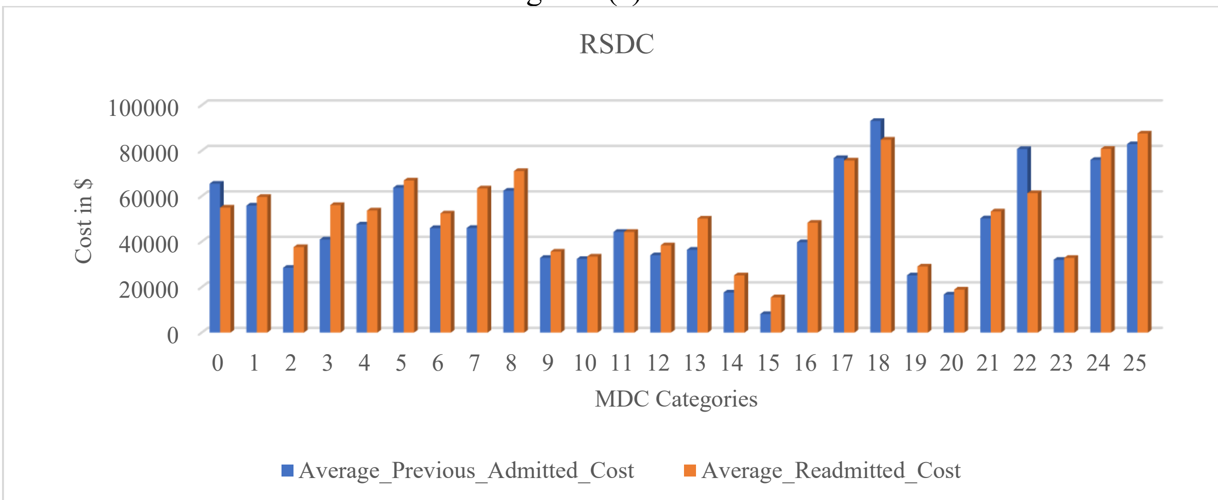
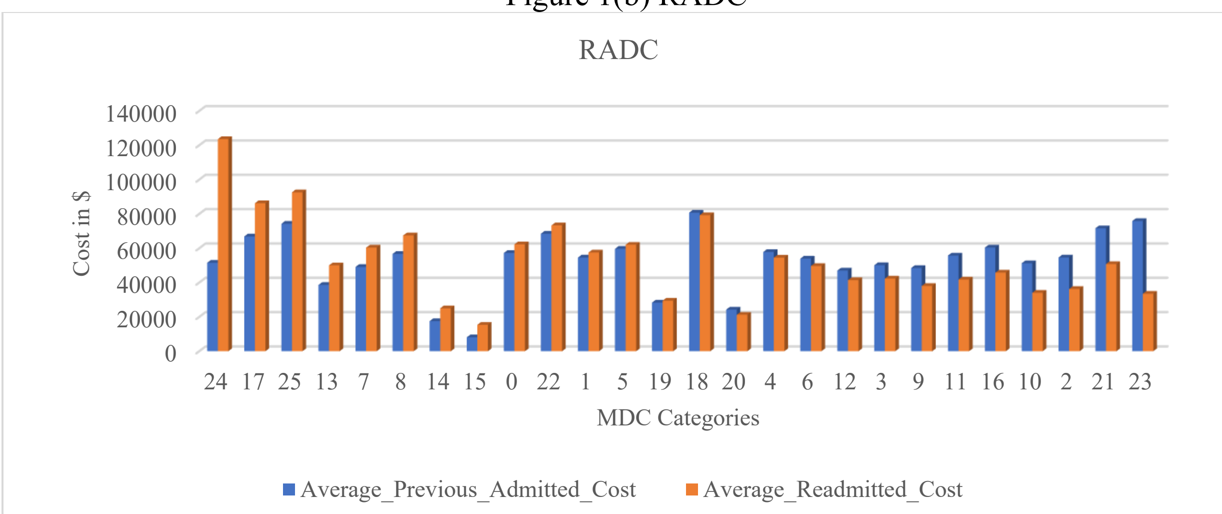


Figure 1(b) RADC



Based on the above analysis, I see that readmitted costs varied from previous admission costs at both individual and aggregated levels. Next, I try applying various, including state-of-the-art, predictive methods to model costs associated with readmissions at an individual level for both RSDC and RADC. One 10-fold cross validation was used, and the test results are shown in Table 4(a) for RSDC and Table 4(b) for RADC. The descriptive statistics of each fold of training and testing is included in Appendix D and Appendix E. The results of each fold for both training and testing for RSDC and RADC are included in the Appendix (Appendix F to U).

Table 4 Test Results Based on Different Performance Measures

Table 4(a) Test Results of RSDC

Model	MAPE (%)	RMSE	MAE	RAE	RRSE	NRMSE1	NRMSE2	MAD
Linear Regression	4.2679	0.5641	0.4313	0.5276	0.5456	0.0553	0.0549	0.0419
Lasso	4.2691	0.5642	0.4314	0.5276	0.5457	0.0553	0.0549	0.0419
Elastic Net	4.2692	0.5642	0.4314	0.5277	0.5457	0.0553	0.0549	0.0419
Ridge	4.2994	0.5651	0.4339	0.5309	0.5466	0.0554	0.0549	0.0422
XGBoost	14.6135	1.5413	1.5014	1.8368	1.4908	0.1512	0.1499	0.1460
Deep Learning	**3.2982**	**0.4374**	**0.3349**	**0.4098**	**0.4230**	**0.0429**	**0.0425**	**0.0325**

Table 4(b) Test Results of RADC

Model	MAPE (%)	RMSE	MAE	RAE	RRSE	NRMSE1	NRMSE2	MAD
Linear Regression	4.2076	0.5582	0.4265	0.5368	0.5538	0.0546	0.0541	0.0413
Lasso	4.2084	0.5582	0.4266	0.5369	0.5539	0.0546	0.0541	0.0414
Elastic Net	4.2085	0.5582	0.4266	0.5369	0.5539	0.0546	0.0541	0.0414
Ridge	4.2398	0.5591	0.4293	0.5309	0.5548	0.0547	0.0542	0.0416
XGBoost	14.5618	1.5387	1.5049	1.8893	1.5266	0.1510	0.1492	0.1455
Deep Learning	**3.2175**	**0.4265**	**0.3273**	**0.4119**	**0.4232**	**0.0420**	**0.0413**	**0.0317**

Figure 2 Model Comparisons Based on Various Metrics

Figure 2(a) Comparison of All Models Based on All Metrics Except MAPE for RSDC

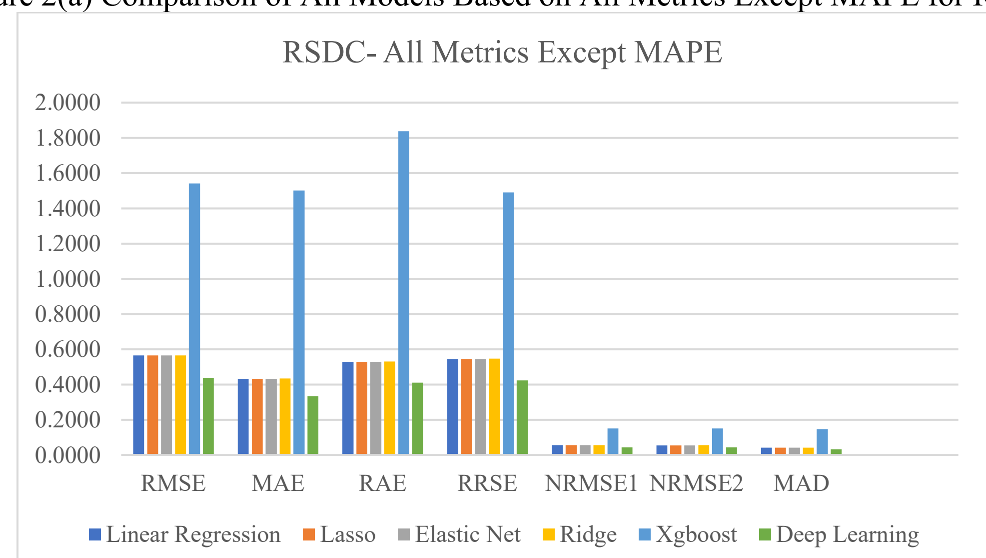


Figure 2(b) Comparison of All Models Based on All Metrics Except MAPE for RADC

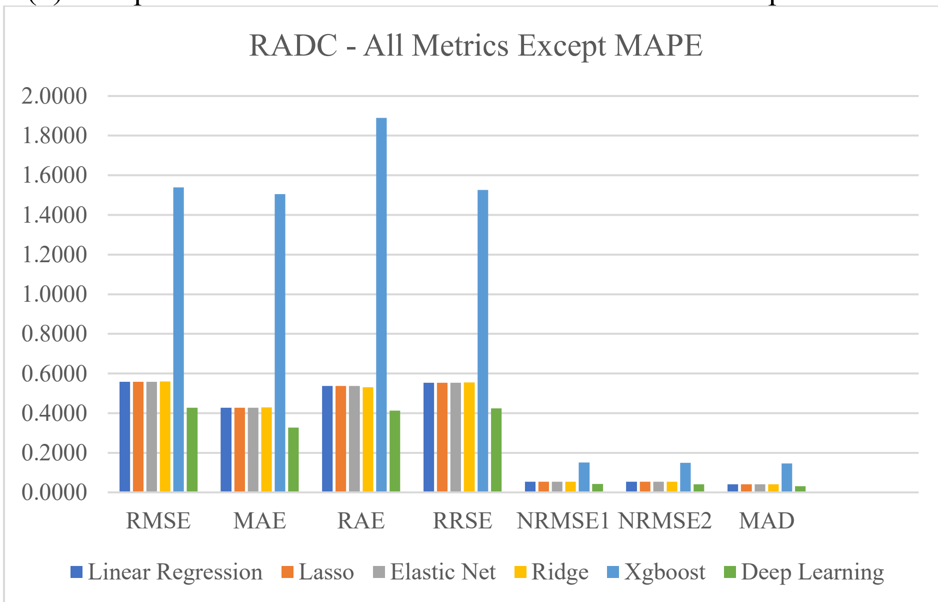


Figure 2(c) Comparison of All Models Based on MAPE for RSDC

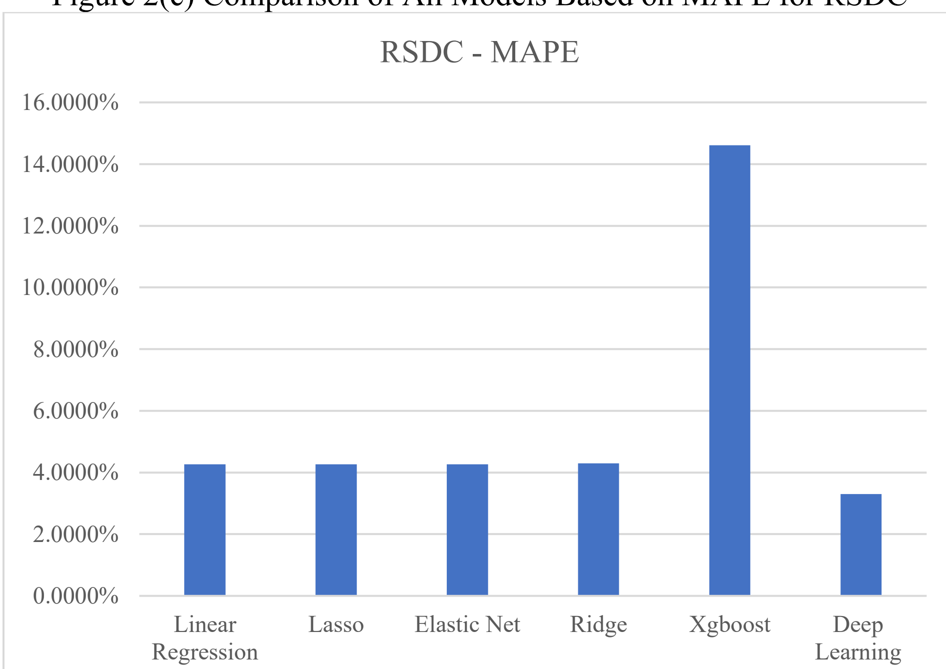
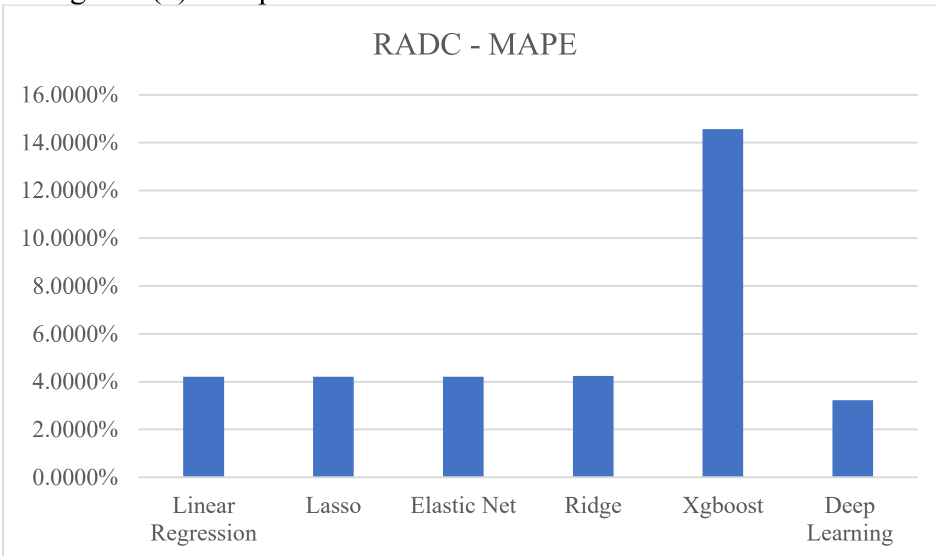


Figure 2(d) Comparison of All Models Based on MAPE for RADC



From Table 4(a), Table 4(b), and Figure 2, I find that the deep learning-based model performed the best for all performance metrics for both RADC and RADC. Some other observations were also made from the results. Firstly, models using regularization techniques (lasso regression, elastic net regression, and ridge regression) performed quite similar to linear regression. Secondly, XGBoost performed the worst in comparison with all other models for this application.

I also performed paired student t-test for each of the metrics to check if the difference between the deep learning model and linear regression model is statistically significant and found that the difference in every metric is significant ($p < 0.05$).

2.5. Contributions

This study makes two important contributions. First, to the best of my knowledge, this would be the first study to apply generalization techniques and deep learning-based models for predicting readmission costs. Deep learning-based models have proven useful in modeling health-related data. Prior studies, even if they predicted readmission costs, used only a few methods, such as linear regression and tree-based models. They also used narrow datasets (~10k samples) with limited features, whose applicability in different geographies is questionable, as they were based on hospitals from a specific region, which might have the issue of dataset drift as explained before.

Second, previous studies that predicted readmitted costs used only the all-cause definition of readmission. This study tried to redefine readmission using MDCs instead of DRGs by giving different criteria to MDC to see what kind of models would be suitable for predicting readmission costs even if the criteria for readmission changes, as readmission policies vary from one nation to another. This could help with generalization to know what kind of models are best suitable for predicting readmission costs even if readmission policies vary.

2.6. Implications

This study has several practical implications. First, this study would help individuals to plan their finances. If an individual has an estimate of the amount billed by the hospital for his/her future readmission, he/she can opt for the right insurance plan and the amount to be deposited to Health Savings Account (HSA). HSA is a tax-free savings account to pay for qualified medical expenses. If an individual opts for HSA combining with High Deductible Health Plan (HDHP), their premium will go down significantly. This would help individuals to save costs.

Second, predicting readmission costs could help hospitals with their financial planning. Under the Affordable Care Act, non-profit hospitals are obliged to provide a financial assistance policy and emergency care policy to low-income groups. Predicting readmission costs can help hospitals equip themselves better in terms of providing financial support to individuals who cannot afford readmission costs by planning with the loan providers armed with less risky estimates for readmission costs. Hospitals can also prepare their patients in advance about future costs that would be incurred by them. This would help patients in planning their own finances.

Third, this study can potentially be useful for healthcare policy makers. For example, after accurately predicting readmission costs at the individual level, the individuals can be put into cost buckets which would in turn be helpful for policy makers to identify high-cost readmitted individuals. Based on that, policy makers could implement new policies or modify existing policies related to readmissions.

2.7. Conclusions, Limitations, and Future Research

The readmission cost is one of the main contributors of the healthcare cost. However, the majority of previous studies have focused mainly on predicting early readmissions. There have been mixed reviews after policies like HRRP have been implemented, making it inconclusive if the healthcare cost has decreased. The goal of this study was to see if readmission costs, which showed to vary from initial admission costs, could be accurately predicted. Results revealed that deep learning-based model performed the best for all performance measures.

This study includes data only from a single year. So, if an individual is readmitted during the month of January, then the individual is considered as an initial admission as the starting month of the dataset would be January. This dataset cannot be linked to any other year or any other external database. Besides, in this study, I have focused only on readmission costs.

Modeling readmission likelihood and the length of stay are also important in the context of readmissions as these outcomes have influence on one another. So, joint modeling of readmission costs along with readmission likelihood and the length of stay might be more beneficial instead of focusing only on modeling readmission costs. In this essay, I had identified readmissions belonging to RSDC and RADC. For my future analysis, I will also deploy the term readmission as Readmission with Different Major Diagnostic Category (RDDC). RDDC will be defined as an admission to a hospital within 30 days of discharge from the same or another hospital with the cause of readmission being different. I will then build predictive models for RDDC. After that, I will compare the built predictive models for RDDC with the predictive models that were built for RSDC. Apart from that, in this essay, I have considered the standard defined categories of major diagnostic category as the cause of readmission. The standard defined categories of major diagnostic category belong to either a single organ system or an etiology. For my future study, I will consider categories that are correlated in terms of causing the set of related health complications that eventually lead to readmissions. These categories may span across multiple major diagnostic categories. I expect that such recategorizations should help in better prediction of costs. The recategorization in terms of correlated categories would be a significant contribution in the field of healthcare economics.

CHAPTER 3

Essay 2- A Multitask Learning Approach for Heterogenous Tasks With Specific Sub-population

3.1. Introduction

Digital revolution was brought into the healthcare industry by passing of the Health Information Technology for Economic and Clinical Health Act (HITECH) by the Congress as part of the American Recovery and Reinvestment Act of 2009. Financial incentives worth $19.2 billion were allocated for physicians and hospitals to implement electronic health record (EHR). The healthcare cost has continued to increase tremendously irrespective of the various efforts taken by the government to modernize the United States (U.S.) healthcare system. As per the data shared by the Center for Medicaid and Medicare Services (CMS), the national health expenditure reached $3.8 trillion in the year 2019 and has been projected to reach $6.2 trillion by 2028. It accounted for 17.7% of the nation's Gross Domestic Product (GDP) (Keehan et al. 2020).

One of the major contributors of these rising healthcare costs is attributed to readmissions. According to CMS, a readmission is defined as an admission to a hospital within 30 days of discharge from the same or another hospital irrespective of the cause of readmission (Commission 2007). In this essay, a readmission is defined as an admission to a hospital within 30 days of discharge from the same or a different hospital with the cause of readmission being the same. The cause of readmission could be grouped either based on diagnostic related groups (DRGs) or major diagnostic category (MDC). MDC is considered as the cause of readmission in this essay. Medicare spending for a year was $56,856 (60% higher) for individuals who were readmitted in comparison with non-readmitted individuals who only spent $35,365 in 2000-2011 (Zheng et al. 2019). Readmission policies vary from country to country. For example, readmission policies of

Germany focused mainly on preventing unintended readmissions of individual patients due to the introduction of DRG-based payments whereas readmission policies of U.S., Denmark, and England focused on both improvement in quality and readmission rates (Kristensen et al. 2015). Readmissions have also been identified as a measure of burden of illness instead of quality of care (Ansari et al. 2018). Prior studies have analyzed the relationship between the length of stay (LOS) and readmission rates as a reduction in LOS was found to be associated with an increase in unintended readmission rates after pancreatectomy (Carey and Lin 2014; Kohlnhofer et al. 2014; Mazmudar et al. 2018). The LOS is commonly used as an indicator for hospital performance and have also been studied with its relation to quality of care (Thomas et al. 1997).

With the increasing interest in big data and analytics and the sparsity of predictive analytics in the Information Systems literature, the use of innovative methods for predictive analytics has become increasingly relevant in Information Systems research (Shmueli and Koppius 2011). Machine learning techniques have been widely used in accurately predicting readmissions, general healthcare costs, length of stays, etc., as they help to improve diagnosis and prognosis, thus leading to improvement in healthcare outcomes (Bardhan et al. 2015; Bertsimas et al. 2008; Hachesu et al. 2013; Hon et al. 2016; Shams et al. 2015). However, most of these models focused on training single tasks independently and not training them simultaneously. Multitask learning-based methods improve generalization by training related tasks in parallel, thereby creating a shared representation (Caruana 1997). Even though a single-task learning may perform reasonably well, it does not consider the information that can be obtained from other related tasks, which might be helpful in learning better outcomes through multitask learning (Caruana 1997).

Deep learning, regularization, and Bayesian-based approaches can be used for multitask learning (Zhang and Yang 2021). Even though deep learning-based models perform superior, however, nowadays more emphasis is on explainable artificial intelligence (XAI) to convert black box models to glass models using model-specific or model-agnostic approaches, especially in the medical domain (Arrieta et al. 2020; Holzinger et al. 2017; Rai 2020). So, in this essay, I focus on using model-specific XAI, i.e., regularization-based multitask learning, to build models that could also be interpretable, rather than just focusing on improving the predictive accuracy. Only limited studies have used multitask learning in Information Systems research, and these studies have only used it for modeling homogenous tasks (classification) (Lin et al. 2017). In this essay, I define TI as the set of both readmitted and non-readmitted individuals. Similarly, I define RI as the subset of only readmitted individuals. Also, in this essay, heterogenous tasks means prediction tasks of both classification and regression. Predicting readmissions would involve TI whereas predicting readmission costs and the readmitted length of stay would involve RI. I use the term specific sub-population in this essay as one task involves TI whereas other two tasks involve RI.

In this essay, the following questions are addressed:

- How can heterogenous tasks with specific sub-population be modeled jointly?
- How did the joint model perform in comparison with single-task models for predicting readmissions, readmission costs, and the readmitted length of stay for different MDCs?
- What common predictor variables were chosen by the joint model to predict readmissions, readmission costs, and the readmitted length of stay for different MDCs?

The rest of the paper is organized as follows. In the next section, I review the existing literature related to predictive analytics in the context of readmissions, healthcare costs, and the length of stay, and explain the research gaps. Then, I explain the methodology used in this study and report on the results that were found. Finally, I discuss the implications, contributions, limitations, and directions for future research.

3.2. Literature Review

U.S. has implemented readmission policies to improve quality of care and reduce excess readmission rates. For example, the Hospital Readmission Reduction Program (HRRP) was established by the CMS under the Affordable Care Act beginning October 2012. It imposes penalties on hospitals with excess all-cause readmissions within 30-days for the following specific chronic conditions (McIlvennan et al. 2015):

* Acute Myocardial Infarction (AMI)

* Heart Failure (HF)

* Pneumonia

* Chronic Obstructive Pulmonary Disease (COPD)

* Elective Primary Total Hip Arthroplasty and/or Total knee Arthroplasty (THA/TKA)

* Coronary Artery Bypass Graft (CABG) surgery

However, the analysis after the implementation of the HRRP has shown mixed results. For example, some studies have found HRRP to be successful in decreasing readmissions whereas others have indicated an increase in mortality (Wadhera et al. 2018; Wasfy et al. 2017).

Predictive analytics for methods specific to readmissions, health care costs, length of stays, and multitask learning in healthcare can be broadly classified into different areas, namely, predicting all-cause 30 days readmissions for all/specific population, predictions specific to general healthcare/readmission costs, predictions specific to the length of stay, and methods used by the existing studies for multitask learning in healthcare.

Prior review study on risk prediction models for hospital readmissions showed that most of the current readmission risk prediction models did not perform well and more efforts are required to improve the predictive performance on readmissions (Kansagara et al. 2011). Studies have used logistic regression, support vector machines, neural networks, and tree-based methods to predict early readmissions (Cui et al. 2018; Schoonover et al. 2014; Shadmi et al. 2015; Sushmita et al. 2016; Wang et al. 2018; Xiao et al. 2018; Yu et al. 2015). These methods outperformed standard tools used by hospitals, such as LACE index and HOSPITAL scores (Sushmita et al. 2016). Recent models focus on creating models that not only have high accuracy but are also interpretable (Xiao et al. 2018). Studies trying to identify time to readmissions used Cox proportional hazard models (Schmutte et al. 2010). The above-mentioned studies were either related to an entire population or specific to a sub-population having congestive health failure, diabetes, psychosis, etc.

Previous studies have shown that data mining techniques were highly successful in predicting general healthcare costs (Bertsimas et al. 2008). Supervised learning techniques, including linear regression, lasso, gradient boosting, random forest, M5 model tree, and classification and regression tree (CART), have been used for predicting healthcare costs (Bertsimas et al. 2008; Duncan et al. 2016; Frees et al. 2013; Kuo et al. 2011; Morid et al. 2017; Sushmita et al. 2015). Prior healthcare costs have been found to be a good predictor for future healthcare costs (Sushmita et al. 2015). There are limited studies that focused on predicting readmission costs. Regression and M5 tree-based models have been used to predict readmission costs (Sushmita et al. 2016).

Supervised learning techniques, such as multiple linear regression, support vector machines, lasso multitask learning, and random forest, have been used to predict the long-term vs short-term length of stay of diabetic patients (Morton et al. 2014). Random forest has been found to be helpful in predicting the prolonged length of stay of general surgery patients (Chuang et al. 2018). Semisupervised techniques have also been found successful in predicting the length of stay with high accuracy compared to supervised learning techniques (Livieris et al. 2018). Regression models have been successful in predicting the length of stay at pediatric emergency departments (Combes et al. 2014).

Multitask learning with a Bayesian approach has been used for risk profiling in chronic care and diabetes (Lin et al. 2017; Liu et al. 2019). Deep learning-based multitask learning approaches have been used to predict patient mortality (Si and Roberts 2019; Suresh et al. 2018). Multitask learning has also been used to predict mortality of diverse rare diseases (Liu et al. 2020).

Based on the literature review, a few research gaps were identified. First, there are very limited studies related to prediction of readmission costs (Sushmita et al. 2016). Second, those existing studies have not jointly learned the predictive models for readmissions, readmission costs, and the readmitted length of stay. Third, even though multitask learning methods have been studied extensively in predictive analytics, to the best of my knowledge, there is no specific multitask learning-based method that focused on heterogenous tasks having specific sub-population. Also, there are very limited studies related to using multitask learning for heterogenous tasks (they did not have specific sub-population with respect to the tasks) (Yang et al. 2009).

3.3. Methodology

3.3.1. Dataset

Healthcare Cost and Utilization Project's Nationwide Readmission Database by Agency for Healthcare Research and Quality was used for analysis in this essay (Databases 2013). The total number of records for the year 2013 in the dataset was 14,325,172. Each admission record consists of demographic (gender, age, median household income, etc.), clinical information (diagnosis, procedure used, etc.), severity information (All Patients Refined Diagnosis Related Groups, i.e., APR DRG in terms of severity of illness, mortality, etc.), hospital details (bed size, teaching or non-teaching hospital, etc.), cost-related and administrative-related data (length of stay). Variables along with their descriptions used in this essay are given in Appendix V (Table 27).

3.3.2. Models

3.3.2.1. Baseline Models

The following models were used as the baseline models:

- Logistic regression for predicting readmissions
- Linear regression for predicting readmission costs
- Linear regression for predicting the length of stay

3.3.2.2. Joint Model

I define s as the index of the readmission task, k as the index of the cost billed by hospitals or the length of stay task, n_s as the total number of individuals, i.e., including both readmitted and non-readmitted, n_k as the total number of readmitted individuals, j as the index of individuals in n_s, l as the index of individuals in n_k, $Y1$ as whether an individual from n_s was readmitted or not, $Y2$ as the cost billed by hospitals for individuals from n_k, $Y3$ as the length of stay of an individual from n_k, $W1$ as the coefficient vector of the readmission task, $W2$ as the coefficient vector for the task of costs billed by hospitals, $W3$ as the coefficient vector for the length of stay task, X as the vector of predictors, $C1$ as the constant of the readmission task, $C2$ as the constant for the task of costs billed by hospitals, and $C3$ as the constant for the length of stay task.

The loss function for predicting the readmission task (L_r) is given as:

$$L_r = \frac{1}{n_s}\sum_{j=1}^{n_s} \log\left(1 + e^{-Y1_{s,j}\left(X_{s,j}W1_s^T + C1_s\right)}\right)$$

The loss function for predicting the readmission cost (L_{rc}) is given as:

$$L_{rc} = \frac{1}{n_k}\sum_{l=1}^{n_k} ||Y2_{k,l} - X_{k,l}W2_k^T - C2_k||_2^2$$

The loss function for predicting the readmitted length of stay (L_{rl}) is given as:

$$L_{rl} = \frac{1}{n_k}\sum_{l=1}^{n_k} ||Y3_{k,l} - X_{k,l}W3_k^T - C3_k||_2^2$$

The objective loss function of the model L_t (i.e., combined loss function for all three tasks) is formulated by combining the three empirical losses:

$$L_t = L_r + L_{rc} + L_{rl}$$

Joint feature selection with multitask learning methods has been used previously for modeling homogenous tasks (either classification or regression tasks) to create joint sparse representations (Cao et al. 2019; Evgeniou and Pontil 2007; Liu et al. 2012) . For modeling the joint heterogenous (classification and regression) problems for specific sub-population, multitask learning with joint feature selection is used by placing constraints on the loss function as follows:

$$Loss\ function = \min(\mathrm{L}_t) + \lambda_1 \Omega(W) + \lambda_2 ||W||_F^2$$

where

t is the number of tasks, i.e., 3 (readmission, readmission cost, and readmitted length of stay),

λ_1 illustrates the strength of the relatedness of tasks, i.e., effect of cross-task regularization,

λ_2 is the penalty of quadratic form of W for regularization,

$\Omega(W)$ is the $||W||_{2,1}$ used to create group sparse structure.

The matrix W has row values corresponding to features and column values corresponding to tasks. The $||W||_{2,1}$ penalizes the $(2,1)$ norm of the matrix W. The $(2,1)$ norm of the matrix W is obtained by first applying the $2-$norm (across the tasks) of the rows w_f corresponding to feature f of the matrix W and then applying 1 norm of the vector by adding the absolute sum of the coefficients of the matrix W (Evgeniou and Pontil 2007) . This ensures that variables that are consistently important to all three tasks (readmissions, readmission costs, and the readmitted length of stay) are selected. Nesterov's accelerated gradient descent method was used as the solver.

3.3.3. Performance Measures Used

Three verification statistics were used to evaluate the performance of the model. I define n as the total number of observations, i.e., including both readmitted and non-readmitted individuals/patients, n_i as the total number of readmission observations, i.e., only readmitted individuals/patients, y_{i1} as whether an individual i was readmitted or not, y_{i2} as the actual value of the cost billed by hospitals for the readmitted individual i, y_{i3} as the actual value of the length of stay of the readmitted individual i, $\hat{y}_{i1}$ as whether an individual i was predicted as being readmitted, $\hat{y}_{i2}$ as the predicted value of the cost billed by hospitals for the readmitted individual i, and $\hat{y}_{i3}$ as the predicted value of the length of stay for the readmitted individual i. Then, the measures are as follows:

Misclassification rate for readmission:

$$Misclassification\ rate = \frac{1}{n}\sum_{i=1}^{n} 1_{(y_{i1} \neq \hat{y}_{i1})}$$

Root Mean Squared Error (RMSE) for readmitted cost:

$$RMSE = \sqrt{\left(\frac{\sum_{i=1}^{n_i}(\hat{y}_{i2} - y_{i2})^2}{n_i}\right)}$$

Mean Absolute Error (MAE) for readmitted cost:

$$MAE = \frac{\sum_{i=1}^{n_i}|\hat{y}_{i2} - y_{i2}|}{n_i}$$

Root Mean Squared Error (RMSE) for readmitted length of stay:

$$RMSE = \sqrt{\left(\frac{\sum_{i=1}^{n_i}(\hat{y}_{i3} - y_{i3})^2}{n_i}\right)}$$

Mean Absolute Error (MAE) for readmitted length of stay:

$$MAE = \frac{\sum_{i=1}^{n_i}|\hat{y}_{i3} - y_{i3}|}{n_i}$$

3.4. Results

For my analysis, I considered top three major diagnostic categories belonging to readmissions with same major diagnostic categories identified in Essay 1. They are listed below:

- Diseases and Disorders of the Circulatory System
- Diseases and Disorders of the Respiratory System
- Diseases and Disorders of the Digestive System

I applied both baseline models and the joint model for the above three major diagnostic categories. I define the previous admission cost as the cost billed by the hospital for the previous admission. I define the readmission cost as the cost billed by the hospital for the admission when the readmission occurred within 30 days belonging to the same major diagnostic category. I define the previous admitted length of stay as the length of stay for the previous admission. I define the readmitted length of stay as the length of stay when the readmission occurred within 30 days belonging to the same major diagnostic category. Selective feature scaling was applied on all numeric independent variables, except for the previous admission cost and the previous admitted length of stay. Log transformation was applied on the previous admission cost and the readmission cost. Log (previous admitted length of stay + 1) transformation was applied on the previous admitted length of stay, and log (readmitted length of stay + 1) was applied on the readmitted length of stay, as they had zero values. The readmission cost and the readmitted length of stay were then divided by their means respectively to bring heterogenous tasks to the same scale. One 10-fold stratified cross-validation was used for training and testing. All together 15,000 samples were considered for each of the above-mentioned major diagnostic categories. Average testing errors are given in Table 5 for all three tasks. Individual training and testing errors for each fold of all three major diagnostic categories are given in the appendix (Appendix W, X, and Y). It was observed that the entire joint model (i.e., for all three tasks) performed slightly better for the Diseases and Disorders of the Digestive System whereas it performed slightly better only for specific tasks in case of Diseases and Disorder of the Circulatory System and Diseases and Disorders of the Respiratory System (see Figure 3 and Table 5). As the joint model was performing differently for each of top three major diagnostic categories, further analysis was done on predictors selected by joint models for these top three major diagnostic categories (Table 6). Apart from this, I plan to conduct future experiments in my next study to identify reasons for the difference in performance of the joint model for specific scenarios.

Table 5 Test Results of Top Three Major Diagnostic Category Groups

Table 5(a) Test Results of Diseases and Disorders of the Circulatory System

	RMSE		MAE		Misclassification Rate	
Tasks	Linear Regression	Joint Model	Linear Regression	Joint Model	Logistic Regression	Joint Model
Readmission	-	-	-	-	0.4331	**0.4213**
Readmission Cost	**0.1168**	0.1204	**0.0929**	0.0939	-	-
Readmitted Length of Stay	**0.4374**	0.4389	**0.3468**	0.3472	-	-

Table 5(b) Test Results of Diseases and Disorders of the Respiratory System

	RMSE		MAE		Misclassification Rate	
Tasks	Linear Regression	Joint Model	Linear Regression	Joint Model	Logistic Regression	Joint Model
Readmission	-	-	-	-	**0.4383**	0.4488
Readmission Cost	0.1859	**0.0943**	0.0769	**0.0728**	-	-
Readmitted Length of Stay	0.3778	**0.3756**	0.2938	**0.2913**	-	-

Table 5(c) Test Results of Diseases and Disorders of the Digestive System

	RMSE		MAE		Misclassification Rate	
Tasks	Linear Regression	Joint Model	Linear Regression	Joint Model	Logistic Regression	Joint Model
Readmission	-	-	-	-	0.4169	**0.4053**
Readmission Cost	0.1028	**0.1013**	0.0799	**0.0782**	-	-
Readmitted Length of Stay	0.3894	**0.3881**	0.3044	**0.3018**	-	-

Figure 3 Comparison of Single-task vs Multitask (Joint Model) for Various MDCs

Figure 3(a) Comparison of the Readmission Task

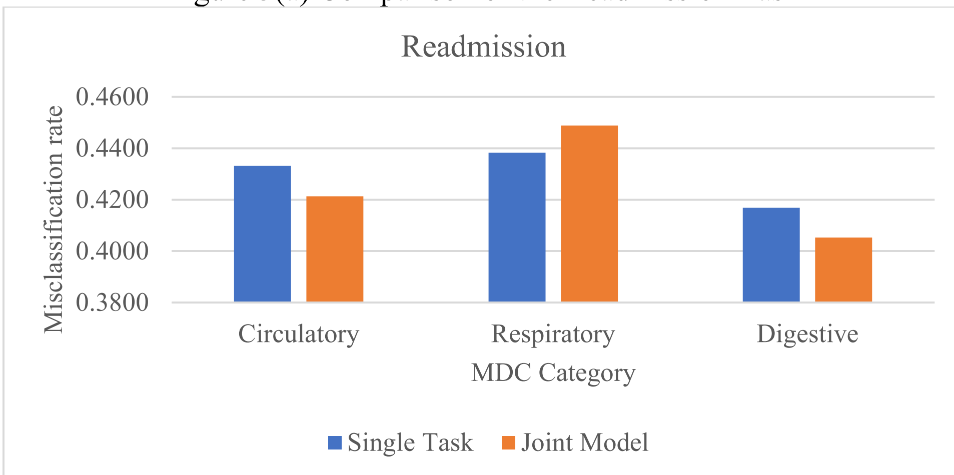


Figure 3(b) Comparison of Readmission Costs Based on RMSE

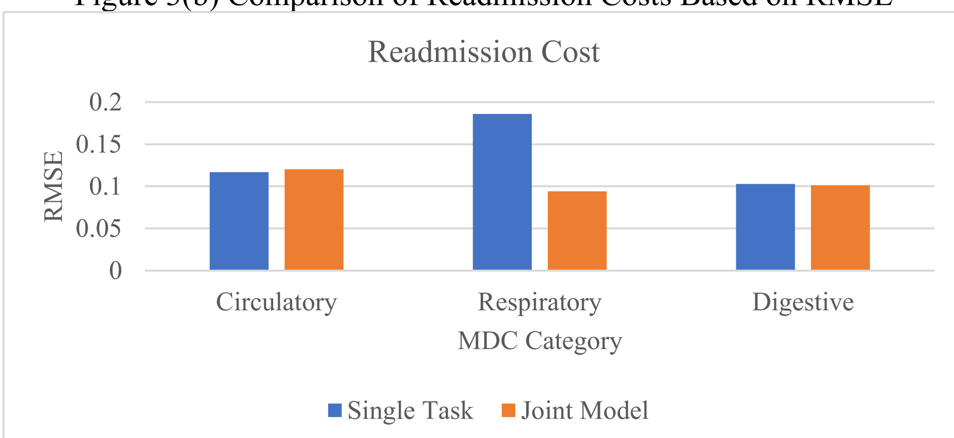


Figure 3(c) Comparison of Readmission Costs Based on MAE

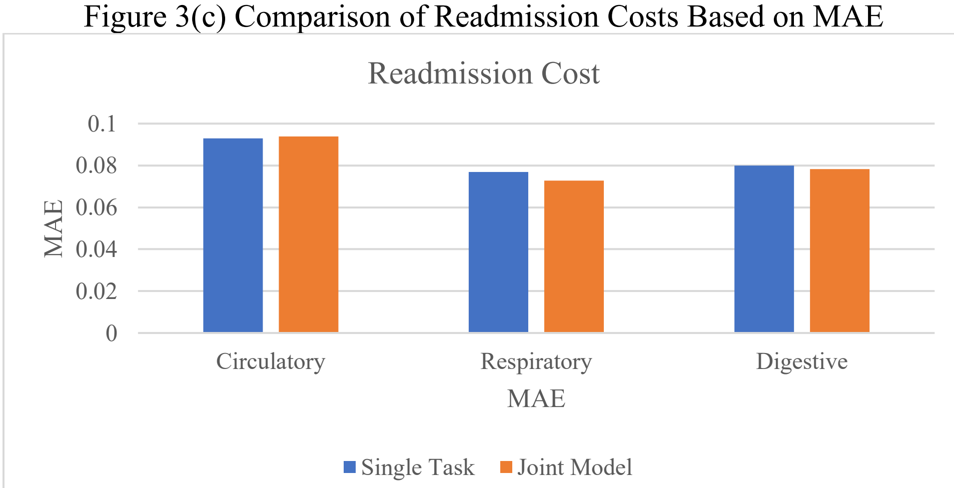


Figure 3(d) Comparison of the Readmitted Length of Stay Based on RMSE

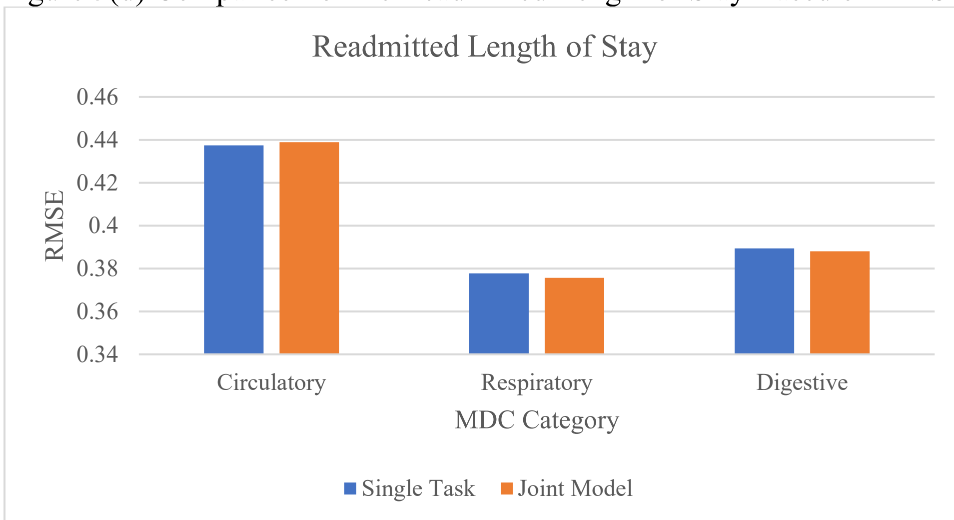


Figure 3(e) Comparison of the Readmitted Length of Stay Based on MAE

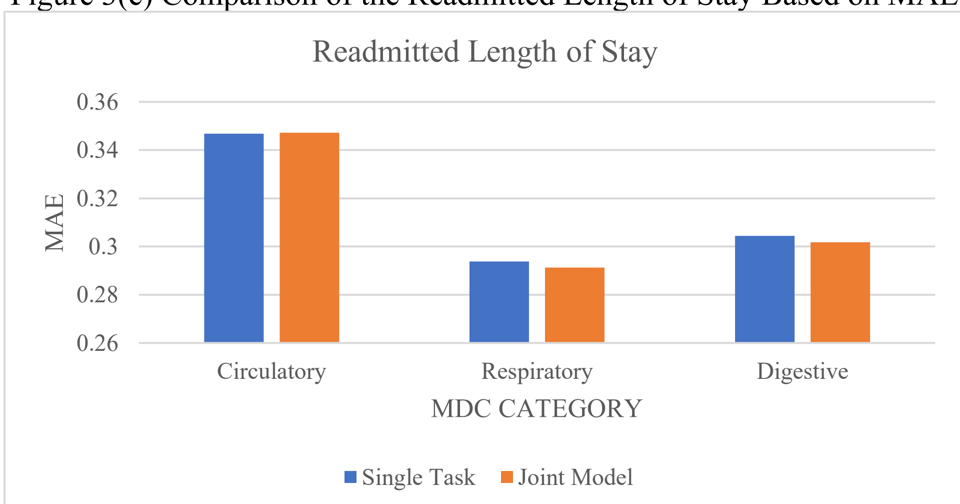


Table 6 Predictors Selected by the Joint Model for Top Three Major Diagnostic Categories

Variable	Diseases and Disorders of the Circulatory System	Diseases and Disorders of the Respiratory System	Diseases and Disorders of the Digestive System
AGE	x	x	x
AWEEKEND	x	x	x
DISCWT	x	x	x
DISPUNIFORM	x	✓	x
DQTR	✓	✓	✓
ELECTIVE	x	x	x
FEMALE	x	x	x
HCUP_ED	x	✓	✓
NCHRONIC	x	x	x
NDX	x	x	x
NECODE	x	x	x
NPR	x	x	x
ORPROC	x	x	x
PAY1	x	x	x
PL_NCHS	✓	✓	✓
REHABTRANSFER	x	x	x
RESIDENT	x	✓	✓
SAMEDAYEVENT	x	x	x
PREV_CHG	✓	✓	✓
ZIPINC_QRTL	x	✓	✓
PREV_LOS	x	✓	✓
HOSP_BEDSIZE	✓	✓	✓
H_CONTRL	x	✓	✓
HOSP_URCAT4	x	x	x
HOSP_UR_TEACH	x	x	x
TOTAL_DISC	x	x	x
APRDRG_RISK_MORTALITY	✓	✓	✓
APRDRG_SEVERITY	✓	✓	✓

3.5. Contributions and Implications

This study has research contributions as well as practical implications. I follow a design science approach where the motive is to develop an Information Technology artifact that would help to solve a practical problem (Hevner et al. 2004). To the best of my knowledge, this is the first study in Information Systems research that deploys a multitask learning-based method for heterogenous tasks with a specific sub-population – this is an important research contribution. The only prior multitask learning approach in Information Systems research was used for homogenous tasks (classification) (Lin et al. 2017). Specifically, in the domain of healthcare analytics, to the best of my knowledge, this is the first study to jointly learn the readmission probability, readmission cost, and readmitted length of stay using multitask learning.

Jointly predicting all three tasks would give us better estimates of effects of predictors than predicting them separately. Thus, the benefit of using a better model would obviously accrue from deploying the joint model. Apart from that, the regularization forces the joint model to use fewer variables. So, hospitals can now focus only on fewer variables that would have been jointly selected for all three tasks. As per Kaiser Health News, the CMS will penalize 2,545 hospitals for the fiscal year 2020 for excess readmissions of Medicare patients (Ellison 2020). So, predicting readmissions would be beneficial to hospitals to avoid penalties. Predicting readmission costs will be helpful for individuals to plan their finances, as a survey has shown that delaying care has been used as a strategy by three in ten Americans (Saad 2018). Federal and state laws provide rules to hospitals for charity care, which would provide financial assistance to individuals who are unable to pay for their treatment. Predicting readmission costs would be helpful for hospitals to provide better financial assistance to individuals. Similarly, predicting the readmitted length of stay would be helpful for hospitals to optimize hospital resources like bed allocations, costs, etc. As multitasking helps to learn better outcomes based on the information obtained from other similar tasks, it will give hospitals the benefit of getting better estimates for tasks in comparison to using single-task models. This will be shown using the counterfactual analysis in the future study.

3.6. Conclusion, Limitation, and Future Research

The joint model performed slightly better for Diseases and Disorders of the Digestive System in comparison to baseline models. As the joint model performed only slightly better for specific tasks in comparison to baseline models for the Disease and Disorders of the Circulatory System and Disease and Disorders of the Respiratory System, further analysis is needed to identify the cause of the difference in performance with respect to specific major diagnostic categories. In this essay, I tried to analyze the difference in behavior by identifying joint predictors with respect to major diagnostic categories (Table 6). The variables that were commonly used by the joint model across all three major diagnostic groups were All Patient Related Diagnostic Related Groups in terms of risk mortality and severity, the quarter in which an individual was admitted, the size of the hospital in terms of number of beds, the location of an individual in terms of urban-rural classification by the National Center for Health Statistics (NCHS), and the cost from previous admission. The variables that were additionally used by Diseases and Disorders of the Digestive System (the major diagnostic group for which the entire joint modeling worked well) were if the discharge included services from emergency department, the residency status (i.e., resident or non-resident) of an individual, median household income of an individual, the length of stay from the previous admission, and the ownership of the hospital (i.e., government, private, etc.).

The future analysis will include an in-depth study of the data and distribution of data, thereby providing recommendations on the way the data needs to be collected for multitask learning-based model to work better for all tasks with respect to each of these major diagnostic categories. The future analysis will also include applying multitask learning on different datasets to examine the generalizability of the current findings. Similarly, the conclusions are based on top three categories of readmissions for RSDC from Essay 1. In future research, I plan to test the joint model for readmissions based on RADC for top major diagnostic categories identified in Essay 1.

REFERENCES

Ansari, S.F., Yan, H., Zou, J., Worth, R.M., and Barbaro, N.M. 2018. "Hospital Length of Stay and Readmission Rate for Neurosurgical Patients," *Neurosurgery* (82:2), pp. 173-181.

Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., and Benjamins, R. 2020. "Explainable Artificial Intelligence (Xai): Concepts, Taxonomies, Opportunities and Challenges toward Responsible Ai," *Information Fusion* (58), pp. 82-115.

Atherton, J. 2011. "Development of the Electronic Health Record," *AMA Journal of Ethics* (13:3), pp. 186-189.

Bardhan, I., Oh, J.-h., Zheng, Z., and Kirksey, K. 2015. "Predictive Analytics for Readmission of Patients with Congestive Heart Failure," *Information Systems Research* (26:1), pp. 19-39.

Bates, D.W., Saria, S., Ohno-Machado, L., Shah, A., and Escobar, G. 2014. "Big Data in Health Care: Using Analytics to Identify and Manage High-Risk and High-Cost Patients," *Health Affairs* (33:7), pp. 1123-1131.

Bertsimas, D., Bjarnadóttir, M.V., Kane, M.A., Kryder, J.C., Pandey, R., Vempala, S., and Wang, G. 2008. "Algorithmic Prediction of Health-Care Costs," *Operations Research* (56:6), pp. 1382-1392.

Byrne, K. 2020. "Potential Average Cost of Covid-19 Treatment with Complications Could Be More Than $20,000."

Cai, X., Perez-Concha, O., Coiera, E., Martin-Sanchez, F., Day, R., Roffe, D., and Gallego, B. 2016. "Real-Time Prediction of Mortality, Readmission, and Length of Stay Using Electronic Health Record Data," *Journal of the American Medical Informatics Association* (23:3), pp. 553-561.

Campanella, P., Lovato, E., Marone, C., Fallacara, L., Mancuso, A., Ricciardi, W., and Specchia, M.L. 2016. "The Impact of Electronic Health Records on Healthcare Quality: A Systematic Review and Meta-Analysis," *The European Journal of Public Health* (26:1), pp. 60-64.

Cao, H., Zhou, J., and Schwarz, E. 2019. "Rmtl: An R Library for Multi-Task Learning," *Bioinformatics* (35:10), pp. 1797-1798.

Carey, K., and Lin, M.-Y. 2014. "Hospital Length of Stay and Readmission: An Early Investigation," *Medical Care Research and Review* (71:1), pp. 99-111.

Caruana, R. 1997. "Multitask Learning," *Machine Learning* (28:1), pp. 41-75.

Chuang, M.T., Hu, Y.h., and Lo, C.L. 2018. "Predicting the Prolonged Length of Stay of General Surgery Patients: A Supervised Learning Approach," *International Transactions in Operational Research* (25:1), pp. 75-90.

Combes, C., Kadri, F., and Chaabane, S. 2014. "Predicting Hospital Length of Stay Using Regression Models: Application to Emergency Department," *10ème Conférence Francophone de Modélisation, Optimisation et Simulation-MOSIM'14*.

Commission, M.P.A. 2007. "Payment Policy for Inpatient Readmissions," *Report to the Congress: Promoting Greater Efficiency in Medicare*: MedPAC Washington, DC, pp. 102-120.

Cotter, P.E., Bhalla, V.K., Wallis, S.J., and Biram, R.W. 2012. "Predicting Readmissions: Poor Performance of the Lace Index in an Older Uk Population," *Age and Ageing* (41:6), pp. 784-789.

Cui, S., Wang, D., Wang, Y., Yu, P.-W., and Jin, Y. 2018. "An Improved Support Vector Machine-Based Diabetic Readmission Prediction," *Computer Methods and Programs in Biomedicine* (166), pp. 123-135.

Databases, H. 2013. "Healthcare Cost and Utilization Project (Hcup)." Agency for Healthcare Research and Quality Rockville, MD.

Duncan, I., Loginov, M., and Ludkovski, M. 2016. "Testing Alternative Regression Frameworks for Predictive Modeling of Health Care Costs," *North American Actuarial Journal* (20:1), pp. 65-87.

Ellison, A. 2020. "Cms Fines 2,545 Hospitals for High Readmissions: 5 Things to Know." from https://www.beckershospitalreview.com/finance/cms-fines-2-545-hospitals-for-high-readmissions-5-things-to-know.html

Evgeniou, A., and Pontil, M. 2007. "Multi-Task Feature Learning," *Advances in Neural Information Processing Systems* (19), p. 41.

Farley, J.F., Harley, C.R., and Devine, J.W. 2006. "A Comparison of Comorbidity Measurements to Predict Healthcare Expenditures," *American Journal of Managed Care* (12:2), pp. 110-118.

Fine, M.J., Pratt, H.M., Obrosky, D.S., Lave, J.R., McIntosh, L.J., Singer, D.E., Coley, C.M., and Kapoor, W.N. 2000. "Relation between Length of Hospital Stay and Costs of Care for Patients with Community-Acquired Pneumonia," *The American Journal of Medicine* (109:5), pp. 378-385.

Fleishman, J.A., and Cohen, J.W. 2010. "Using Information on Clinical Conditions to Predict High-Cost Patients," *Health Services Research* (45:2), pp. 532-552.

Frakt, A. 2016. "The Hidden Financial Incentives Behind Your Shorter Hospital Stay." *The New Health Care* Retrieved 03/05/2020, 2020, from https://www.nytimes.com/2016/01/05/upshot/the-hidden-financial-incentives-behind-your-shorter-hospital-stay.html

Frees, E.W., Jin, X., and Lin, X. 2013. "Actuarial Applications of Multivariate Two-Part Regression Models," *Annals of Actuarial Science* (7:2), p. 258.

Greenblatt, D.Y., Weber, S.M., O'Connor, E.S., LoConte, N.K., Liou, J.-I., and Smith, M.A. 2010. "Readmission after Colectomy for Cancer Predicts One-Year Mortality," *Annals of Surgery* (251:4), p. 659.

Gupta, A., Allen, L.A., Bhatt, D.L., Cox, M., DeVore, A.D., Heidenreich, P.A., Hernandez, A.F., Peterson, E.D., Matsouaka, R.A., and Yancy, C.W. 2018. "Association of the Hospital Readmissions Reduction Program Implementation with Readmission and Mortality Outcomes in Heart Failure," *JAMA Cardiology* (3:1), pp. 44-53.

Hachesu, P.R., Ahmadi, M., Alizadeh, S., and Sadoughi, F. 2013. "Use of Data Mining Techniques to Determine and Predict Length of Stay of Cardiac Patients," *Healthcare Informatics Research* (19:2), p. 121.

Hammoudeh, A., Al-Naymat, G., Ghannam, I., and Obied, N. 2018. "Predicting Hospital Readmission among Diabetics Using Deep Learning," *Procedia Computer Science* (141), pp. 484-489.

Hasan, O., Meltzer, D.O., Shaykevich, S.A., Bell, C.M., Kaboli, P.J., Auerbach, A.D., Wetterneck, T.B., Arora, V.M., Zhang, J., and Schnipper, J.L. 2010. "Hospital Readmission in General Medicine Patients: A Prediction Model," *Journal of General Internal Medicine* (25:3), pp. 211-219.

Hevner, A.R., March, S.T., Park, J., and Ram, S. 2004. "Design Science in Information Systems Research," *MIS Quarterly*, pp. 75-105.

Hines, A.L., Barrett, M.L., Jiang, H.J., and Steiner, C.A. 2014. "Conditions with the Largest Number of Adult Hospital Readmissions by Payer, 2011: Statistical Brief# 172,".

Holzinger, A., Biemann, C., Pattichis, C.S., and Kell, D.B. 2017. "What Do We Need to Build Explainable Ai Systems for the Medical Domain?," *arXiv preprint arXiv:1712.09923*.

Hon, C.P., Pereira, M., Sushmita, S., Teredesai, A., and De Cock, M. 2016. "Risk Stratification for Hospital Readmission of Heart Failure Patients: A Machine Learning Approach," *Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, pp. 491-492.

Jamei, M., Nisnevich, A., Wetchler, E., Sudat, S., and Liu, E. 2017. "Predicting All-Cause Risk of 30-Day Hospital Readmission Using Artificial Neural Networks," *PloS one* (12:7).

Jamoom, E., and Yang, N. 2016. "Table of Electronic Health Record Adoption and Use among Office-Based Physicians in the Us, by State: 2015 National Electronic Health Records Survey," *Hyattsville, MD: National Center for Health Statistics*.

Jovanovic, M., Radovanovic, S., Vukicevic, M., Van Poucke, S., and Delibasic, B. 2016. "Building Interpretable Predictive Models for Pediatric Hospital Readmission Using Tree-Lasso Logistic Regression," *Artificial intelligence in Medicine* (72), pp. 12-21.

Kansagara, D., Englander, H., Salanitro, A., Kagen, D., Theobald, C., Freeman, M., and Kripalani, S. 2011. "Risk Prediction Models for Hospital Readmission: A Systematic Review," *JAMA* (306:15), pp. 1688-1698.

Keehan, S.P., Cuckler, G.A., Poisal, J.A., Sisko, A.M., Smith, S.D., Madison, A.J., Rennie, K.E., Fiore, J.A., and Hardesty, J.C. 2020. "National Health Expenditure Projections, 2019–28: Expected Rebound in Prices Drives Rising Spending Growth: National Health Expenditure Projections for the Period 2019–2028," *Health Affairs* (39:4), pp. 704-714.

Kelly, C.J., Karthikesalingam, A., Suleyman, M., Corrado, G., and King, D. 2019. "Key Challenges for Delivering Clinical Impact with Artificial Intelligence," *BMC medicine* (17:1), pp. 1-9.

Kelly, M., Sharp, L., Dwane, F., Kelleher, T., and Comber, H. 2012. "Factors Predicting Hospital Length-of-Stay and Readmission after Colorectal Resection: A Population-Based Study of Elective and Emergency Admissions," *BMC Health Services Research* (12:1), p. 77.

Kohlnhofer, B.M., Tevis, S.E., Weber, S.M., and Kennedy, G.D. 2014. "Multiple Complications and Short Length of Stay Are Associated with Postoperative Readmissions," *The American Journal of Surgery* (207:4), pp. 449-456.

Kommers, A.-M. 2019. "Average Hospital Readmission Costs for 18 Diagnoses." Retrieved 03/19/2020, 2020, from https://www.beckershospitalreview.com/rankings-and-ratings/average-hospital-readmission-costs-for-18-diagnoses.html

Kristensen, S.R., Bech, M., and Quentin, W. 2015. "A Roadmap for Comparing Readmission Policies with Application to Denmark, England, Germany and the United States," *Health Policy* (119:3), pp. 264-273.

Kuehn, B.M. 2020. "Hospital Readmission Is Common among Covid-19 Survivors," *JAMA* (324:24), pp. 2477-2477.

Kuo, R.N., Dong, Y.-H., Liu, J.-P., Chang, C.-H., Shau, W.-Y., and Lai, M.-S. 2011. "Predicting Healthcare Utilization Using a Pharmacy-Based Metric with the Who's Anatomic Therapeutic Chemical Algorithm," *Medical Care*, pp. 1031-1039.

LaPointe, J. 2020. "Cdc: 1 in 11 Covid-19 Inpatients Experience a Hospital Readmission." from https://revcycleintelligence.com/news/cdc-1-in-11-covid-19-inpatients-experience-a-hospital-readmission

Leigh, J.P., Hubert, H.B., and Romano, P.S. 2005. "Lifestyle Risk Factors Predict Healthcare Costs in an Aging Cohort," *American Journal of Preventive Medicine* (29:5), pp. 379-387.

Lin, Y.-K., Chen, H., Brown, R.A., Li, S.-H., and Yang, H.-J. 2017. "Healthcare Predictive Analytics for Risk Profiling in Chronic Care: A Bayesian Multitask Learning Approach," *MIS Quarterly* (41:2).

Liu, B., Li, Y., Ghosh, S., Sun, Z., Ng, K., and Hu, J. 2019. "Complication Risk Profiling in Diabetes Care: A Bayesian Multi-Task and Feature Relationship Learning Approach," *IEEE Transactions on Knowledge and Data Engineering* (32:7), pp. 1276-1289.

Liu, J., Ji, S., and Ye, J. 2012. "Multi-Task Feature Learning Via Efficient L2, 1-Norm Minimization," *arXiv preprint arXiv:1205.2631*.

Liu, L., Liu, Z., Wu, H., Wang, Z., Shen, J., Song, Y., and Zhang, M. 2020. "Multi-Task Learning Via Adaptation to Similar Tasks for Mortality Prediction of Diverse Rare Diseases," *arXiv preprint arXiv:2004.05318*.

Livieris, I.E., Dimopoulos, I.F., Kotsilieris, T., and Pintelas, P. 2018. "Predicting Length of Stay in Hospitalized Patients Using Ssl Algorithms," *Proceedings of the 8th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion*, pp. 16-22.

Mainous, A.G., Diaz, V.A., Everett, C.J., and Knoll, M.E. 2011. "Impact of Insurance and Hospital Ownership on Hospital Length of Stay among Patients with Ambulatory Care–Sensitive Conditions," *The Annals of Family Medicine* (9:6), pp. 489-495.

Mazmudar, A., Castle, J., Yang, A.D., and Bentrem, D.J. 2018. "The Association of Length of Hospital Stay with Readmission after Elective Pancreatic Resection," *Journal of Surgical Oncology* (118:1), pp. 7-14.

McIlvennan, C.K., Eapen, Z.J., and Allen, L.A. 2015. "Hospital Readmissions Reduction Program," *Circulation* (131:20), pp. 1796-1803.

Morid, M.A., Kawamoto, K., Ault, T., Dorius, J., and Abdelrahman, S. 2017. "Supervised Learning Methods for Predicting Healthcare Costs: Systematic Literature Review and Empirical Evaluation," *AMIA Annual Symposium Proceedings*: American Medical Informatics Association, p. 1312.

Morris, P.E., Griffin, L., Thompson, C., Hite, R.D., Haponik, E., Berry, M., Winkelman, C., Hopkins, R.O., Ross, A., and Dixon, L. 2011. "Receiving Early Mobility During an Intensive Care Unit Admission Is a Predictor of Improved Outcomes in Acute Respiratory Failure," *The American Journal of the Medical Sciences* (341:5), pp. 373-377.

Morton, A., Marzban, E., Giannoulis, G., Patel, A., Aparasu, R., and Kakadiaris, I.A. 2014. "A Comparison of Supervised Machine Learning Techniques for Predicting Short-Term in-Hospital Length of Stay among Diabetic Patients," *2014 13th International Conference on Machine Learning and Applications*: IEEE, pp. 428-431.

Mudge, A.M., Kasper, K., Clair, A., Redfern, H., Bell, J.J., Barras, M.A., Dip, G., and Pachana, N.A. 2011. "Recurrent Readmissions in Medical Patients: A Prospective Study," *Journal of Hospital Medicine* (6:2), pp. 61-67.

OECD. 2019. *Health at a Glance 2019*.

OECD. 2020. "Oecd (2020), Length of Hospital Stay (Indicator)." Retrieved 16 April 2020

Piccialli, F., Di Somma, V., Giampaolo, F., Cuomo, S., and Fortino, G. 2021. "A Survey on Deep Learning in Medicine: Why, How and When?," *Information Fusion* (66), pp. 111-137.

Picker, D., Heard, K., Bailey, T.C., Martin, N.R., LaRossa, G.N., and Kollef, M.H. 2015. "The Number of Discharge Medications Predicts Thirty-Day Hospital Readmission: A Cohort Study," *BMC Health Services Research* (15:1), p. 282.

Plantier, M., Havet, N., Durand, T., Caquot, N., Amaz, C., Biron, P., Philip, I., and Perrier, L. 2017. "Does Adoption of Electronic Health Records Improve the Quality of Care Management in France? Results from the French E-Si (Preps-Sips) Study," *International Journal of Medical Informatics* (102), pp. 156-165.

Radovanovic, S., Vukicevic, M., Kovacevic, A., Stiglic, G., and Obradovic, Z. 2015. "Domain Knowledge Based Hierarchical Feature Selection for 30-Day Hospital Readmission Prediction," *Conference on Artificial Intelligence in Medicine in Europe*: Springer, pp. 96-100.

Rai, A. 2020. "Explainable Ai: From Black Box to Glass Box," *Journal of the Academy of Marketing Science* (48:1), pp. 137-141.

Saad, L. 2018. "Delaying Care a Healthcare Strategy for Three in 10 Americans." 2021, from https://news.gallup.com/poll/245486/delaying-care-healthcare-strategy-three-americans.aspx

Schmutte, T., Dunn, C.L., and Sledge, W.H. 2010. "Predicting Time to Readmission in Patients with Recent Histories of Recurrent Psychiatric Hospitalization: A Matched-Control Survival Analysis," *The Journal of Nervous and Mental Disease* (198:12), pp. 860-863.

Schoonover, H., Corbett, C.F., Weeks, D.L., Willson, M.N., and Setter, S.M. 2014. "Predicting Potential Postdischarge Adverse Drug Events and 30-Day Unplanned Hospital Readmissions from Medication Regimen Complexity," *Journal of Patient Safety* (10:4), pp. 186-191.

Shadmi, E., Flaks-Manov, N., Hoshen, M., Goldman, O., Bitterman, H., and Balicer, R.D. 2015. "Predicting 30-Day Readmissions with Preadmission Electronic Health Record Data," *Medical Care* (53:3), pp. 283-289.

Shameer, K., Johnson, K.W., Yahi, A., Miotto, R., Li, L., Ricks, D., Jebakaran, J., Kovatch, P., Sengupta, P.P., and Gelijns, S. 2017. "Predictive Modeling of Hospital Readmission Rates Using Electronic Medical Record-Wide Machine Learning: A Case-Study Using Mount Sinai Heart Failure Cohort," *Pacific Symposium on Biocomputing 2017*: World Scientific, pp. 276-287.

Shams, I., Ajorlou, S., and Yang, K. 2015. "A Predictive Analytics Approach to Reducing 30-Day Avoidable Readmissions among Patients with Heart Failure, Acute Myocardial Infarction, Pneumonia, or Copd," *Healthcare Management Science* (18:1), pp. 19-34.

Shmueli, G., and Koppius, O.R. 2011. "Predictive Analytics in Information Systems Research," *MIS Quarterly*, pp. 553-572.

Si, Y., and Roberts, K. 2019. "Deep Patient Representation of Clinical Notes Via Multi-Task Learning for Mortality Prediction," *AMIA Summits on Translational Science Proceedings* (2019), p. 779.

Sud, M., Yu, B., Wijeysundera, H.C., Austin, P.C., Ko, D.T., Braga, J., Cram, P., Spertus, J.A., Domanski, M., and Lee, D.S. 2017. "Associations between Short or Long Length of Stay and 30-Day Readmission and Mortality in Hospitalized Patients with Heart Failure," *JACC: Heart Failure* (5:8), pp. 578-588.

Suresh, H., Gong, J.J., and Guttag, J.V. 2018. "Learning Tasks for Multitask Learning: Heterogenous Patient Populations in the Icu," *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 802-810.

Sushmita, S., Khulbe, G., Hasan, A., Newman, S., Ravindra, P., Roy, S.B., De Cock, M., and Teredesai, A. 2016. "Predicting 30-Day Risk and Cost of" All-Cause" Hospital Readmissions," *30th AAAI Conference on Artificial Intelligence, AAAI 2016*: AI Access Foundation, pp. 453-461.

Sushmita, S., Newman, S., Marquardt, J., Ram, P., Prasad, V., Cock, M.D., and Teredesai, A. 2015. "Population Cost Prediction on Public Healthcare Datasets," *Proceedings of the 5th International Conference on Digital Health 2015*, pp. 87-94.

Tabata, M., Shimizu, R., Kamekawa, D., Kato, M., Kamiya, K., Akiyama, A., Kamada, Y., Tanaka, S., Noda, C., and Masuda, T. 2014. "Six-Minute Walk Distance Is an Independent Predictor of Hospital Readmission in Patients with Chronic Heart Failure," *International Heart Journal*, pp. 13-224.

Thomas, J.W., Guire, K.E., and Horvat, G.G. 1997. "Is Patient Length of Stay Related to Quality of Care?," *Journal of Healthcare Management* (42:4), p. 489.

Wadhera, R.K., Maddox, K.E.J., Wasfy, J.H., Haneuse, S., Shen, C., and Yeh, R.W. 2018. "Association of the Hospital Readmissions Reduction Program with Mortality among Medicare Beneficiaries Hospitalized for Heart Failure, Acute Myocardial Infarction, and Pneumonia," *JAMA* (320:24), pp. 2542-2552.

Wang, H., Cui, Z., Chen, Y., Avidan, M., Abdallah, A.B., and Kronzer, A. 2018. "Predicting Hospital Readmission Via Cost-Sensitive Deep Learning," *IEEE/ACM Transactions on Computational Biology and Bioinformatics* (15:6), pp. 1968-1978.

Wasfy, J.H., Zigler, C.M., Choirat, C., Wang, Y., Dominici, F., and Yeh, R.W. 2017. "Readmission Rates after Passage of the Hospital Readmissions Reduction Program: A Pre–Post Analysis," *Annals of Internal Medicine* (166:5), pp. 324-331.

Wilson, L. 2019. "Ma Patients' Readmission Rates Higher Than Traditional Medicare, Study Finds." Retrieved 01/11/2021, 2021, from https://www.healthcaredive.com/news/ma-patients-readmission-rates-higher-than-traditional-medicare-study-find/557694/

Xiao, C., Ma, T., Dieng, A.B., Blei, D.M., and Wang, F. 2018. "Readmission Prediction Via Deep Contextual Embedding of Clinical Concepts," *PloS one* (13:4), p. e0195024.

Yang, X., Kim, S., and Xing, E. 2009. "Heterogeneous Multitask Learning with Joint Sparsity Constraints," *Advances in Neural Information Processing Systems* (22), pp. 2151-2159.

Yu, S., Farooq, F., Van Esbroeck, A., Fung, G., Anand, V., and Krishnapuram, B. 2015. "Predicting Readmission Risk with Institution-Specific Prediction Models," *Artificial Intelligence in Medicine* (65:2), pp. 89-96.

Zhang, Y., and Yang, Q. 2021. "A Survey on Multi-Task Learning," *IEEE Transactions on Knowledge and Data Engineering*.

Zheng, B., Zhang, J., Yoon, S.W., Lam, S.S., Khasawneh, M., and Poranki, S. 2015. "Predictive Modeling of Hospital Readmissions Using Metaheuristics and Data Mining," *Expert Systems with Applications* (42:20), pp. 7110-7120.

Zheng, S., Hanchate, A., and Shwartz, M. 2019. "One-Year Costs of Medical Admissions with and without a 30-Day Readmission and Enhanced Risk Adjustment," *BMC Health Services Research* (19:1), pp. 1-10.

APPENDIX A:

Literature Review of Predictive Analytics on Readmissions and Healthcare Costs

Table 7 Detailed Literature Review of Essay 1

Study	**Models Used**	**Dataset**
Predicting All-cause Hospital Readmissions		
(Wang et al. 2018)	Convolution neural network and multilayer perceptron	Barnes Jewish Hospital
(Jamei et al. 2017)	Artificial neural network	Hospital stays in California from Sutter Health's EHR system
(Cai et al. 2016)	Bayesian analysis	Sydney metropolitan hospital
	For readmission analysis: support vector machine, decision trees, random forest, logistic regression, generalized boosting model For readmission cost: Linear regression and tree-based models	Chain of northwestern hospital
(Picker et al. 2015)	Multiple logistic regression	Barnes Jewish Hospital
(Zheng et al. 2015)	Support vector machine, neural network, and random forest	Data derived from medical records
(Yu et al. 2015)	Support vector machine, and Cox regression model	Three U.S. based hospitals
(Shadmi et al. 2015)	Decision tree, and neural network	Clalit Health Services, admitted to internal medicine ward in Israel
(Schoonover et al. 2014)	Logistic regression	-
(Morris et al. 2011)	Multiple logistic regression	Academic medical center
Predicting Readmissions Specific to a Population		
(Shameer et al. 2017)	Naïve Bayes	Mount Sinai Heart Failure Cohort
(Jovanovic et al. 2016)	Tree lasso logistic regression	Hospital discharge records from California
(Bardhan et al. 2015)	Beta geometric Erlang-2 model	67 hospitals from North Texas
(Radovanovic et al. 2015)	Lasso regularization with group-level feature selection	Pediatric patient data from California (Healthcare Cost Utilization Project)
(Tabata et al. 2014)	Multivariate logistic regression	Kitasato University Hospital
(Cotter et al. 2012)	Logistic regression	Department of Medicine for the elderly
(Kelly et al. 2012)	Logistic regression	All acute public hospitals in Ireland
(Hasan et al. 2010)	Logistic regression	General medicine services from six academic medical centers
(Greenblatt et al. 2010)	Multivariate logistic regression	SEER
Predicting Time to Readmissions		
(Schmutte et al. 2010)	Multivariate Cox proportional hazard model	A large public-private health system

Predicting General Healthcare Costs		
(Sushmita et al. 2015)	M5 model tree	State in-patient database and survey specific to Washington State
(Bertsimas et al. 2008)	Classification trees, and clustering	Commercially insured population
(Farley et al. 2006)	Linear regression	Claims data from managed care organization
(Leigh et al. 2005)	Linear regression	Annual survey
Predicting Patients With High Healthcare Costs		
(Fleishman and Cohen 2010)	Logistic regression	Medical expenditure panel survey

APPENDIX B:

Variables Used in Essay 1 Along With Their Descriptions and Descriptive Statistics

(Variable Names and Their Explanations are Retrieved From AHRQ Website)[2]

Table 8 Variable Descriptions and Descriptive Statistics of Essay 1

Variable Name	**Explanation**	**Descriptive Statistics for RSDC**	**Descriptive Statistics for RADC**
AGE	Age in years of a patient	56 (mean)	60 (mean)
AWEEKEND	Indicates if the admission took place on a weekend: (1) yes, (0) no	0 – 78.72% 1 – 21.28%	0 – 77.79% 1 – 22.21%
DISCWT	NRD discharge weight to be used for calculating national estimates	2.38 (mean)	2.36(mean)
DISPUNIFORM	Indicates the disposition status of a patient: (1) routine, (2) transfer to short term hospital, (5) other transfers, including skilled nursing facility, intermediate care, and another type of facility, (6) home health care, (7) against medical advice, (20) died in hospital, (99) discharged alive, destination unknown	1 – 59.75% (highest category) 99 – 0.11% (lowest category)	1 – 51.01% (highest category) 99 – 0.15% (lowest category)
DQTR	Indicates the quarter of the year: (1) Jan–Mar, (2) Apr–Jun, (3) Jul–Sep, (4) Oct–Dec	1- 19.98% 2- 25.36% 3- 26% 4- 28.66%	1- 19.75% 2- 24.92% 3- 25.76% 4- 29.58%
ELECTIVE	Indicates the elective status of an admission: (1) yes, (0) no	0 – 83.61% 1- 16.39%	0 – 87.41% 1- 12.59%
FEMALE	Indicates the sex of a patient:(0) male, (1) female.	0- 46.26% 1- 53.74%	0- 46.55% 1- 53.45%
HCUP_ED	Indicates if the discharge record included emergency department (ED) services: (0) record does not meet any HCUP ED criteria, (1) ED revenue code was on State Inpatient Database (SID) record, (2) ED charge reported on SID record, (3) ED CPT procedure code on SID record, (4) other indication of ED services	1- 39.76% 3-0.0002%	1-43.6% 3-0.0003%
LOS	Length of stay	6	6
MDC	MDC that was in use on the discharge date. The details of MDC categories are given in Table 9 of Appendix C.	5-20.35% (highest category)	5-16.83% (highest category)

[2] https://www.hcup-us.ahrq.gov/db/nation/nrd/nrddde.jsp

		24-0.01% (lowest category)	24-0.05% (lowest category)
NCHRONIC	Number of chronic conditions	5 (mean)	5 (mean)
NDX	Number of diagnoses coded	11(mean)	12 (mean)
NPR	Number of procedures coded	1 (mean)	1 (mean)
NECODE	Number of external causes of injury codes coded	0 (mean)	0 (mean)
NRD_STRATUM	NRD stratum for post-stratification based on geographic region, urban/rural location, teaching status, size of hospital based on number of beds, and control/ownership.	1-7.70% (highest category) 81-0.01% (lowest category)	1-7.18% (highest category) 23,41,43, and 81-0.02% (lowest category)
ORPROC	Indicates if the discharge record has a major operating room procedure: (1) yes, (0) no	0-77.80% 1- 22.21%	0-81.14% 1-18.86%
PAY1	Indicates the type of insurance: (1) Medicare, (2) Medicaid, (3) private insurance, (4) uninsured (self-pay), (5) uninsured (no charge), (6) other	1- 49.03% (highest category) 5-0.65% (lowest category)	1-56.16% (highest category) 5-0.57% (lowest category)
PL_NCHS	Indicates the patient location using the National Center for Health Statistics (NCHS) urban-rural classification scheme for U.S. counties: (1) "Central" counties of metro areas of >=1 million population, (2) "Fringe" counties of metro areas of >=1 million population, (3) counties in metro areas of 250,000–999,999 population, (4) counties in metro areas of 50,000–249,999 population, (5) micropolitan counties, (6) not metropolitan or micropolitan counties	1-32.59% (highest category) 6-5.88% (lowest category)	1-31.63% (highest category) 6-6.09% (lowest category)
REHABTRANSFER	Indicates if the record had transfer to rehabilitation, evaluation, or other aftercare: (1) yes, (0) no	0- 99.08% 1-0.92%	0-98.95% 1-1.05%
RESIDENT	Indicates if a patient is a resident of the State in which he or she received hospital care: (1) yes, (0) no	0-3.35% 1-96.65%	0-3.31% 1- 96.69%
SAMEDAYEVENT	Indicates the same day event status of a patient: (0) not a combined transfer or other same-day stay record, (1) combined transfer involving two discharges from different hospitals, (2)	0-96.48% (highest category)	0-96.21% (highest category)

	combined same-day stay involving two discharges at different hospitals, (3) combined same-day stay involving two discharges at the same hospital, (4) combined same-day stay involving three or more discharges at same or different hospitals	4-0.32% (lowest category)	4-0.35% (lowest category)
ZIPINC_QRTL	Indicates the median household income quartiles for patient's ZIP code. For 2013, the median income quartiles are defined as: (1) $1 - $37,999; (2) $38,000 - $47,999; (3) $48,000 - $63,999; and (4) $64,000 or more.	1-30.70% (highest category) 4-19.21% (lowest category)	1-30.08% (highest category) 4-19.61% (lowest category)
TOTCHG	Costs billed by the hospital for the readmission	$52715 (mean) 10.29 (mean log value)	$53041 (mean) 10.32 (mean log value)
PREVCHG1 (Variable defined by me for this essay)	Costs billed by the hospital for the last previous admission	$48004 (mean) 10.19 (mean log value)	$54925 (mean) 10.34 (mean log value)
AVG_PREVCHG (Variable defined by me for this essay)	Average of costs billed by hospitals for all previous admissions	$48273 (mean) 10.23 (mean log value)	$55774 (mean) 10.39 (mean log value)
SUM_PREVCHG (Variable defined by me for this essay)	Sum of costs billed by hospitals for all previous admissions	$64102 (mean) 10.41 (mean log value)	$84215 (mean) 10.66 (mean log value)
HOSP_BEDSIZE	Indicates the size of a hospital based on the number of beds: (1) small, (2) medium, (3) large.	3-67.98% (highest category) 1-8.98% (lowest category)	3-67.23% (highest category) 1-9.36% (lowest category)
HOSP_UR_TEACH	Indicates the teaching status of the hospital: (0) metropolitan non-teaching, (1) metropolitan teaching, (2) non-metropolitan	1-53.20% (highest category) 2-7.43% (lowest category)	1-51.05% (highest category) 2-8.05% (lowest category)
HOSP_URCAT4	Indicates the urban-rural location of the hospital: (1) large metropolitan areas with at least 1 million residents, (2) small metropolitan areas with less than 1 million residents, (3) micropolitan areas, (4) not metropolitan or micropolitan	1-59.01% (highest category) 4-1.71% (lowest category)	1-58.00% (highest category) 4-1.91% (lowest category)

H_CONTRL	Indicates the control/ownership of the hospital: (1) government, nonfederal [public]; (2) private, not-for-profit [voluntary]; (3) private, investor-owned [proprietary]	2-68.51% (highest category) 1-14.59% (lowest category)	2-68.60% (highest category) 1-13.99% (lowest category)
TOTAL_DISC	Total number of discharges for this hospital in the NRD	19806 (mean)	20311 (mean)
APRDRG_Risk_Mortality	Indicates the 3M All Patient Refined DRG: Risk of mortality subclass: (0) No class specified, (1) minor likelihood of dying, (2) moderate likelihood of dying, (3) major likelihood of dying, (4) extreme likelihood of dying	1-42.83% (highest category) 0-0.03% (lowest category)	1-34.42% (highest category) 0-0.04% (lowest category)
APRDRG_Severity	Indicates the 3M All Patient Refined DRG: Severity of illness subclass: (0) No class specified, (1) minor loss of function (includes cases with no comorbidity or complications), (2) moderate loss of function, (3) major loss of function, (4) extreme loss of function	2-40.83% (highest category) 0-0.03% (lowest category)	2-37.48% (highest category) 0-0.04% (lowest category)
CM_AIDS	Indicates if the AHRQ comorbidity measure - acquired immune deficiency syndrome is present: (1) yes, (0) no	0-99.75% 1-0.25%	0-99.73% 1-0.27%
CM_ALCOHOL	Indicates if the AHRQ comorbidity measure - alcohol abuse is present: (1) yes, (0) no	0-94.75% 1-5.25%	0-94.97% 1-5.03%
CM_ANEMDEF	Indicates if the AHRQ comorbidity measure - deficiency anemias are present: (1) yes, (0) no	0-77.94% 1-22.06%	0-74.98% 1-25.02%
CM_ARTH	Indicates if the AHRQ comorbidity measure - rheumatoid arthritis/collagen vascular diseases are present: (1) yes, (0) no	0-97.29% 1-2.71%	0-96.85% 1-3.15%
CM_BLDLOSS	Indicates if the AHRQ comorbidity measure - chronic blood loss anemia is present: (1) yes, (0) no	0-97.44% 1-2.56%	0-97.91% 1-2.09%
CM_CHF	Indicates if the AHRQ comorbidity measure - congestive heart failure is present: (1) yes, (0) no	0-90.40% 1-9.60%	0-85.99% 1-14.01%
CM_CHRNLUNG	Indicates if the AHRQ comorbidity measure - chronic pulmonary disease is present: (1) yes, (0) no	0-78.49% 1-21.51%	0-76.87% 1-23.13%
CM_COAG	Indicates if the AHRQ comorbidity measure - coagulopathy is present: (1) yes, (0) no	0-94.11% 1-5.89%	0-93.37% 1-6.63%

CM_DEPRESS	Indicates if the AHRQ comorbidity measure - depression is present: (1) yes, (0) no	0-88.79% 1-11.21%	0-88.01% 1-11.99%
CM_DM	Indicates if the AHRQ comorbidity measure - diabetes (uncomplicated) is present: (1) yes, (0) no	0-79.36% 1-20.64%	0-77.65% 1-22.35%
CM_DMCX	Indicates if the AHRQ comorbidity measure - diabetes (with chronic complications) is present: (1) yes, (0) no	0-94.86% 1-5.14%	0-93.81% 1-6.19%
CM_DRUG	Indicates if the AHRQ comorbidity measure - drug abuse is present: (1) yes, (0) no	0-93.73% 1-6.27%	0-94.75% 1-5.25%
CM_HTN_C	Indicates if the AHRQ comorbidity measure - hypertension (uncomplicated and complicated) is present: (1) yes, (0) no	0-50.48% 1-49.52%	0-45.63% 1-54.37%
CM_HYPOTHY	Indicates if the AHRQ comorbidity measure - hypothyroidism is present: (1) yes, (0) no	0-88.77% 1-11.23%	0-87.35% 1-12.65%
CM_LIVER	Indicates if the AHRQ comorbidity measure - liver disease is present: (1) yes, (0) no	0-96.49% 1-3.51%	0-95.78% 1-4.22%
CM_LYMPH	Indicates if the AHRQ comorbidity measure – lymphoma is present: (1) yes, (0) no	0-99.07% 1-0.93%	0-98.60% 1-1.40%
CM_LYTES	Indicates if the AHRQ comorbidity measure - fluid and electrolyte disorders are present: (1) yes, (0) no	0-72.95% 1-27.05%	0-68.07% 1-31.93%
CM_METS	Indicates if the AHRQ comorbidity measure - metastatic cancer is present: (1) yes, (0) no	0-96.89% 1-3.11%	0-95.77% 1-4.23%
CM_NEURO	Indicates if the AHRQ comorbidity measure - other neurological disorders are present: (1) yes, (0) no	0-92.01% 1-7.99%	0-90.05% 1-9.95%
CM_OBESE	Indicates if the AHRQ comorbidity measure – obesity is present: (1) yes, (0) no	0-88.39% 1-11.61%	0-88.39% 1-11.61%
CM_PARA	Indicates if the AHRQ comorbidity measure - paralysis is present: (1) yes, (0) no	0-97.20% 1-2.80%	0-96.44% 1-3.56%
CM_PERIVASC	Indicates if the AHRQ comorbidity measure - peripheral vascular disorders are present: (1) yes, (0) no	0- 93.51% 1-6.49%	0- 92.53% 1-7.47%
CM_PSYCH	Indicates if the AHRQ comorbidity measure - psychosis is present: (1) yes, (0) no	0- 94.68% 1-5.32%	0- 94.17% 1-5.83%
CM_PULMCIRC	Indicates if the AHRQ comorbidity measure - pulmonary circulation disorders are present: (1) yes, (0) no	0- 97.40% 1-2.60%	0- 96.35% 1-3.65%

CM_RENLFAIL	Indicates if the AHRQ comorbidity measure - renal failure is present: (1) yes, (0) no	0- 84.06% 1-15.94%	0- 81.39% 1-18.61%
CM_TUMOR	Indicates if the AHRQ comorbidity measure - solid tumor without metastasis is present: (1) yes, (0) no	0- 97.21% 1-2.79%	0- 96.38% 1-3.62%
CM_ULCER	Indicates if the AHRQ comorbidity measure - peptic ulcer disease excluding bleeding is present: (1) yes, (0) no	0- 99.95% 1-0.05%	0- 99.95% 1-0.05%
CM_VALVE	Indicates if the AHRQ comorbidity measure - valvular disease is present: (1) yes, (0) no	0- 96.68% 1-3.32%	0- 95.17% 1-4.83%
CM_WGHTLOSS	Indicates if the AHRQ comorbidity measure - weight loss is present: (1) yes, (0) no	0-92.87% 1-7.13%	0- 91.40% 1-8.60%

APPENDIX C:

Major Diagnostic Categories and Their Descriptions

(Retrieved From the CMS Website)[3]

Table 9 MDC and its Description

MDC	Description
0	Pre-MDC
1	Diseases and Disorders of the Nervous System
2	Diseases and Disorders of the Eye
3	Diseases and Disorders of the Ear, Nose, Mouth and Throat
4	Diseases and Disorders of the Respiratory System
5	Diseases and Disorders of the Circulatory System
6	Diseases and Disorders of the Digestive System
7	Diseases and Disorders of the Hepatobiliary System and Pancreas
8	Diseases and Disorders of the Musculoskeletal System and Connective Tissue
9	Diseases and Disorders of the Skin, Subcutaneous Tissue and Breast
10	Diseases and Disorders of the Endocrine, Nutritional and Metabolic System
11	Diseases and Disorders of the Kidney and Urinary Tract
12	Diseases and Disorders of the Male Reproductive System
13	Diseases and Disorders of the Female Reproductive System
14	Pregnancy, Childbirth and Puerperium
15	Newborn and Other Neonates (Perinatal Period)
16	Diseases and Disorders of the Blood and Blood Forming Organs and Immunological Disorders
17	Myeloproliferative DDs (Poorly Differentiated Neoplasms)
18	Infectious and Parasitic DDs (Systemic or unspecified sites)
19	Mental Diseases and Disorders
20	Alcohol/Drug Use or Induced Mental Disorders
21	Injuries, Poison and Toxic Effect of Drugs
22	Burns
23	Factors Influencing Health Status and Other Contacts with Health Services
24	Multiple Significant Trauma
25	Human Immunodeficiency Virus Infection

[3] https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Acute-Inpatient-Files-for-Download-Items/CMS1247844

APPENDIX D:

Descriptive Statistics of Each Fold of 10-fold Cross Validation for Readmissions Belonging to Same Major Diagnostic Group Related Analysis

Table 10 Descriptive Statistics of Each Fold of RSDC data

Table 10(a) Training Set RSDC

Fold (Train)	Mean (Readmission Cost)	Max (Readmission Cost)	Min (Readmission Cost)
1	10.2726	15.4250	4.7622
2	10.2854	15.4250	4.7622
3	10.2864	15.3516	4.7622
4	10.2855	15.4250	4.7622
5	10.2859	15.4250	4.7622
6	10.2865	15.4250	4.7622
7	10.2862	15.4250	4.8040
8	10.2860	15.4250	4.7622
9	10.2859	15.4250	4.7622
10	10.2865	15.4250	4.7622

Table 10(b) Testing Set RSDC

Fold (Test)	Mean (Readmission Cost)	Max (Readmission Cost)	Min (Readmission Cost)
1	10.3933	15.2981	5.4294
2	10.2787	15.2832	5.5215
3	10.2695	15.4250	5.1240
4	10.2771	15.0776	4.8040
5	10.2739	15.1679	4.8904
6	10.2687	15.2613	5.0563
7	10.2712	15.1452	4.7622
8	10.2725	15.0915	4.8122
9	10.2733	15.1208	4.8978
10	10.2688	15.3516	4.9273

APPENDIX E:

Descriptive Statistics of Each Fold of 10-fold Cross Validation for Readmissions Belonging to Different Major Diagnostic Group Related Analysis

Table 11 Descriptive Statistics of Each Fold of RADC data

Table 11(a) Training Set RADC

Fold (Train)	Mean (Readmission Cost)	Max (Readmission Cost)	Min (Readmission Cost)
1	10.3035	15.4250	4.6821
2	10.3171	15.4250	4.6821
3	10.3172	15.4250	4.7005
4	10.3171	15.4250	4.6821
5	10.3172	15.4250	4.6821
6	10.3179	15.4250	4.6821
7	10.3170	15.4250	4.6821
8	10.3178	15.4250	4.6821
9	10.3175	15.3622	4.6821
10	10.3173	15.4250	4.6821

Table 11(b) Testing Set RADC

Fold (Test)	Mean (Readmission Cost)	Max (Readmission Cost)	Min (Readmission Cost)
1	10.4279	15.2529	5.2204
2	10.3059	14.8843	4.7875
3	10.3050	15.1184	4.6821
4	10.3060	15.2515	4.7875
5	10.3045	15.0095	5.0563
6	10.2987	15.2981	5.0626
7	10.3064	15.1792	5.2983
8	10.2995	15.3622	5.1591
9	10.3020	15.4250	4.8978
10	10.3038	15.1452	4.7005

APPENDIX F:

Results Based on MAPE of Each Fold of 10-fold Cross Validation for Readmissions Belonging to Same Major Diagnostic Group Related Analysis

Table 12 MAPE for Models Based on RSDC

Table 12(a) MAPE (%) of RSDC for Regression and Generalization Based Models

Model	**Linear Regression**		**Lasso Regression**		**Elastic Net Regression**		**Ridge Regression**	
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	4.2736	4.1824	4.2749	4.1824	4.2750	4.1825	4.3049	4.2066
2	4.2651	4.2860	4.2662	4.2873	4.2664	4.2876	4.2966	4.3179
3	4.2645	4.2779	4.2656	4.2786	4.2657	4.2788	4.2960	4.3105
4	4.2654	4.2525	4.2665	4.2542	4.2668	4.2545	4.2970	4.2848
5	4.2646	4.2983	4.2656	4.3003	4.2657	4.3004	4.2961	4.3323
6	4.2637	4.2818	4.2648	4.2836	4.2650	4.2836	4.2952	4.3146
7	4.2683	4.2668	4.2696	4.2674	4.2691	4.2672	4.2999	4.2994
8	4.2630	4.2898	4.2643	4.2909	4.2638	4.2905	4.2947	4.3203
9	4.2679	4.2442	4.2692	4.2445	4.2694	4.2446	4.2995	4.2751
10	4.2631	4.2998	4.2641	4.3014	4.2645	4.3017	4.2947	4.3320
MEAN	**4.2659**	**4.2679**	**4.2671**	**4.2691**	**4.2672**	**4.2692**	**4.2975**	**4.2994**

Table 12(b) MAPE (%) of RSDC for Hyper Tuning From Lasso towards Elastic Net Based Models

Model	**Hyper tuning From Lasso Towards Elastic Net Regression**							
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	4.2749	4.1824	4.2749	4.1824	4.2750	4.1824	4.2749	4.1825
2	4.2663	4.2875	4.2663	4.2875	4.2662	4.2873	4.2663	4.2875
3	4.2655	4.2785	4.2657	4.2787	4.2656	4.2786	4.2657	4.2787
4	4.2665	4.2542	4.2665	4.2541	4.2666	4.2542	4.2665	4.2542
5	4.2657	4.3004	4.2657	4.3003	4.2657	4.3005	4.2658	4.3005
6	4.2644	4.2831	4.2650	4.2835	4.2650	4.2834	4.2648	4.2835
7	4.2696	4.2674	4.2695	4.2674	4.2692	4.2672	4.2696	4.2675
8	4.2643	4.2909	4.2638	4.2904	4.2643	4.2909	4.2644	4.2910
9	4.2691	4.2445	4.2691	4.2445	4.2692	4.2445	4.2692	4.2445
10	4.2642	4.3015	4.2642	4.3015	4.2642	4.3014	4.2643	4.3016
MEAN	**4.2670**	**4.2690**	**4.2671**	**4.2690**	**4.2671**	**4.2691**	**4.2672**	**4.2691**

Table 12(c) MAPE (%) of RSDC for Hyper Tuning From Elastic Net Towards Ridge Based Models

Model	Hyper tuning From Elastic Net Towards Ridge Regression							
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	4.2750	4.1825	4.2752	4.1826	4.2754	4.1828	4.2760	4.1833
2	4.2665	4.2877	4.2666	4.2878	4.2669	4.2881	4.2675	4.2886
3	4.2658	4.2788	4.2661	4.2791	4.2662	4.2792	4.2669	4.2800
4	4.2668	4.2545	4.2669	4.2545	4.2672	4.2549	4.2678	4.2555
5	4.2660	4.3008	4.2659	4.3006	4.2663	4.3011	4.2670	4.3019
6	4.2650	4.2829	4.2645	4.2839	4.2653	4.2842	4.2662	4.2846
7	4.2691	4.2672	4.2701	4.2679	4.2702	4.2681	4.2698	4.2679
8	4.2644	4.2910	4.2647	4.2912	4.2649	4.2915	4.2654	4.2919
9	4.2693	4.2447	4.2695	4.2448	4.2698	4.2451	4.2695	4.2451
10	4.2644	4.3016	4.2645	4.3017	4.2649	4.3021	4.2655	4.3027
MEAN	**4.2672**	**4.2692**	**4.2674**	**4.2694**	**4.2677**	**4.2697**	**4.2682**	**4.2701**

Table 12(d) MAPE (%) of RSDC for Tree and Deep Learning-based Models

Model	XGBoost		Deep Learning	
Fold	**Training**	**Testing**	**Training**	**Testing**
1	14.6281	14.5031	***3.3363***	***3.2256***
2	14.6132	14.6339	***3.2576***	***3.2883***
3	14.6102	14.6096	***3.2400***	***3.2887***
4	14.6114	14.6263	***3.2756***	***3.3119***
5	14.6124	14.6378	***3.2837***	***3.3042***
6	14.6101	14.6243	***3.2867***	***3.3115***
7	14.6124	14.6409	***3.3015***	***3.3277***
8	14.6112	14.6317	***3.3211***	***3.3460***
9	14.6116	14.5992	***3.2612***	***3.2619***
10	14.6097	14.6277	***3.2665***	***3.3162***
MEAN	**14.6130**	**14.6135**	***3.2830***	***3.2982***

APPENDIX G:
Results based on MAPE of Each Fold of 10-fold Cross Validation for Readmissions Belonging to Different Major Diagnostic Group Related Analysis

Table 13 MAPE for Models Based on RADC

Table 13(a) MAPE (%) of RADC for Regression and Generalization Based Models

Model	Linear Regression		Lasso Regression		Elastic Net Regression		Ridge Regression	
Fold	Training	Testing	Training	Testing	Training	Testing	Training	Testing
1	4.2201	4.0789	4.2210	4.0795	4.2211	4.0796	4.2522	4.1052
2	4.2059	4.2096	4.2069	4.2106	4.2071	4.2108	4.2383	4.2420
3	4.2068	4.2272	4.2077	4.2284	4.2078	4.2285	4.2391	4.2617
4	4.2048	4.2082	4.2058	4.2091	4.2059	4.2092	4.2373	4.2412
5	4.2044	4.2137	4.2054	4.2148	4.2054	4.2148	4.2369	4.2462
6	4.2021	4.2345	4.2032	4.2344	4.2033	4.2346	4.2346	4.2667
7	4.2064	4.2080	4.2073	4.2093	4.2075	4.2095	4.2387	4.2406
8	4.2039	4.2446	4.2048	4.2457	4.2049	4.2459	4.2362	4.2781
9	4.2050	4.2234	4.2060	4.2241	4.2061	4.2242	4.2375	4.2565
10	4.2033	4.2276	4.2043	4.2285	4.2043	4.2286	4.2358	4.2599
MEAN	**4.2063**	**4.2076**	**4.2072**	**4.2084**	**4.2073**	**4.2085**	**4.2387**	**4.2398**

Table 13(b) MAPE (%) of RADC for Hyper Tuning From Lasso Towards Elastic Net Based Models

Model	Hyper tuning From Lasso Towards Elastic Net Regression							
Fold	Training	Testing	Training	Testing	Training	Testing	Training	Testing
1	4.2211	4.0795	4.2211	4.0795	4.2211	4.0795	4.2211	4.0795
2	4.2068	4.2105	4.2070	4.2107	4.2068	4.2105	4.2070	4.2107
3	4.2077	4.2284	4.2077	4.2284	4.2078	4.2285	4.2077	4.2284
4	4.2058	4.2091	4.2058	4.2091	4.2058	4.2090	4.2058	4.2091
5	4.2054	4.2148	4.2054	4.2147	4.2054	4.2148	4.2055	4.2149
6	4.2032	4.2345	4.2032	4.2345	4.2033	4.2345	4.2033	4.2345
7	4.2073	4.2093	4.2072	4.2092	4.2074	4.2094	4.2073	4.2093
8	4.2049	4.2459	4.2047	4.2457	4.2048	4.2458	4.2049	4.2459
9	4.2061	4.2241	4.2061	4.2241	4.2061	4.2241	4.2061	4.2241
10	4.2043	4.2285	4.2043	4.2285	4.2043	4.2286	4.2044	4.2286
MEAN	**4.2073**	**4.2084**	**4.2072**	**4.2084**	**4.2073**	**4.2085**	**4.2073**	**4.2085**

Table 13(c) MAPE (%) of RADC for Hyper Tuning From Elastic Net Towards Ridge Based Models

Model	Hyper tuning From Elastic Towards Ridge Regression							
Fold	Training	Testing	Training	Testing	Training	Testing	Training	Testing
1	4.2212	4.0796	4.2213	4.0797	4.2214	4.0798	4.2218	4.0802
2	4.2070	4.2107	4.2071	4.2108	4.2072	4.2109	4.2077	4.2114
3	4.2078	4.2285	4.2079	4.2286	4.2081	4.2288	4.2085	4.2292
4	4.2059	4.2092	4.2060	4.2093	4.2062	4.2095	4.2066	4.2099
5	4.2056	4.2149	4.2056	4.2150	4.2058	4.2151	4.2061	4.2155
6	4.2034	4.2346	4.2035	4.2347	4.2036	4.2348	4.2039	4.2352
7	4.2074	4.2094	4.2075	4.2095	4.2077	4.2097	4.2081	4.2100
8	4.2049	4.2459	4.2050	4.2460	4.2052	4.2462	4.2056	4.2466
9	4.2061	4.2242	4.2062	4.2243	4.2064	4.2245	4.2068	4.2249
10	4.2044	4.2287	4.2045	4.2287	4.2046	4.2289	4.2050	4.2292
MEAN	**4.2074**	**4.2086**	**4.2075**	**4.2087**	**4.2076**	**4.2088**	**4.2080**	**4.2092**

Table 13(d) MAPE (%) of RADC for Tree and Deep Learning-based Models

Model	XGBoost		Deep Learning	
Fold	Training	Testing	Training	Testing
1	14.5779	14.4319	***3.1916***	***3.2075***
2	14.5611	14.5827	***3.2038***	***3.2246***
3	14.5610	14.5827	***3.1790***	***3.2031***
4	14.5598	14.5742	***3.2179***	***3.2196***
5	14.5586	14.5708	***3.1793***	***3.2189***
6	14.5598	14.5713	***3.1850***	***3.2211***
7	14.5614	14.5723	***3.1809***	***3.1983***
8	14.5591	14.5739	***3.1746***	***3.2014***
9	14.5590	14.5855	***3.1837***	***3.2194***
10	14.5601	14.5726	***3.2069***	***3.2612***
MEAN	**14.5618**	**14.5618**	***3.1903***	***3.2175***

APPENDIX H:

Results Based on RMSE of Each Fold of 10-fold Cross Validation for Readmissions Belonging to Same Major Diagnostic Group Related Analysis

Table 14 RMSE for Models Based on RSDC

Table 14(a) RMSE of RSDC for Regression and Generalization Based Models

Model	**Linear Regression**		**Lasso Regression**		**Elastic Net Regression**		**Ridge Regression**	
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.5643	0.5598	0.5644	0.5599	0.5644	0.5599	0.5653	0.5595
2	0.5639	0.5635	0.5640	0.5636	0.5640	0.5636	0.5649	0.5649
3	0.5638	0.5646	0.5639	0.5647	0.5639	0.5647	0.5648	0.5656
4	0.5635	0.5668	0.5636	0.5670	0.5636	0.5670	0.5645	0.5675
5	0.5639	0.5634	0.5640	0.5636	0.5640	0.5636	0.5649	0.5652
6	0.5635	0.5671	0.5636	0.5672	0.5636	0.5672	0.5645	0.5679
7	0.5641	0.5618	0.5642	0.5619	0.5641	0.5618	0.5651	0.5635
8	0.5635	0.5666	0.5637	0.5667	0.5636	0.5666	0.5646	0.5674
9	0.5641	0.5616	0.5642	0.5616	0.5642	0.5616	0.5651	0.5622
10	0.5636	0.5660	0.5637	0.5661	0.5637	0.5662	0.5646	0.5673
MEAN	**0.5638**	**0.5641**	**0.5639**	**0.5642**	**0.5639**	**0.5642**	**0.5648**	**0.5651**

Table 14(b) RMSE of RSDC for Hyper Tuning From Lasso Towards Elastic Net Based Models

Model	**From Lasso Towards Elastic Net Regression**							
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.5644	0.5599	0.5644	0.5599	0.5644	0.5599	0.5644	0.5599
2	0.5640	0.5636	0.5640	0.5636	0.5640	0.5636	0.5640	0.5636
3	0.5639	0.5647	0.5639	0.5647	0.5639	0.5647	0.5639	0.5647
4	0.5636	0.5670	0.5636	0.5670	0.5636	0.5670	0.5636	0.5670
5	0.5640	0.5636	0.5640	0.5636	0.5640	0.5636	0.5640	0.5636
6	0.5636	0.5672	0.5636	0.5672	0.5636	0.5672	0.5636	0.5672
7	0.5642	0.5619	0.5642	0.5619	0.5642	0.5618	0.5642	0.5619
8	0.5637	0.5667	0.5636	0.5666	0.5637	0.5667	0.5637	0.5667
9	0.5642	0.5616	0.5642	0.5616	0.5642	0.5616	0.5642	0.5616
10	0.5637	0.5662	0.5637	0.5662	0.5637	0.5661	0.5637	0.5662
MEAN	**0.5639**	**0.5642**	**0.5639**	**0.5642**	**0.5639**	**0.5642**	**0.5639**	**0.5642**

Table 14(c) RMSE of RSDC for Hyper Tuning from Elastic Net Towards Ridge Based Models

Model	From Elastic Net Towards Ridge Regression							
Fold	Training	Testing	Training	Testing	Training	Testing	Training	Testing
1	0.5644	0.5599	0.5644	0.5599	0.5644	0.5599	0.5644	0.5598
2	0.5640	0.5636	0.5640	0.5636	0.5640	0.5636	0.5640	0.5636
3	0.5639	0.5647	0.5639	0.5647	0.5639	0.5647	0.5639	0.5647
4	0.5636	0.5670	0.5636	0.5670	0.5636	0.5670	0.5636	0.5670
5	0.5640	0.5637	0.5640	0.5636	0.5640	0.5637	0.5640	0.5637
6	0.5636	0.5671	0.5636	0.5672	0.5636	0.5672	0.5636	0.5672
7	0.5641	0.5618	0.5642	0.5619	0.5642	0.5619	0.5642	0.5618
8	0.5637	0.5667	0.5637	0.5667	0.5637	0.5667	0.5637	0.5667
9	0.5642	0.5616	0.5642	0.5616	0.5642	0.5616	0.5642	0.5616
10	0.5637	0.5661	0.5637	0.5661	0.5637	0.5662	0.5637	0.5662
MEAN	**0.5639**	**0.5642**	**0.5639**	**0.5642**	**0.5639**	**0.5642**	**0.5639**	**0.5642**

Table 14(d) RMSE of RSDC for Tree and Deep Learning-based Models

Model	XGBoost		Deep Learning	
Fold	Training	Testing	Training	Testing
1	1.5293	1.5424	***0.4373***	***0.4305***
2	1.5294	1.5421	***0.4309***	***0.4346***
3	1.5292	1.5388	***0.4292***	***0.4334***
4	1.5291	1.5419	***0.4332***	***0.4374***
5	1.5293	1.5442	***0.4357***	***0.4377***
6	1.5292	1.5392	***0.4371***	***0.4403***
7	1.5293	1.5421	***0.4393***	***0.4425***
8	1.5293	1.5429	***0.4425***	***0.4457***
9	1.5291	1.5386	***0.4333***	***0.4330***
10	1.5291	1.5409	***0.4337***	***0.4387***
MEAN	**1.5292**	**1.5413**	***0.4352***	***0.4374***

APPENDIX I:
Results Based on RMSE of Each Fold of 10-fold Cross Validation for Readmissions Belonging to Different Major Diagnostic Group Related Analysis

Table 15 RMSE for Models Based on RADC

Table 15(a) RMSE of RADC for Regression and Generalization Based Models

Model	**Linear Regression**		**Lasso Regression**		**Elastic Net Regression**		**Ridge Regression**	
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.5592	0.5474	0.5593	0.5474	0.5593	0.5474	0.5602	0.5474
2	0.5580	0.5584	0.5580	0.5585	0.5580	0.5585	0.5590	0.5594
3	0.5582	0.5562	0.5583	0.5563	0.5583	0.5563	0.5592	0.5581
4	0.5578	0.5601	0.5578	0.5601	0.5578	0.5601	0.5588	0.5608
5	0.5577	0.5603	0.5578	0.5604	0.5578	0.5604	0.5587	0.5611
6	0.5575	0.5624	0.5576	0.5623	0.5576	0.5623	0.5585	0.5631
7	0.5580	0.5581	0.5581	0.5582	0.5581	0.5582	0.5590	0.5592
8	0.5579	0.5593	0.5579	0.5594	0.5579	0.5594	0.5589	0.5607
9	0.5579	0.5587	0.5580	0.5588	0.5580	0.5588	0.5589	0.5599
10	0.5577	0.5607	0.5578	0.5608	0.5578	0.5607	0.5587	0.5615
MEAN	**0.5580**	**0.5582**	**0.5581**	**0.5582**	**0.5581**	**0.5582**	**0.5590**	**0.5591**

Table 15(b) RMSE of RADC for Hyper Tuning From Lasso Towards Elastic Net Based Models

Model	**From Lasso Towards Elastic Net Regression**							
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.5593	0.5474	0.5593	0.5474	0.5593	0.5474	0.5593	0.5474
2	0.5580	0.5585	0.5580	0.5585	0.5580	0.5585	0.5580	0.5585
3	0.5583	0.5563	0.5583	0.5563	0.5583	0.5563	0.5583	0.5563
4	0.5578	0.5601	0.5578	0.5601	0.5578	0.5601	0.5578	0.5601
5	0.5578	0.5604	0.5578	0.5604	0.5578	0.5604	0.5578	0.5604
6	0.5576	0.5623	0.5576	0.5623	0.5576	0.5623	0.5576	0.5623
7	0.5581	0.5582	0.5581	0.5582	0.5581	0.5582	0.5581	0.5582
8	0.5579	0.5594	0.5579	0.5594	0.5579	0.5594	0.5579	0.5594
9	0.5580	0.5588	0.5580	0.5588	0.5580	0.5588	0.5580	0.5588
10	0.5578	0.5608	0.5578	0.5608	0.5578	0.5608	0.5578	0.5608
MEAN	**0.5581**	**0.5582**	**0.5581**	**0.5582**	**0.5581**	**0.5582**	**0.5581**	**0.5582**

Table 15(c) RMSE of RADC for Hyper Tuning From Elastic Net Towards Ridge Based Models

Model	**From Elastic Net Towards Ridge Regression**							
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.5593	0.5474	0.5593	0.5474	0.5593	0.5474	0.5593	0.5474
2	0.5580	0.5585	0.5580	0.5585	0.5580	0.5585	0.5580	0.5585
3	0.5583	0.5563	0.5583	0.5563	0.5583	0.5563	0.5583	0.5563
4	0.5578	0.5601	0.5578	0.5601	0.5578	0.5601	0.5578	0.5601
5	0.5578	0.5604	0.5578	0.5604	0.5578	0.5604	0.5578	0.5604
6	0.5576	0.5623	0.5576	0.5623	0.5576	0.5623	0.5576	0.5623
7	0.5581	0.5582	0.5581	0.5582	0.5581	0.5582	0.5581	0.5582
8	0.5579	0.5594	0.5579	0.5594	0.5579	0.5594	0.5579	0.5594
9	0.5580	0.5588	0.5580	0.5588	0.5580	0.5588	0.5580	0.5588
10	0.5578	0.5608	0.5578	0.5607	0.5578	0.5607	0.5578	0.5607
MEAN	**0.5581**	**0.5582**	**0.5581**	**0.5582**	**0.5581**	**0.5582**	**0.5581**	**0.5582**

Table 15(d) RMSE of RADC for Tree and Deep Learning Based Models

Model	**XGBoost**		**Deep Learning**	
Fold	**Training**	**Testing**	**Training**	**Testing**
1	1.5312	1.5370	***0.4234***	***0.4308***
2	1.5309	1.5394	***0.4270***	***0.4298***
3	1.5310	1.5395	***0.4214***	***0.4246***
4	1.5309	1.5381	***0.4289***	***0.4287***
5	1.5311	1.5382	***0.4215***	***0.4255***
6	1.5310	1.5378	***0.4232***	***0.4266***
7	1.5312	1.5439	***0.4220***	***0.4241***
8	1.5311	1.5374	***0.4200***	***0.4224***
9	1.5310	1.5380	***0.4209***	***0.4251***
10	1.5310	1.5374	***0.4219***	***0.4275***
MEAN	**1.5310**	**1.5387**	***0.4230***	***0.4265***

APPENDIX J:

Results Based on MAE of Each Fold of 10-fold Cross Validation for Readmissions Belonging to Same Major Diagnostic Group Related Analysis

Table 16 MAE for Models Based on RSDC

Table 16(a) MAE of RSDC for Regression and Generalization Based Models

Model	**Linear Regression**		**Lasso Regression**		**Elastic Net Regression**		**Ridge Regression**	
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.4313	0.4279	0.4314	0.4279	0.4314	0.4279	0.4340	0.4298
2	0.4310	0.4327	0.4311	0.4328	0.4311	0.4328	0.4337	0.4354
3	0.4310	0.4316	0.4311	0.4316	0.4311	0.4316	0.4337	0.4343
4	0.4310	0.4297	0.4311	0.4298	0.4312	0.4299	0.4337	0.4324
5	0.4310	0.4334	0.4311	0.4335	0.4311	0.4336	0.4337	0.4363
6	0.4309	0.4320	0.4310	0.4322	0.4310	0.4322	0.4336	0.4348
7	0.4314	0.4305	0.4315	0.4305	0.4314	0.4305	0.4341	0.4333
8	0.4308	0.4332	0.4309	0.4333	0.4309	0.4333	0.4335	0.4358
9	0.4313	0.4283	0.4314	0.4283	0.4314	0.4283	0.4340	0.4309
10	0.4309	0.4335	0.4310	0.4337	0.4310	0.4337	0.4336	0.4363
MEAN	**0.4311**	**0.4313**	**0.4312**	**0.4314**	**0.4312**	**0.4314**	**0.4337**	**0.4339**

Table 16(b) MAE of RSDC for Hyper Tuning From Lasso Towards Elastic Net Based Models

Model	**Hyper tuning From Lasso Towards Elastic Net Regression**							
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.4314	0.4279	0.4314	0.4279	0.4314	0.4279	0.4314	0.4279
2	0.4311	0.4328	0.4311	0.4328	0.4311	0.4328	0.4311	0.4328
3	0.4311	0.4316	0.4311	0.4316	0.4311	0.4316	0.4311	0.4316
4	0.4311	0.4298	0.4311	0.4298	0.4311	0.4298	0.4311	0.4298
5	0.4311	0.4336	0.4311	0.4336	0.4311	0.4336	0.4311	0.4336
6	0.4310	0.4321	0.4310	0.4322	0.4310	0.4321	0.4310	0.4321
7	0.4315	0.4305	0.4315	0.4305	0.4314	0.4305	0.4315	0.4305
8	0.4309	0.4333	0.4309	0.4333	0.4309	0.4333	0.4309	0.4333
9	0.4314	0.4283	0.4314	0.4283	0.4314	0.4283	0.4314	0.4283
10	0.4310	0.4337	0.4310	0.4337	0.4310	0.4337	0.4310	0.4337
MEAN	**0.4312**	**0.4314**	**0.4312**	**0.4314**	**0.4312**	**0.4314**	**0.4312**	**0.4314**

Table 16(c) MAE of RSDC for Hyper Tuning From Elastic Net Towards Ridge Based Models

Model	Hyper tuning From Elastic Net Towards Ridge Regression							
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.4314	0.4279	0.4315	0.4279	0.4315	0.4279	0.4315	0.4280
2	0.4311	0.4328	0.4312	0.4328	0.4312	0.4328	0.4312	0.4329
3	0.4311	0.4316	0.4311	0.4316	0.4311	0.4316	0.4312	0.4317
4	0.4312	0.4299	0.4312	0.4299	0.4312	0.4299	0.4312	0.4300
5	0.4311	0.4336	0.4311	0.4336	0.4311	0.4336	0.4312	0.4337
6	0.4310	0.4321	0.4310	0.4322	0.4310	0.4322	0.4311	0.4322
7	0.4314	0.4305	0.4315	0.4305	0.4315	0.4305	0.4315	0.4305
8	0.4309	0.4333	0.4310	0.4333	0.4310	0.4334	0.4310	0.4334
9	0.4314	0.4283	0.4314	0.4284	0.4315	0.4284	0.4314	0.4284
10	0.4310	0.4337	0.4310	0.4337	0.4310	0.4337	0.4311	0.4338
MEAN	**0.4312**	**0.4314**	**0.4312**	**0.4314**	**0.4312**	**0.4314**	**0.4313**	**0.4315**

Table 16(d) MAE of RSDC for Tree and Deep Learning-based Models

Model	XGBoost		Deep Learning	
Fold	**Training**	**Testing**	**Training**	**Testing**
1	1.5004	1.5045	***0.3356***	***0.3309***
2	1.5006	1.5027	***0.3297***	***0.3326***
3	1.5005	1.4996	***0.3285***	***0.3326***
4	1.5004	1.5020	***0.3318***	***0.3347***
5	1.5006	1.5001	***0.3337***	***0.3352***
6	1.5004	1.4994	***0.3346***	***0.3357***
7	1.5005	1.5020	***0.3362***	***0.3385***
8	1.5005	1.5031	***0.3396***	***0.3414***
9	1.5004	1.4993	***0.3321***	***0.3313***
10	1.5004	1.5011	***0.3316***	***0.3356***
MEAN	**1.5005**	**1.5014**	***0.3333***	***0.3349***

APPENDIX K:
Results Based on MAE of Each Fold of 10-fold Cross Validation for Readmissions Belonging to Different Major Diagnostic Group Related Analysis

Table 17 MAE for Models Based on RADC

Table 17(a) MAE of RADC for Regression and Generalization Based Models

Model	**Linear Regression**		**Lasso Regression**		**Elastic Net Regression**		**Ridge Regression**	
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.4273	0.4186	0.4273	0.4186	0.4273	0.4186	0.4300	0.4207
2	0.4264	0.4265	0.4265	0.4266	0.4265	0.4266	0.4292	0.4293
3	0.4265	0.4279	0.4266	0.4280	0.4266	0.4280	0.4293	0.4310
4	0.4263	0.4263	0.4264	0.4263	0.4264	0.4264	0.4291	0.4291
5	0.4262	0.4268	0.4263	0.4269	0.4263	0.4269	0.4290	0.4296
6	0.4261	0.4283	0.4262	0.4283	0.4262	0.4283	0.4289	0.4311
7	0.4264	0.4263	0.4265	0.4264	0.4265	0.4264	0.4292	0.4291
8	0.4262	0.4292	0.4263	0.4292	0.4263	0.4293	0.4290	0.4321
9	0.4263	0.4274	0.4264	0.4274	0.4264	0.4274	0.4292	0.4302
10	0.4262	0.4279	0.4262	0.4280	0.4262	0.4280	0.4290	0.4307
MEAN	**0.4264**	**0.4265**	**0.4265**	**0.4266**	**0.4265**	**0.4266**	**0.4292**	**0.4293**

Table 17(b) MAE of RADC for Hyper Tuning From Lasso towards Elastic Net Based Models

Model	**Hyper tuning From Lasso Towards Elastic Net Regression**							
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.4273	0.4186	0.4273	0.4186	0.4273	0.4186	0.4273	0.4186
2	0.4264	0.4266	0.4265	0.4266	0.4264	0.4266	0.4265	0.4266
3	0.4266	0.4280	0.4265	0.4280	0.4266	0.4280	0.4266	0.4280
4	0.4264	0.4263	0.4264	0.4263	0.4264	0.4263	0.4264	0.4263
5	0.4263	0.4269	0.4263	0.4269	0.4263	0.4269	0.4263	0.4269
6	0.4262	0.4283	0.4262	0.4283	0.4262	0.4283	0.4262	0.4283
7	0.4265	0.4264	0.4265	0.4264	0.4265	0.4264	0.4265	0.4264
8	0.4263	0.4293	0.4263	0.4292	0.4263	0.4292	0.4263	0.4293
9	0.4264	0.4274	0.4264	0.4274	0.4264	0.4274	0.4264	0.4274
10	0.4262	0.4280	0.4262	0.4280	0.4262	0.4280	0.4262	0.4280
MEAN	**0.4265**	**0.4266**	**0.4265**	**0.4266**	**0.4265**	**0.4266**	**0.4265**	**0.4266**

Table 17(c) MAE of RADC for Hyper Tuning From Elastic Net Towards Ridge Based Models

Model	**Hyper tuning From Elastic Net Towards Ridge Regression**							
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.4273	0.4186	0.4273	0.4186	0.4274	0.4187	0.4274	0.4187
2	0.4265	0.4266	0.4265	0.4266	0.4265	0.4266	0.4265	0.4267
3	0.4266	0.4280	0.4266	0.4280	0.4266	0.4281	0.4266	0.4281
4	0.4264	0.4264	0.4264	0.4264	0.4264	0.4264	0.4264	0.4264
5	0.4263	0.4269	0.4263	0.4269	0.4263	0.4269	0.4264	0.4269
6	0.4262	0.4283	0.4262	0.4283	0.4262	0.4283	0.4262	0.4284
7	0.4265	0.4264	0.4265	0.4264	0.4265	0.4265	0.4266	0.4265
8	0.4263	0.4293	0.4263	0.4293	0.4263	0.4293	0.4264	0.4293
9	0.4264	0.4274	0.4264	0.4274	0.4265	0.4274	0.4265	0.4275
10	0.4263	0.4280	0.4263	0.4280	0.4263	0.4280	0.4263	0.4281
MEAN	**0.4265**	**0.4266**	**0.4265**	**0.4266**	**0.4265**	**0.4266**	**0.4265**	**0.4267**

Table 17(d) MAE of RADC for Tree and Deep Learning-based Models

Model	**XGBoost**		**Deep Learning**	
Fold	**Training**	**Testing**	**Training**	**Testing**
1	1.5005	1.5022	***0.3246***	***0.3326***
2	1.5006	1.5394	***0.3278***	***0.3293***
3	1.5006	1.5023	***0.3232***	***0.3254***
4	1.5006	1.5014	***0.3294***	***0.3286***
5	1.5005	1.5007	***0.3235***	***0.3266***
6	1.5005	1.5003	***0.3248***	***0.3271***
7	1.5007	1.5013	***0.3238***	***0.3253***
8	1.5006	1.5000	***0.3222***	***0.3240***
9	1.5005	1.5008	***0.3233***	***0.3260***
10	1.5005	1.5004	***0.3236***	***0.3281***
MEAN	**1.5006**	**1.5049**	***0.3246***	***0.3273***

APPENDIX L:
Results Based on RAE of Each Fold of 10-fold Cross Validation for Readmissions Belonging to Same Major Diagnostic Group Related Analysis

Table 18 RAE for Models Based on RSDC

Table 18(a) RAE of RSDC for Regression and Generalization Based Models

Model	**Linear Regression**		**Lasso Regression**		**Elastic Net Regression**		**Ridge Regression**	
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.5285	0.5162	0.5286	0.5162	0.5286	0.5162	0.5318	0.5185
2	0.5270	0.5292	0.5271	0.5293	0.5271	0.5293	0.5303	0.5325
3	0.5268	0.5294	0.5269	0.5294	0.5269	0.5294	0.5301	0.5327
4	0.5268	0.5272	0.5269	0.5274	0.5270	0.5274	0.5301	0.5306
5	0.5273	0.5272	0.5274	0.5275	0.5274	0.5275	0.5305	0.5308
6	0.5266	0.5308	0.5267	0.5310	0.5267	0.5310	0.5298	0.5342
7	0.5273	0.5279	0.5274	0.5280	0.5273	0.5279	0.5305	0.5314
8	0.5267	0.5303	0.5268	0.5304	0.5268	0.5303	0.5300	0.5334
9	0.5269	0.5277	0.5270	0.5277	0.5271	0.5277	0.5302	0.5309
10	0.5268	0.5304	0.5269	0.5305	0.5269	0.5305	0.5301	0.5337
MEAN	**0.5271**	**0.5276**	**0.5272**	**0.5277**	**0.5272**	**0.5277**	**0.5303**	**0.5309**

Table 18(b) RAE of RSDC for Hyper Tuning From Lasso Towards Elastic Net Based Models

Model	**Hyper Tuning from Lasso Towards Elastic Net Regression**							
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.5286	0.5162	0.5286	0.5162	0.5286	0.5162	0.5286	0.5162
2	0.5271	0.5293	0.5271	0.5293	0.5271	0.5293	0.5271	0.5293
3	0.5269	0.5294	0.5269	0.5294	0.5269	0.5294	0.5269	0.5294
4	0.5269	0.5274	0.5269	0.5274	0.5269	0.5274	0.5269	0.5274
5	0.5274	0.5275	0.5274	0.5275	0.5274	0.5275	0.5274	0.5275
6	0.5266	0.5309	0.5267	0.5310	0.5267	0.5310	0.5267	0.5310
7	0.5274	0.5280	0.5274	0.5280	0.5273	0.5279	0.5274	0.5280
8	0.5268	0.5304	0.5268	0.5303	0.5268	0.5304	0.5268	0.5304
9	0.5270	0.5277	0.5270	0.5277	0.5270	0.5277	0.5270	0.5277
10	0.5269	0.5305	0.5269	0.5305	0.5269	0.5305	0.5269	0.5305
MEAN	**0.5272**	**0.5277**	**0.5272**	**0.5277**	**0.5272**	**0.5277**	**0.5272**	**0.5277**

Table 18(c) RAE of RSDC for Hyper Tuning From Elastic Net Towards Ridge Based Models

Model	**Hyper tuning from Elastic Net towards Ridge Regression**							
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.5286	0.5162	0.5287	0.5162	0.5287	0.5162	0.5287	0.5163
2	0.5271	0.5293	0.5271	0.5294	0.5272	0.5294	0.5272	0.5294
3	0.5269	0.5294	0.5269	0.5295	0.5270	0.5295	0.5270	0.5295
4	0.5270	0.5274	0.5270	0.5274	0.5270	0.5275	0.5271	0.5275
5	0.5274	0.5275	0.5274	0.5275	0.5274	0.5275	0.5275	0.5276
6	0.5267	0.5309	0.5267	0.5310	0.5267	0.5310	0.5268	0.5311
7	0.5273	0.5279	0.5274	0.5280	0.5274	0.5280	0.5274	0.5280
8	0.5268	0.5304	0.5268	0.5304	0.5269	0.5304	0.5269	0.5305
9	0.5271	0.5277	0.5271	0.5277	0.5271	0.5278	0.5271	0.5278
10	0.5269	0.5305	0.5269	0.5305	0.5270	0.5306	0.5270	0.5306
MEAN	**0.5272**	**0.5277**	**0.5272**	**0.5278**	**0.5272**	**0.5278**	**0.5273**	**0.5278**

Table 18(d) RAE of RSDC for Tree and Deep Learning-based Models

Model	**XGBoost**		**Deep Learning**	
Fold	**Training**	**Testing**	**Training**	**Testing**
1	1.8384	1.8149	***0.4112***	***0.3992***
2	1.8346	1.8379	***0.4031***	***0.4068***
3	1.8339	1.8394	***0.4015***	***0.4093***
4	1.8338	1.8429	***0.4055***	***0.4106***
5	1.8357	1.8250	***0.4083***	***0.4079***
6	1.8335	1.8422	***0.4089***	***0.4124***
7	1.8340	1.8420	***0.4109***	***0.4151***
8	1.8344	1.8398	***0.4151***	***0.4179***
9	1.8330	1.8472	***0.4057***	***0.4081***
10	1.8344	1.8363	***0.4055***	***0.4106***
MEAN	**1.8346**	**1.8368**	***0.4076***	***0.4098***

APPENDIX M:

Results Based on RAE of Each Fold of 10-fold Cross Validation for Readmissions Belonging to Different Major Diagnostic Group Related Analysis

Table 19 RAE for Models Based on RADC

Table 19(a) RAE of RADC for Regression and Generalization Based Models

Model	**Linear Regression**		**Lasso Regression**		**Elastic Net Regression**		**Ridge Regression**	
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.5381	0.5241	0.5381	0.5242	0.5382	0.5242	0.5415	0.5268
2	0.5363	0.5367	0.5364	0.5368	0.5364	0.5368	0.5398	0.5402
3	0.5365	0.5380	0.5365	0.5381	0.5366	0.5381	0.5400	0.5418
4	0.5360	0.5381	0.5361	0.5382	0.5361	0.5382	0.5395	0.5417
5	0.5361	0.5371	0.5362	0.5373	0.5362	0.5373	0.5396	0.5407
6	0.5358	0.5399	0.5359	0.5399	0.5359	0.5399	0.5393	0.5435
7	0.5362	0.5378	0.5363	0.5379	0.5363	0.5379	0.5397	0.5413
8	0.5361	0.5400	0.5362	0.5402	0.5362	0.5402	0.5396	0.5437
9	0.5363	0.5375	0.5364	0.5375	0.5364	0.5376	0.5398	0.5411
10	0.5360	0.5389	0.5361	0.5390	0.5361	0.5390	0.5395	0.5424
MEAN	**0.5363**	**0.5368**	**0.5364**	**0.5369**	**0.5364**	**0.5369**	**0.5398**	**0.5403**

Table 19(b) RAE of RADC for Hyper Tuning From Lasso Towards Elastic Net Based Models

Model	**Hyper Tuning From Lasso Towards Elastic Net Regression**							
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.5381	0.5242	0.5382	0.5242	0.5382	0.5242	0.5382	0.5242
2	0.5364	0.5368	0.5364	0.5368	0.5364	0.5368	0.5364	0.5368
3	0.5366	0.5381	0.5365	0.5381	0.5366	0.5381	0.5365	0.5381
4	0.5361	0.5382	0.5361	0.5382	0.5361	0.5382	0.5361	0.5382
5	0.5362	0.5373	0.5362	0.5373	0.5362	0.5373	0.5362	0.5373
6	0.5359	0.5399	0.5359	0.5399	0.5359	0.5399	0.5359	0.5399
7	0.5363	0.5379	0.5363	0.5379	0.5363	0.5379	0.5363	0.5379
8	0.5362	0.5402	0.5362	0.5402	0.5362	0.5402	0.5362	0.5402
9	0.5364	0.5375	0.5364	0.5375	0.5364	0.5375	0.5364	0.5375
10	0.5361	0.5390	0.5361	0.5390	0.5361	0.5390	0.5361	0.5390
MEAN	**0.5364**	**0.5369**	**0.5364**	**0.5369**	**0.5364**	**0.5369**	**0.5364**	**0.5369**

Table 19(c) RAE of RADC for Hyper Tuning From Elastic Net Towards Ridge Based Models

Model	**Hyper Tuning From Elastic Net Towards Ridge Regression**							
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.5382	0.5242	0.5382	0.5242	0.5382	0.5242	0.5382	0.5242
2	0.5364	0.5368	0.5364	0.5368	0.5364	0.5368	0.5365	0.5369
3	0.5366	0.5381	0.5366	0.5381	0.5366	0.5382	0.5366	0.5382
4	0.5361	0.5382	0.5361	0.5382	0.5361	0.5382	0.5362	0.5383
5	0.5362	0.5373	0.5362	0.5373	0.5362	0.5373	0.5363	0.5373
6	0.5359	0.5399	0.5359	0.5400	0.5359	0.5400	0.5360	0.5400
7	0.5363	0.5379	0.5363	0.5379	0.5363	0.5380	0.5364	0.5380
8	0.5362	0.5402	0.5362	0.5402	0.5362	0.5402	0.5363	0.5403
9	0.5364	0.5376	0.5364	0.5376	0.5364	0.5376	0.5365	0.5376
10	0.5361	0.5390	0.5361	0.5390	0.5361	0.5390	0.5361	0.5391
MEAN	**0.5364**	**0.5369**	**0.5364**	**0.5369**	**0.5365**	**0.5370**	**0.5365**	**0.5370**

Table 19(d) RAE of RADC for Tree and Deep Learning-based Models

Model	**XGBoost**		**Deep Learning**	
Fold	**Training**	**Testing**	**Training**	**Testing**
1	1.8896	1.8809	***0.4087***	***0.4165***
2	1.8874	1.8900	***0.4123***	***0.4144***
3	1.8876	1.8888	***0.4065***	***0.4092***
4	1.8867	1.8953	***0.4142***	***0.4148***
5	1.8872	1.8888	***0.4069***	***0.4111***
6	1.8869	1.8914	***0.4084***	***0.4124***
7	1.8870	1.8938	***0.4072***	***0.4103***
8	1.8874	1.8876	***0.4053***	***0.4078***
9	1.8874	1.8876	***0.4066***	***0.4100***
10	1.8872	1.8894	***0.4070***	***0.4131***
MEAN	**1.8874**	**1.8893**	***0.4083***	***0.4119***

APPENDIX N:

Results Based on RRSE of Each Fold of 10-fold Cross Validation for Readmissions Belonging to Same Major Diagnostic Group Related Analysis

Table 20 RRSE for Models Based on RSDC

Table 20(a) RRSE of RSDC for Regression and Generalization Based Models

Model	**Linear Regression**		**Lasso Regression**		**Elastic Net Regression**		**Ridge Regression**	
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.5466	0.5342	0.5467	0.5342	0.5467	0.5342	0.5476	0.5339
2	0.5450	0.5447	0.5452	0.5448	0.5452	0.5448	0.5460	0.5461
3	0.5447	0.5479	0.5448	0.5480	0.5448	0.5480	0.5457	0.5488
4	0.5446	0.5484	0.5447	0.5485	0.5448	0.5486	0.5456	0.5490
5	0.5453	0.5430	0.5453	0.5432	0.5453	0.5432	0.5462	0.5447
6	0.5444	0.5503	0.5445	0.5504	0.5445	0.5504	0.5454	0.5511
7	0.5451	0.5445	0.5452	0.5446	0.5451	0.5446	0.5460	0.5462
8	0.5446	0.5491	0.5447	0.5492	0.5446	0.5491	0.5455	0.5498
9	0.5448	0.5468	0.5449	0.5468	0.5449	0.5468	0.5458	0.5474
10	0.5447	0.5475	0.5448	0.5476	0.5449	0.5476	0.5457	0.5487
MEAN	**0.5450**	**0.5456**	**0.5451**	**0.5457**	**0.5451**	**0.5457**	**0.5460**	**0.5466**

Table 20(b) RRSE of RSDC for Hyper Tuning From Lasso Towards Elastic Net Based Models

Model	**Hyper Tuning From Lasso Towards Elastic Net Regression**							
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.5467	0.5342	0.5467	0.5342	0.5467	0.5342	0.5467	0.5342
2	0.5452	0.5448	0.5452	0.5448	0.5451	0.5448	0.5452	0.5448
3	0.5448	0.5480	0.5448	0.5480	0.5448	0.5480	0.5448	0.5480
4	0.5447	0.5485	0.5447	0.5485	0.5447	0.5485	0.5447	0.5485
5	0.5453	0.5432	0.5453	0.5432	0.5453	0.5432	0.5453	0.5432
6	0.5445	0.5504	0.5446	0.5504	0.5446	0.5504	0.5445	0.5504
7	0.5452	0.5446	0.5452	0.5446	0.5451	0.5446	0.5452	0.5446
8	0.5447	0.5492	0.5446	0.5491	0.5447	0.5491	0.5447	0.5492
9	0.5449	0.5468	0.5449	0.5468	0.5449	0.5468	0.5449	0.5468
10	0.5449	0.5476	0.5449	0.5476	0.5448	0.5476	0.5449	0.5476
MEAN	**0.5451**	**0.5457**	**0.5451**	**0.5457**	**0.5451**	**0.5457**	**0.5451**	**0.5457**

Table 20(c) RRSE of RSDC for Hyper Tuning From Elastic Net Towards Ridge Based Models

Model	**Hyper Tuning From Elastic Net Towards Ridge Regression**							
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.5467	0.5342	0.5467	0.5342	0.5467	0.5342	0.5467	0.5342
2	0.5452	0.5448	0.5452	0.5448	0.5452	0.5448	0.5452	0.5448
3	0.5448	0.5480	0.5448	0.5480	0.5448	0.5480	0.5448	0.5480
4	0.5448	0.5486	0.5447	0.5485	0.5448	0.5486	0.5448	0.5486
5	0.5454	0.5432	0.5453	0.5432	0.5453	0.5432	0.5454	0.5432
6	0.5445	0.5503	0.5445	0.5504	0.5445	0.5504	0.5446	0.5504
7	0.5451	0.5446	0.5452	0.5446	0.5452	0.5446	0.5451	0.5446
8	0.5447	0.5491	0.5447	0.5492	0.5447	0.5492	0.5447	0.5491
9	0.5449	0.5468	0.5449	0.5468	0.5449	0.5468	0.5449	0.5468
10	0.5448	0.5476	0.5448	0.5476	0.5449	0.5476	0.5449	0.5476
MEAN	**0.5451**	**0.5457**	**0.5451**	**0.5457**	**0.5451**	**0.5457**	**0.5451**	**0.5457**

Table 20(d) RRSE of RSDC for Tree and Deep Learning-based Models

Model	**XGBoost**		**Deep Learning**	
Fold	**Training**	**Testing**	**Training**	**Testing**
1	1.4813	1.4718	***0.4236***	***0.4108***
2	1.4782	1.4908	***0.4165***	***0.4202***
3	1.4774	1.4933	***0.4146***	***0.4206***
4	1.4779	1.4918	***0.4186***	***0.4231***
5	1.4787	1.4881	***0.4213***	***0.4218***
6	1.4775	1.4936	***0.4223***	***0.4273***
7	1.4778	1.4947	***0.4245***	***0.4289***
8	1.4777	1.4952	***0.4283***	***0.4319***
9	1.4768	1.4982	***0.4185***	***0.4216***
10	1.4779	1.4904	***0.4192***	***0.4243***
MEAN	**1.4781**	**1.4908**	***0.4208***	***0.4230***

APPENDIX O:

Results Based on RRSE of Each Fold of 10-fold Cross Validation for Readmissions Belonging to Different Major Diagnostic Group Related Analysis

Table 21 RRSE for Models Based on RADC

Table 21(a) RRSE of RADC for Regression and Generalization Based Models

Model	**Linear Regression**		**Lasso Regression**		**Elastic Net Regression**		**Ridge Regression**	
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.5550	0.5418	0.5550	0.5418	0.5550	0.5418	0.5559	0.5418
2	0.5532	0.5535	0.5533	0.5536	0.5533	0.5536	0.5542	0.5545
3	0.5534	0.5516	0.5535	0.5517	0.5535	0.5517	0.5544	0.5535
4	0.5528	0.5575	0.5529	0.5575	0.5529	0.5575	0.5538	0.5582
5	0.5530	0.5552	0.5531	0.5553	0.5531	0.5553	0.5540	0.5559
6	0.5527	0.5585	0.5528	0.5585	0.5528	0.5585	0.5537	0.5593
7	0.5532	0.5541	0.5532	0.5542	0.5532	0.5542	0.5542	0.5552
8	0.5531	0.5550	0.5531	0.5551	0.5531	0.5551	0.5541	0.5564
9	0.5532	0.5540	0.5532	0.5541	0.5532	0.5541	0.5542	0.5552
10	0.5529	0.5567	0.5529	0.5567	0.5529	0.5567	0.5539	0.5575
MEAN	**0.5532**	**0.5538**	**0.5533**	**0.5539**	**0.5533**	**0.5539**	**0.5542**	**0.5548**

Table 21(b) RRSE of RADC for Hyper Tuning From Lasso Towards Elastic Net Based Models

Model	**Hyper Tuning From Lasso Towards Elastic Net Regression**							
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.5550	0.5418	0.5550	0.5418	0.5550	0.5418	0.5550	0.5418
2	0.5533	0.5536	0.5533	0.5536	0.5533	0.5536	0.5533	0.5536
3	0.5535	0.5517	0.5535	0.5517	0.5535	0.5517	0.5535	0.5517
4	0.5529	0.5575	0.5529	0.5575	0.5529	0.5575	0.5529	0.5575
5	0.5531	0.5553	0.5531	0.5553	0.5531	0.5553	0.5531	0.5553
6	0.5528	0.5585	0.5528	0.5585	0.5528	0.5585	0.5528	0.5585
7	0.5532	0.5542	0.5532	0.5541	0.5532	0.5542	0.5532	0.5542
8	0.5531	0.5551	0.5531	0.5551	0.5531	0.5551	0.5531	0.5551
9	0.5532	0.5541	0.5532	0.5541	0.5532	0.5541	0.5532	0.5541
10	0.5529	0.5567	0.5529	0.5567	0.5529	0.5567	0.5529	0.5567
MEAN	**0.5533**	**0.5539**	**0.5533**	**0.5539**	**0.5533**	**0.5539**	**0.5533**	**0.5539**

Table 21(c) RRSE of RADC for Hyper Tuning From Elastic Net Towards Ridge Based Models

Model	**Hyper Tuning From Elastic Net Towards Ridge Regression**							
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.5550	0.5418	0.5550	0.5418	0.5550	0.5418	0.5550	0.5418
2	0.5533	0.5536	0.5533	0.5536	0.5533	0.5536	0.5533	0.5536
3	0.5535	0.5517	0.5535	0.5517	0.5535	0.5517	0.5535	0.5518
4	0.5529	0.5575	0.5529	0.5575	0.5529	0.5575	0.5529	0.5575
5	0.5531	0.5553	0.5531	0.5553	0.5531	0.5553	0.5531	0.5552
6	0.5528	0.5585	0.5528	0.5585	0.5528	0.5585	0.5528	0.5585
7	0.5532	0.5542	0.5532	0.5542	0.5532	0.5542	0.5532	0.5542
8	0.5531	0.5551	0.5531	0.5551	0.5531	0.5551	0.5531	0.5551
9	0.5532	0.5541	0.5532	0.5541	0.5532	0.5541	0.5532	0.5541
10	0.5529	0.5567	0.5529	0.5567	0.5529	0.5567	0.5529	0.5567
MEAN	**0.5533**	**0.5539**	**0.5533**	**0.5539**	**0.5533**	**0.5539**	**0.5533**	**0.5539**

Table 21(d) RRSE of RADC for Tree and Deep Learning-based Models

Model	**XGBoost**		**Deep Learning**	
Fold	**Training**	**Testing**	**Training**	**Testing**
1	1.5196	1.5212	***0.4202***	***0.4264***
2	1.5179	1.5259	***0.4233***	***0.4261***
3	1.5180	1.5268	***0.4178***	***0.4211***
4	1.5172	1.5310	***0.4251***	***0.4267***
5	1.5181	1.5241	***0.4179***	***0.4216***
6	1.5177	1.5272	***0.4195***	***0.4237***
7	1.5179	1.5327	***0.4183***	***0.4210***
8	1.5179	1.5257	***0.4164***	***0.4192***
9	1.5180	1.5250	***0.4173***	***0.4215***
10	1.5178	1.5264	***0.4183***	***0.4245***
MEAN	**1.5180**	**1.5266**	***0.4194***	***0.4232***

APPENDIX P:
Results Based on NRMSE1 of Each Fold of 10-fold Cross Validation for Readmissions Belonging to Same Major Diagnostic Group Related Analysis

Table 22 NRMSE1 for Models Based on RSDC

Table 22(a) NRMSE1 of RSDC for Regression and Generalization Based Models

Model	**Linear Regression**		**Lasso Regression**		**Elastic Net Regression**		**Ridge Regression**	
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.0529	0.0567	0.0529	0.0567	0.0529	0.0567	0.0530	0.0567
2	0.0529	0.0577	0.0529	0.0577	0.0529	0.0577	0.0530	0.0579
3	0.0532	0.0548	0.0532	0.0548	0.0532	0.0548	0.0533	0.0549
4	0.0529	0.0552	0.0529	0.0552	0.0529	0.0552	0.0529	0.0552
5	0.0529	0.0548	0.0529	0.0548	0.0529	0.0548	0.0530	0.0550
6	0.0528	0.0556	0.0529	0.0556	0.0529	0.0556	0.0529	0.0557
7	0.0531	0.0541	0.0531	0.0541	0.0531	0.0541	0.0532	0.0543
8	0.0529	0.0551	0.0529	0.0551	0.0529	0.0551	0.0529	0.0552
9	0.0529	0.0549	0.0529	0.0549	0.0529	0.0549	0.0530	0.0550
10	0.0529	0.0543	0.0529	0.0543	0.0529	0.0543	0.0530	0.0544
MEAN	**0.0529**	**0.0553**	**0.0529**	**0.0553**	**0.0529**	**0.0553**	**0.0530**	**0.0554**

Table 22(b) NRMSE1 of RSDC for Hyper Tuning From Lasso Towards Elastic Net Based Models

Model	**Hyper tuning from Lasso towards Elastic Net Regression**							
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.0529	0.0567	0.0529	0.0539	0.0529	0.0567	0.0529	0.0567
2	0.0529	0.0577	0.0529	0.0577	0.0529	0.0577	0.0529	0.0577
3	0.0532	0.0548	0.0533	0.0548	0.0532	0.0548	0.0532	0.0548
4	0.0529	0.0552	0.0529	0.0552	0.0529	0.0552	0.0529	0.0552
5	0.0529	0.0548	0.0529	0.0548	0.0529	0.0548	0.0529	0.0548
6	0.0529	0.0556	0.0529	0.0556	0.0529	0.0556	0.0529	0.0556
7	0.0531	0.0541	0.0531	0.0541	0.0531	0.0541	0.0531	0.0541
8	0.0529	0.0551	0.0529	0.0551	0.0529	0.0551	0.0529	0.0551
9	0.0529	0.0549	0.0529	0.0549	0.0529	0.0549	0.0529	0.0549
10	0.0529	0.0543	0.0529	0.0543	0.0529	0.0543	0.0529	0.0543
MEAN	**0.0529**	**0.0553**	**0.0529**	**0.0551**	**0.0529**	**0.0553**	**0.0529**	**0.0553**

Table 22(c) NRMSE1 of RSDC for Hyper Tuning From Elastic Net Towards Ridge Based Models

Model	**Hyper Tuning From Elastic Net Towards Ridge Regression**							
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.0529	0.0567	0.0529	0.0567	0.0529	0.0567	0.0529	0.0567
2	0.0529	0.0577	0.0529	0.0577	0.0529	0.0577	0.0529	0.0577
3	0.0532	0.0548	0.0533	0.0548	0.0532	0.0548	0.0533	0.0548
4	0.0529	0.0552	0.0529	0.0552	0.0529	0.0552	0.0529	0.0552
5	0.0529	0.0548	0.0529	0.0548	0.0529	0.0548	0.0529	0.0548
6	0.0529	0.0556	0.0529	0.0556	0.0529	0.0556	0.0529	0.0556
7	0.0531	0.0541	0.0531	0.0541	0.0531	0.0541	0.0531	0.0541
8	0.0529	0.0551	0.0529	0.0551	0.0529	0.0551	0.0529	0.0551
9	0.0529	0.0549	0.0529	0.0549	0.0529	0.0549	0.0529	0.0549
10	0.0529	0.0543	0.0529	0.0543	0.0529	0.0543	0.0529	0.0543
MEAN	**0.0529**	**0.0553**	**0.0529**	**0.0553**	**0.0529**	**0.0553**	**0.0529**	**0.0553**

Table 22(d) NRMSE1 of RSDC for Tree and Deep Learning-based Models

Model	**XGBoost**		**Deep Learning**	
Fold	**Training**	**Testing**	**Training**	**Testing**
1	0.1434	0.1563	***0.0410***	***0.0436***
2	0.1434	0.1580	***0.0404***	***0.0445***
3	0.1444	0.1494	***0.0405***	***0.0421***
4	0.1434	0.1501	***0.0406***	***0.0426***
5	0.1434	0.1502	***0.0409***	***0.0426***
6	0.1434	0.1508	***0.0410***	***0.0431***
7	0.1440	0.1485	***0.0414***	***0.0426***
8	0.1434	0.1501	***0.0416***	***0.0434***
9	0.1434	0.1505	***0.0406***	***0.0424***
10	0.1434	0.1478	***0.0407***	***0.0421***
MEAN	**0.1436**	**0.1512**	***0.0409***	***0.0429***

APPENDIX Q:

Results Based on NRMSE1 of Each Fold of 10-fold Cross Validation for Readmissions Belonging to Different Major Diagnostic Group Related Analysis

Table 23 NRMSE1 for Models Based on RADC

Table 23(a) NRMSE1 of RADC for Regression and Generalization Based Models

Model	Linear Regression		Lasso Regression		Elastic Net Regression		Ridge Regression	
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.0521	0.0546	0.0521	0.0546	0.0521	0.0546	0.0521	0.0546
2	0.0519	0.0553	0.0519	0.0553	0.0519	0.0553	0.0520	0.0554
3	0.0520	0.0533	0.0521	0.0533	0.0521	0.0533	0.0521	0.0535
4	0.0519	0.0535	0.0519	0.0535	0.0519	0.0535	0.0520	0.0536
5	0.0519	0.0563	0.0519	0.0563	0.0519	0.0563	0.0520	0.0564
6	0.0519	0.0549	0.0519	0.0549	0.0519	0.0549	0.0520	0.0550
7	0.0519	0.0565	0.0519	0.0565	0.0519	0.0565	0.0520	0.0566
8	0.0519	0.0548	0.0519	0.0548	0.0519	0.0548	0.0520	0.0549
9	0.0522	0.0531	0.0522	0.0531	0.0522	0.0531	0.0523	0.0532
10	0.0519	0.0537	0.0519	0.0537	0.0519	0.0537	0.0520	0.0538
MEAN	**0.0520**	**0.0546**	**0.0520**	**0.0546**	**0.0520**	**0.0546**	**0.0521**	**0.0547**

Table 23(b) NRMSE1 of RADC for Hyper Tuning From Lasso Towards Elastic Net Based Models

Model	Hyper Tuning From Lasso Towards Elastic Net Regression							
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.0521	0.0546	0.0521	0.0546	0.0521	0.0546	0.0521	0.0546
2	0.0519	0.0553	0.0519	0.0553	0.0519	0.0553	0.0519	0.0553
3	0.0521	0.0533	0.0521	0.0533	0.0521	0.0533	0.0521	0.0533
4	0.0519	0.0535	0.0519	0.0535	0.0519	0.0535	0.0519	0.0535
5	0.0519	0.0563	0.0519	0.0563	0.0519	0.0563	0.0519	0.0563
6	0.0519	0.0549	0.0519	0.0549	0.0519	0.0549	0.0519	0.0549
7	0.0519	0.0565	0.0519	0.0565	0.0519	0.0565	0.0519	0.0565
8	0.0519	0.0548	0.0519	0.0548	0.0519	0.0548	0.0519	0.0548
9	0.0522	0.0531	0.0522	0.0531	0.0522	0.0531	0.0522	0.0531
10	0.0519	0.0537	0.0519	0.0537	0.0519	0.0537	0.0519	0.0537
MEAN	**0.0520**	**0.0546**	**0.0520**	**0.0546**	**0.0520**	**0.0546**	**0.0520**	**0.0546**

Table 23(c) NRMSE1 of RADC for Hyper Tuning From Elastic Net Towards Ridge Based Models

Model	**Hyper Tuning From Elastic Net Towards Ridge Regression**							
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.0521	0.0546	0.0521	0.0546	0.0521	0.0546	0.0521	0.0546
2	0.0519	0.0553	0.0519	0.0553	0.0519	0.0553	0.0519	0.0553
3	0.0521	0.0533	0.0521	0.0533	0.0521	0.0533	0.0521	0.0533
4	0.0519	0.0535	0.0519	0.0535	0.0519	0.0535	0.0519	0.0535
5	0.0519	0.0563	0.0519	0.0563	0.0519	0.0563	0.0519	0.0563
6	0.0519	0.0549	0.0519	0.0549	0.0519	0.0549	0.0519	0.0549
7	0.0519	0.0565	0.0519	0.0565	0.0519	0.0565	0.0519	0.0565
8	0.0519	0.0548	0.0519	0.0548	0.0519	0.0548	0.0519	0.0548
9	0.0522	0.0531	0.0522	0.0531	0.0522	0.0531	0.0522	0.0531
10	0.0519	0.0537	0.0519	0.0537	0.0519	0.0537	0.0519	0.0537
MEAN	**0.0520**	**0.0546**	**0.0520**	**0.0546**	**0.0520**	**0.0546**	**0.0520**	**0.0546**

Table 23(d) NRMSE1 of RADC for Tree and Deep Learning-based Models

Model	**XGBoost**		**Deep Learning**	
Fold	**Training**	**Testing**	**Training**	**Testing**
1	0.1436	0.1538	***0.0394***	***0.0429***
2	0.1435	0.1530	***0.0397***	***0.0426***
3	0.1438	0.1480	***0.0393***	***0.0407***
4	0.1435	0.1475	***0.0399***	***0.0410***
5	0.1435	0.1551	***0.0392***	***0.0427***
6	0.1435	0.1507	***0.0394***	***0.0417***
7	0.1435	0.1563	***0.0393***	***0.0429***
8	0.1435	0.1512	***0.0391***	***0.0440***
9	0.1444	0.1467	***0.0394***	***0.0404***
10	0.1435	0.1477	***0.0393***	***0.0409***
MEAN	**0.1436**	**0.1510**	***0.0394***	***0.0420***

APPENDIX R:
Results Based on NRMSE2 of Each Fold of 10-fold Cross Validation for Readmissions Belonging to Same Major Diagnostic Group Related Analysis

Table 24 NRMSE2 for RSDC Models

Table 24(a) NRMSE2 of RSDC for Regression and Generalization Based Models

Model	**Linear Regression**		**Lasso Regression**		**Elastic Net Regression**		**Ridge Regression**	
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.0549	0.0539	0.0549	0.0539	0.0549	0.0539	0.0550	0.0538
2	0.0548	0.0548	0.0548	0.0548	0.0548	0.0548	0.0549	0.0550
3	0.0548	0.0550	0.0548	0.0550	0.0548	0.0550	0.0549	0.0551
4	0.0548	0.0552	0.0548	0.0552	0.0548	0.0552	0.0549	0.0552
5	0.0548	0.0548	0.0548	0.0549	0.0548	0.0549	0.0549	0.0550
6	0.0548	0.0552	0.0548	0.0552	0.0548	0.0552	0.0549	0.0553
7	0.0548	0.0547	0.0549	0.0547	0.0548	0.0547	0.0549	0.0549
8	0.0548	0.0552	0.0548	0.0552	0.0548	0.0552	0.0549	0.0552
9	0.0548	0.0547	0.0549	0.0547	0.0549	0.0547	0.0549	0.0547
10	0.0548	0.0551	0.0548	0.0551	0.0548	0.0551	0.0549	0.0552
MEAN	**0.0548**	**0.0549**	**0.0548**	**0.0549**	**0.0548**	**0.0549**	**0.0549**	**0.0549**

Table 24(b) NRMSE2 of RSDC for Hyper Tuning From Lasso Towards Elastic Net Based Models

Model	**Hyper Tuning From Lasso Towards Elastic Net Regression**							
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.0549	0.0539	0.0549	0.0539	0.0549	0.0539	0.0549	0.0539
2	0.0548	0.0548	0.0548	0.0548	0.0548	0.0548	0.0548	0.0548
3	0.0548	0.0550	0.0548	0.0550	0.0548	0.0550	0.0548	0.0550
4	0.0548	0.0552	0.0548	0.0552	0.0548	0.0552	0.0548	0.0552
5	0.0548	0.0549	0.0548	0.0549	0.0548	0.0549	0.0548	0.0549
6	0.0548	0.0552	0.0548	0.0552	0.0548	0.0552	0.0548	0.0552
7	0.0549	0.0547	0.0548	0.0547	0.0548	0.0547	0.0548	0.0547
8	0.0548	0.0552	0.0548	0.0552	0.0548	0.0552	0.0548	0.0552
9	0.0549	0.0547	0.0549	0.0547	0.0549	0.0547	0.0549	0.0547
10	0.0548	0.0551	0.0548	0.0551	0.0548	0.0551	0.0548	0.0551
MEAN	**0.0548**	**0.0549**	**0.0548**	**0.0549**	**0.0548**	**0.0549**	**0.0548**	**0.0549**

Table 24(c) NRMSE2 of RSDC for Hyper Tuning From Elastic Net Towards Ridge Based Models

Model	**Hyper Tuning From Elastic Net Towards Ridge Regression**							
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.0549	0.0539	0.0549	0.0539	0.0549	0.0539	0.0549	0.0539
2	0.0548	0.0548	0.0548	0.0548	0.0548	0.0548	0.0548	0.0548
3	0.0548	0.0550	0.0548	0.0550	0.0548	0.0550	0.0548	0.0550
4	0.0548	0.0552	0.0548	0.0552	0.0548	0.0552	0.0548	0.0552
5	0.0548	0.0549	0.0548	0.0549	0.0548	0.0549	0.0548	0.0549
6	0.0548	0.0552	0.0548	0.0552	0.0548	0.0552	0.0548	0.0552
7	0.0548	0.0547	0.0549	0.0547	0.0549	0.0547	0.0548	0.0547
8	0.0548	0.0552	0.0548	0.0552	0.0548	0.0552	0.0548	0.0552
9	0.0549	0.0547	0.0549	0.0547	0.0549	0.0547	0.0548	0.0547
10	0.0548	0.0551	0.0548	0.0551	0.0548	0.0551	0.0548	0.0551
MEAN	**0.0548**	**0.0549**	**0.0548**	**0.0549**	**0.0548**	**0.0549**	**0.0548**	**0.0549**

Table 24(d) NRMSE2 of RSDC for Tree and Deep Learning-based Models

Model	**XGBoost**		**Deep Learning**	
Fold	**Training**	**Testing**	**Training**	**Testing**
1	0.1489	0.1484	***0.0425***	***0.0414***
2	0.1487	0.1500	***0.0419***	***0.0423***
3	0.1487	0.1498	***0.0417***	***0.0422***
4	0.1487	0.1500	***0.0421***	***0.0426***
5	0.1487	0.1503	***0.0424***	***0.0426***
6	0.1487	0.1499	***0.0425***	***0.0429***
7	0.1487	0.1501	***0.0427***	***0.0431***
8	0.1487	0.1502	***0.0431***	***0.0434***
9	0.1487	0.1498	***0.0421***	***0.0421***
10	0.1487	0.1501	***0.0422***	***0.0427***
MEAN	**0.1487**	**0.1499**	***0.0423***	***0.0425***

APPENDIX S:

Results Based on NRMSE2 of Each Fold of 10-fold Cross Validation for Readmissions Belonging to Different Major Diagnostic Group Related Analysis

Table 25 NRMSE2 for RADC Models

Table 25(a) NRMSE2 of RADC for Regression and Generalization Based Models

Model	**Linear Regression**		**Lasso Regression**		**Elastic Net Regression**		**Ridge Regression**	
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.0543	0.0525	0.0543	0.0525	0.0543	0.0525	0.0544	0.0525
2	0.0541	0.0542	0.0541	0.0542	0.0541	0.0542	0.0542	0.0543
3	0.0541	0.0540	0.0541	0.0540	0.0541	0.0540	0.0542	0.0542
4	0.0541	0.0543	0.0541	0.0544	0.0541	0.0544	0.0542	0.0544
5	0.0541	0.0544	0.0541	0.0544	0.0541	0.0544	0.0542	0.0545
6	0.0540	0.0546	0.0540	0.0546	0.0540	0.0546	0.0541	0.0547
7	0.0541	0.0542	0.0541	0.0542	0.0541	0.0542	0.0542	0.0543
8	0.0541	0.0543	0.0541	0.0543	0.0541	0.0543	0.0542	0.0544
9	0.0541	0.0542	0.0541	0.0542	0.0541	0.0542	0.0542	0.0544
10	0.0541	0.0544	0.0541	0.0544	0.0541	0.0544	0.0542	0.0545
MEAN	**0.0541**	**0.0541**	**0.0541**	**0.0541**	**0.0541**	**0.0541**	**0.0542**	**0.0542**

Table 25(b) NRMSE2 of RADC for Hyper Tuning From Lasso Towards Elastic Net Based Models

Model	**Hyper Tuning From Lasso Towards Elastic Net Regression**							
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.0543	0.0525	0.0543	0.0525	0.0543	0.0525	0.0543	0.0525
2	0.0541	0.0542	0.0541	0.0542	0.0541	0.0542	0.0541	0.0542
3	0.0541	0.0540	0.0541	0.0540	0.0541	0.0540	0.0541	0.0540
4	0.0541	0.0544	0.0541	0.0544	0.0541	0.0543	0.0541	0.0544
5	0.0541	0.0544	0.0541	0.0544	0.0541	0.0544	0.0541	0.0544
6	0.0540	0.0546	0.0540	0.0546	0.0540	0.0546	0.0540	0.0546
7	0.0541	0.0542	0.0541	0.0542	0.0541	0.0542	0.0541	0.0542
8	0.0541	0.0543	0.0541	0.0543	0.0541	0.0543	0.0541	0.0543
9	0.0541	0.0542	0.0541	0.0542	0.0541	0.0542	0.0541	0.0542
10	0.0541	0.0544	0.0541	0.0544	0.0541	0.0544	0.0541	0.0544
MEAN	**0.0541**	**0.0541**	**0.0541**	**0.0541**	**0.0541**	**0.0541**	**0.0541**	**0.0541**

Table 25(c) NRMSE2 of RADC for Hyper Tuning From Elastic Net Towards Ridge Based Models

Model	**Hyper Tuning From Elastic Net Towards Ridge Regression**							
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.0543	0.0525	0.0543	0.0525	0.0543	0.0525	0.0543	0.0525
2	0.0541	0.0542	0.0541	0.0542	0.0541	0.0542	0.0541	0.0542
3	0.0541	0.0540	0.0541	0.0540	0.0541	0.0540	0.0541	0.0540
4	0.0541	0.0544	0.0541	0.0544	0.0541	0.0544	0.0541	0.0544
5	0.0541	0.0544	0.0541	0.0544	0.0541	0.0544	0.0541	0.0544
6	0.0540	0.0546	0.0540	0.0546	0.0540	0.0546	0.0540	0.0546
7	0.0541	0.0542	0.0541	0.0542	0.0541	0.0542	0.0541	0.0542
8	0.0541	0.0543	0.0541	0.0543	0.0541	0.0543	0.0541	0.0543
9	0.0541	0.0542	0.0541	0.0542	0.0541	0.0542	0.0541	0.0542
10	0.0541	0.0544	0.0541	0.0544	0.0541	0.0544	0.0541	0.0544
MEAN	**0.0541**	**0.0541**	**0.0541**	**0.0541**	**0.0541**	**0.0541**	**0.0541**	**0.0541**

Table 25(d) NRMSE2 of RADC for Tree and Deep Learning-based Models

Model	**XGBoost**		**Deep Learning**	
Fold	**Training**	**Testing**	**Training**	**Testing**
1	0.1486	0.1474	***0.0411***	***0.0413***
2	0.1484	0.1494	***0.0414***	***0.0417***
3	0.1484	0.1494	***0.0408***	***0.0412***
4	0.1484	0.1492	***0.0416***	***0.0416***
5	0.1484	0.1493	***0.0409***	***0.0413***
6	0.1484	0.1493	***0.0410***	***0.0414***
7	0.1484	0.1498	***0.0409***	***0.0411***
8	0.1484	0.1493	***0.0407***	***0.0410***
9	0.1484	0.1493	***0.0408***	***0.0413***
10	0.1484	0.1492	***0.0409***	***0.0415***
MEAN	**0.1484**	**0.1492**	***0.0410***	***0.0413***

APPENDIX T:
Results Based on MAD of Each Fold of 10-fold Cross Validation for Readmissions Belonging to Same Major Diagnostic Group Related Analysis

Table 26 MAD for RSDC Models

Table 26(a) MAD of RSDC for Regression and Generalization Based Models

Model	**Linear Regression**		**Lasso Regression**		**Elastic Net Regression**		**Ridge Regression**	
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.0420	0.0412	0.0420	0.0412	0.0420	0.0412	0.0422	0.0414
2	0.0419	0.0421	0.0419	0.0421	0.0419	0.0421	0.0422	0.0424
3	0.0419	0.0420	0.0419	0.0420	0.0419	0.0420	0.0422	0.0423
4	0.0419	0.0418	0.0419	0.0418	0.0419	0.0418	0.0422	0.0421
5	0.0419	0.0422	0.0419	0.0422	0.0419	0.0422	0.0422	0.0425
6	0.0419	0.0421	0.0419	0.0421	0.0419	0.0421	0.0422	0.0423
7	0.0419	0.0419	0.0419	0.0419	0.0419	0.0419	0.0422	0.0422
8	0.0419	0.0422	0.0419	0.0422	0.0419	0.0422	0.0421	0.0424
9	0.0419	0.0417	0.0419	0.0417	0.0419	0.0417	0.0422	0.0419
10	0.0419	0.0422	0.0419	0.0422	0.0419	0.0422	0.0421	0.0425
MEAN	**0.0419**	**0.0419**	**0.0419**	**0.0419**	**0.0419**	**0.0419**	**0.0422**	**0.0422**

Table 26(b) MAD of RSDC for Hyper Tuning From Lasso Towards Elastic Net Based Models

Model	**Hyper Tuning From Lasso Towards Elastic Net Regression**							
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.0420	0.0412	0.0420	0.0412	0.0420	0.0412	0.0420	0.0412
2	0.0419	0.0421	0.0419	0.0421	0.0419	0.0421	0.0419	0.0421
3	0.0419	0.0420	0.0419	0.0420	0.0419	0.0420	0.0419	0.0420
4	0.0419	0.0418	0.0419	0.0418	0.0419	0.0418	0.0419	0.0418
5	0.0419	0.0422	0.0419	0.0422	0.0419	0.0422	0.0419	0.0422
6	0.0419	0.0421	0.0419	0.0421	0.0419	0.0421	0.0419	0.0421
7	0.0419	0.0419	0.0419	0.0419	0.0419	0.0419	0.0419	0.0419
8	0.0419	0.0422	0.0419	0.0422	0.0419	0.0422	0.0419	0.0422
9	0.0419	0.0417	0.0419	0.0417	0.0419	0.0417	0.0419	0.0417
10	0.0419	0.0422	0.0419	0.0422	0.0419	0.0422	0.0419	0.0422
MEAN	**0.0419**	**0.0419**	**0.0419**	**0.0419**	**0.0419**	**0.0419**	**0.0419**	**0.0419**

Table 26(c) MAD of RSDC for Hyper Tuning From Elastic Net Towards Ridge Based Models

Model	**Hyper Tuning From Elastic Net Towards Ridge Regression**							
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.0420	0.0412	0.0420	0.0412	0.0420	0.0412	0.0420	0.0412
2	0.0419	0.0421	0.0419	0.0421	0.0419	0.0421	0.0419	0.0421
3	0.0419	0.0420	0.0419	0.0420	0.0419	0.0420	0.0419	0.0420
4	0.0419	0.0418	0.0419	0.0418	0.0419	0.0418	0.0419	0.0418
5	0.0419	0.0422	0.0419	0.0422	0.0419	0.0422	0.0419	0.0422
6	0.0419	0.0421	0.0419	0.0421	0.0419	0.0421	0.0419	0.0421
7	0.0419	0.0419	0.0420	0.0419	0.0420	0.0419	0.0419	0.0419
8	0.0419	0.0422	0.0419	0.0422	0.0419	0.0422	0.0419	0.0422
9	0.0419	0.0417	0.0419	0.0417	0.0419	0.0417	0.0419	0.0417
10	0.0419	0.0422	0.0419	0.0422	0.0419	0.0422	0.0419	0.0422
MEAN	**0.0419**	**0.0419**	**0.0419**	**0.0419**	**0.0419**	**0.0419**	**0.0419**	**0.0420**

Table 26(d) MAD of RSDC for Tree and Deep Learning-based Models

Model	**XGBoost**		**Deep Learning**	
Fold	**Training**	**Testing**	**Training**	**Testing**
1	0.1461	0.1448	***0.0327***	***0.0312***
2	0.1459	0.1462	***0.0321***	***0.0324***
3	0.1459	0.1460	***0.0319***	***0.0324***
4	0.1459	0.1461	***0.0323***	***0.0326***
5	0.1459	0.1460	***0.0324***	***0.0326***
6	0.1459	0.1460	***0.0325***	***0.0327***
7	0.1459	0.1462	***0.0327***	***0.0330***
8	0.1459	0.1463	***0.0330***	***0.0332***
9	0.1459	0.1459	***0.0323***	***0.0322***
10	0.1459	0.1462	***0.0322***	***0.0327***
MEAN	**0.1459**	**0.1460**	***0.0324***	***0.0325***

APPENDIX U:
Results based on MAD of Each Fold of 10-fold Cross Validation for Readmissions Belonging to Different Major Diagnostic Group Related Analysis

Table 27 MAD for RADC Models

Table 27(a) MAD of RADC for Regression and Generalization Based Models

Model	Linear Regression		Lasso Regression		Elastic Net Regression		Ridge Regression	
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.0415	0.0401	0.0415	0.0401	0.0415	0.0401	0.0417	0.0403
2	0.0413	0.0414	0.0413	0.0414	0.0413	0.0414	0.0416	0.0417
3	0.0413	0.0415	0.0413	0.0415	0.0413	0.0415	0.0416	0.0418
4	0.0413	0.0414	0.0413	0.0414	0.0413	0.0414	0.0416	0.0416
5	0.0413	0.0414	0.0413	0.0414	0.0413	0.0414	0.0416	0.0417
6	0.0413	0.0416	0.0413	0.0416	0.0413	0.0416	0.0416	0.0419
7	0.0413	0.0414	0.0413	0.0414	0.0413	0.0414	0.0416	0.0414
8	0.0413	0.0417	0.0413	0.0417	0.0413	0.0417	0.0416	0.0419
9	0.0413	0.0415	0.0413	0.0415	0.0413	0.0415	0.0416	0.0418
10	0.0413	0.0415	0.0413	0.0415	0.0413	0.0415	0.0416	0.0418
MEAN	**0.0413**	**0.0413**	**0.0413**	**0.0414**	**0.0413**	**0.0414**	**0.0416**	**0.0416**

Table 27(b) MAD of RADC for Hyper Tuning From Lasso Towards Elastic Net Based Models

Model	Hyper Tuning From Lasso Towards Elastic Net Regression							
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.0415	0.0401	0.0415	0.0401	0.0415	0.0401	0.0415	0.0401
2	0.0413	0.0414	0.0413	0.0414	0.0413	0.0414	0.0413	0.0414
3	0.0413	0.0415	0.0413	0.0415	0.0413	0.0415	0.0413	0.0415
4	0.0413	0.0414	0.0413	0.0414	0.0413	0.0414	0.0413	0.0414
5	0.0413	0.0414	0.0413	0.0414	0.0413	0.0414	0.0413	0.0414
6	0.0413	0.0416	0.0413	0.0416	0.0413	0.0416	0.0413	0.0416
7	0.0413	0.0414	0.0413	0.0414	0.0413	0.0414	0.0413	0.0414
8	0.0413	0.0417	0.0413	0.0417	0.0413	0.0417	0.0413	0.0417
9	0.0413	0.0415	0.0413	0.0415	0.0413	0.0415	0.0413	0.0415
10	0.0413	0.0415	0.0413	0.0415	0.0413	0.0415	0.0413	0.0415
MEAN	**0.0413**	**0.0414**	**0.0413**	**0.0414**	**0.0413**	**0.0414**	**0.0413**	**0.0414**

Table 27(c) MAD of RADC for Hyper Tuning From Elastic Net Towards Ridge Based Models

Model	**Hyper Tuning From Elastic Net Towards Ridge Regression**							
Fold	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
1	0.0415	0.0401	0.0415	0.0401	0.0415	0.0401	0.0415	0.0402
2	0.0413	0.0414	0.0413	0.0414	0.0413	0.0414	0.0413	0.0414
3	0.0413	0.0415	0.0413	0.0415	0.0413	0.0415	0.0414	0.0415
4	0.0413	0.0414	0.0413	0.0414	0.0413	0.0414	0.0413	0.0414
5	0.0413	0.0414	0.0413	0.0414	0.0413	0.0414	0.0413	0.0414
6	0.0413	0.0416	0.0413	0.0416	0.0413	0.0416	0.0413	0.0416
7	0.0413	0.0414	0.0413	0.0414	0.0413	0.0414	0.0413	0.0414
8	0.0413	0.0417	0.0413	0.0417	0.0413	0.0417	0.0413	0.0417
9	0.0413	0.0415	0.0413	0.0415	0.0413	0.0415	0.0413	0.0415
10	0.0413	0.0415	0.0413	0.0415	0.0413	0.0415	0.0413	0.0415
MEAN	**0.0413**	**0.0414**	**0.0413**	**0.0414**	**0.0413**	**0.0414**	**0.0413**	**0.0414**

Table 27(d) MAD of RADC for Tree and Deep Learning-based Models

Model	**XGBoost**		**Deep Learning**	
Fold	**Training**	**Testing**	**Training**	**Testing**
1	0.1456	0.1441	***0.0315***	***0.0319***
2	0.1454	0.1457	***0.0318***	***0.0320***
3	0.1455	0.1458	***0.0313***	***0.0316***
4	0.1454	0.1457	***0.0319***	***0.0319***
5	0.1454	0.1456	***0.0314***	***0.0317***
6	0.1454	0.1457	***0.0315***	***0.0318***
7	0.1455	0.1457	***0.0314***	***0.0316***
8	0.1454	0.1456	***0.0312***	***0.0315***
9	0.1454	0.1457	***0.0313***	***0.0316***
10	0.1454	0.1456	***0.0314***	***0.0318***
MEAN	**0.1455**	**0.1455**	***0.0315***	***0.0317***

APPENDIX V:

Variables Used in Essay 2 Along With Their Descriptions

(Variable Names and Their Explanations are Retrieved From AHRQ Website)[4]

Table 28 Variables and Their Descriptions Used in Essay 2

Variable Name	**Explanation**
AGE	Age in years of a patient
AWEEKEND	Indicates if the admission took place on a weekend: (1) yes, (0) no
DISCWT	NRD discharge weight to be used for calculating national estimates
DISPUNIFORM	Indicates the disposition status of a patient: (1) routine, (2) transfer to short term hospital, (5) other transfers, including skilled nursing facility, intermediate care, and another type of facility, (6) home health care, (7) against medical advice, (20) died in hospital, (99) discharged alive, destination unknown
DQTR	Indicates the quarter of the year: (1) Jan–Mar, (2) Apr–Jun, (3) Jul–Sep, (4) Oct–Dec
ELECTIVE	Indicates the elective status of an admission: (1) yes, (0) no
FEMALE	Indicates the sex of a patient:(0) male, (1) female.
HCUP_ED	Indicates if the discharge record included emergency department (ED) services: (0) record does not meet any HCUP ED criteria, (1) ED revenue code was on State Inpatient Database (SID) record, (2) ED charge reported on SID record, (3) ED CPT procedure code on SID record, (4) other indication of ED services
PREV_LOS	Length of stay from previous admission
MDC	MDC that was in use on the discharge date. The details of the MDC categories are given in Table 9 of Appendix C.
NCHRONIC	Number of chronic conditions
NDX	Number of diagnoses coded
NPR	Number of procedures coded
NECODE	Number of external causes of injury codes coded
ORPROC	Indicates if the discharge record has a major operating room procedure: (1) yes, (0) no
PAY1	Indicates the type of insurance: (1) Medicare, (2) Medicaid, (3) private insurance, (4) uninsured (self-pay), (5) uninsured (no charge), (6) other
PL_NCHS	Indicates the patient location using the National Center for Health Statistics (NCHS) urban-rural classification scheme for U.S. counties: (1) "Central" counties of metro areas of >=1 million population, (2) "Fringe" counties of metro areas of >=1 million population, (3) counties in metro areas of 250,000–999,999 population, (4) counties in metro areas of 50,000–249,999

[4] https://www.hcup-us.ahrq.gov/db/nation/nrd/nrddde.jsp

	population, (5) micropolitan counties, (6) not metropolitan or micropolitan counties
REHABTRANSFER	Indicates if the record had transfer to a rehabilitation, evaluation, or other aftercare: (1) yes, (0) no
RESIDENT	Indicates if a patient is a resident of the State in which he or she received hospital care: (1) yes, (0) no
SAMEDAYEVENT	Indicates the same day event status of a patient: (0) not a combined transfer or other same-day stay record, (1) combined transfer involving two discharges from different hospitals, (2) combined same-day stay involving two discharges at different hospitals, (3) combined same-day stay involving two discharges at the same hospital, (4) combined same-day stay involving three or more discharges at same or different hospitals
ZIPINC_QRTL	Indicates the median household income quartiles for patient's ZIP code. For 2013, the median income quartiles are defined as: (1) $1 - $37,999; (2) $38,000 - $47,999; (3) $48,000 - $63,999; and (4) $64,000 or more.
TOTCHG	Costs billed by the hospital for readmissions
PREVCHG	Costs billed by the hospital for the last previous admission
HOSP_BEDSIZE	Indicates the size of a hospital based on the number of beds: (1) small, (2) medium, (3) large.
HOSP_UR_TEACH	Indicates the teaching status of the hospital: (0) metropolitan non-teaching, (1) metropolitan teaching, (2) non-metropolitan
HOSP_URCAT4	Indicates the urban-rural location of the hospital: (1) large metropolitan areas with at least 1 million residents, (2) small metropolitan areas with less than 1 million residents, (3) micropolitan areas, (4) not metropolitan or micropolitan
H_CONTRL	Indicates the control/ownership of the hospital: (1) government, nonfederal [public]; (2) private, not-for-profit [voluntary]; (3) private, investor-owned [proprietary]
TOTAL_DISC	Total number of discharges for this hospital in the NRD
APRDRG_Risk_Mortality	Indicates the 3M All Patient Refined DRG: Risk of mortality subclass: (0) No class specified, (1) Minor likelihood of dying, (2) Moderate likelihood of dying, (3) Major likelihood of dying, (4) Extreme likelihood of dying
APRDRG_Severity	Indicates the 3M All Patient Refined DRG: Severity of illness subclass: (0) No class specified, (1) Minor loss of function (includes cases with no comorbidity or complications), (2) Moderate loss of function, (3) Major loss of function, (4) Extreme loss of function

APPENDIX W:

Results of Essay 2 for Each Fold of 10-fold Cross Validation for Readmissions Belonging to Diseases and Disorders of the Circulatory System

Table 29 Overall Results of Essay 2 for Circulatory MDC

		Multitask Learning Algorithm (lambda1= 0.221 and lambda2 = 0.225)						Linear Regression				Logistic Regression	
		Metric											
		RMSE		MAE		Misclassification Error		RMSE		MAE		Misclassification Error	
Fold	Task	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing
1	R	-	-	-	-	0.419	0.444	-	-	-	-	0.429	0.445
1	RC	0.120	0.122	0.094	0.096	-	-	0.117	0.120	0.092	0.096	-	-
1	LOS	0.438	0.453	0.347	0.354	-	-	0.434	0.450	0.344	0.354	-	-
2	R	-	-	-	-	0.431	0.448	-	-	-	-	0.427	0.457
2	RC	0.120	0.119	0.094	0.092	-	-	0.117	0.115	0.093	0.089	-	-
2	LOS	0.439	0.438	0.346	0.355	-	-	0.436	0.435	0.344	0.354	-	-
3	R	-	-	-	-	0.423	0.420	-	-	-	-	0.434	0.420
3	RC	0.120	0.117	0.094	0.093	-	-	0.117	0.114	0.093	0.091	-	-
3	LOS	0.442	0.421	0.344	0.332	-	-	0.435	0.438	0.354	0.347	-	-
4	R	-	-	-	-	0.419	0.395	-	-	-	-	0.432	0.419
4	RC	0.120	0.120	0.094	0.094	-	-	0.117	0.120	0.092	0.096	-	-
4	LOS	0.438	0.448	0.347	0.351	-	-	0.435	0.446	0.345	0.349	-	-
5	R	-	-	-	-	0.414	0.423	-	-	-	-	0.428	0.453
5	RC	0.121	0.115	0.094	0.090	-	-	0.117	0.112	0.093	0.088	-	-
5	LOS	0.440	0.436	0.348	0.342	-	-	0.436	0.432	0.346	0.338	-	-
6	R	-	-	-	-	0.424	0.415	-	-	-	-	0.432	0.427
6	RC	0.120	0.119	0.094	0.092	-	-	0.117	0.115	0.093	0.095	-	-
6	LOS	0.440	0.433	0.348	0.344	-	-	0.437	0.427	0.345	0.342	-	-
7	R	-	-	-	-	0.431	0.419	-	-	-	-	0.432	0.415
7	RC	0.120	0.127	0.094	0.094	-	-	0.117	0.118	0.092	0.093	-	-
7	LOS	0.438	0.452	0.346	0.358	-	-	0.434	0.448	0.344	0.354	-	-
8	R	-	-	-	-	0.428	0.415	-	-	-	-	0.431	0.437
8	RC	0.120	0.122	0.094	0.097	-	-	0.117	0.118	0.093	0.094	-	-
8	LOS	0.441	0.425	0.348	0.342	-	-	0.437	0.421	0.346	0.338	-	-
9	R	-	-	-	-	0.420	0.424	-	-	-	-	0.431	0.433
9	RC	0.120	0.120	0.094	0.096	-	-	0.117	0.118	0.092	0.094	-	-
9	LOS	0.441	0.429	0.348	0.338	-	-	0.437	0.427	0.346	0.338	-	-
10	R	-	-	-	-	0.414	0.411	-	-	-	-	0.432	0.426
10	RC	0.120	0.123	0.094	0.095	-	-	0.117	0.119	0.092	0.095	-	-
10	LOS	0.438	0.455	0.346	0.357	-	-	0.434	0.450	0.344	0.354	-	-

R – Readmissions; RC – Readmission Costs; LOS – Readmitted Length of Stay

APPENDIX X:

Results of Essay 2 for Each Fold of 10-fold Cross Validation for Readmissions Belonging to Diseases and Disorders of the Respiratory System

Table 30 Overall Results of Essay 2 for Respiratory MDC

		Multitask Learning Algorithm (lambda1= 0.125 and lambda2 = 0.125)						Linear Regression				Logistic Regression	
		RMSE		MAE		Misclassification Error		RMSE		MAE		Misclassification Error	
Fold	Task	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing
1	R	-	-	-	-	0.442	0.459	-	-	-	-	0.434	0.441
1	RC	0.094	0.095	0.073	0.075	-	-	0.098	0.103	0.077	0.080	-	-
1	LOS	0.377	0.363	0.292	0.286	-	-	0.379	0.369	0.294	0.291	-	-
2	R	-	-	-	-	0.453	0.447	-	-	-	-	0.441	0.426
2	RC	0.094	0.095	0.072	0.074	-	-	0.099	0.098	0.077	0.078	-	-
2	LOS	0.375	0.380	0.291	0.298	-	-	0.378	0.380	0.294	0.298	-	-
3	R	-	-	-	-	0.450	0.441	-	-	-	-	0.439	0.442
3	RC	0.094	0.096	0.073	0.073	-	-	0.099	0.098	0.077	0.076	-	-
3	LOS	0.374	0.388	0.291	0.298	-	-	0.377	0.389	0.293	0.300	-	-
4	R	-	-	-	-	0.448	0.437	-	-	-	-	0.439	0.443
4	RC	0.095	0.093	0.073	0.071	-	-	0.099	0.097	0.077	0.077	-	-
4	LOS	0.377	0.365	0.292	0.287	-	-	0.379	0.367	0.294	0.290	-	-
5	R	-	-	-	-	0.453	0.422	-	-	-	-	0.438	0.414
5	RC	0.094	0.099	0.072	0.076	-	-	0.098	0.104	0.077	0.081	-	-
5	LOS	0.375	0.379	0.291	0.289	-	-	0.378	0.382	0.294	0.293	-	-
6	R	-	-	-	-	0.446	0.463	-	-	-	-	0.437	0.453
6	RC	0.095	0.093	0.073	0.071	-	-	0.099	0.097	0.077	0.073	-	-
6	LOS	0.376	0.376	0.291	0.292	-	-	0.378	0.379	0.294	0.295	-	-
7	R	-	-	-	-	0.446	0.456	-	-	-	-	0.438	0.445
7	RC	0.094	0.097	0.073	0.076	-	-	0.099	0.100	0.077	0.078	-	-
7	LOS	0.375	0.383	0.291	0.299	-	-	0.378	0.385	0.293	0.301	-	-
8	R	-	-	-	-	0.443	0.441	-	-	-	-	0.436	0.436
8	RC	0.095	0.091	0.073	0.071	-	-	0.099	0.969	0.077	0.077	-	-
8	LOS	0.378	0.356	0.293	0.277	-	-	0.380	0.360	0.295	0.281	-	-
9	R	-	-	-	-	0.446	0.457	-	-	-	-	0.436	0.435
9	RC	0.094	0.094	0.073	0.072	-	-	0.099	0.097	0.077	0.075	-	-
9	LOS	0.374	0.395	0.291	0.299	-	-	0.376	0.396	0.293	0.301	-	-
10	R	-	-	-	-	0.443	0.464	-	-	-	-	0.436	0.447
10	RC	0.095	0.091	0.073	0.071	-	-	0.099	0.097	0.077	0.075	-	-
10	LOS	0.377	0.370	0.292	0.288	-	-	0.379	0.372	0.294	0.290	-	-

R– Readmissions; RC – Readmission Costs; LOS – Readmitted Length of Stay

APPENDIX Y:

Results of Essay 2 for Each Fold of 10-fold Cross Validation for Readmissions Belonging to Diseases and Disorders of the Digestive System

Table 31 Overall Results of Essay 2 for Digestive MDC

		Multitask Learning Algorithm (lambda1= 0.13 and lambda2 = 0.13)						Linear Regression				Logistic Regression	
		RMSE		MAE		Misclassification Error		RMSE		MAE		Misclassification Error	
Fold	Task	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing
1	R	-	-	-	-	0.404	0.418	-	-	-	-	0.415	0.435
1	RC	0.100	0.106	0.077	0.082	-	-	0.102	0.107	0.080	0.083	-	-
1	LOS	0.388	0.394	0.301	0.310	-	-	0.389	0.396	0.304	0.310	-	-
2	R	-	-	-	-	0.404	0.395	-	-	-	-	0.418	0.420
2	RC	0.102	0.100	0.078	0.079	-	-	0.103	0.102	0.080	0.080	-	-
2	LOS	0.388	0.395	0.301	0.306	-	-	0.389	0.396	0.304	0.310	-	-
3	R	-	-	-	-	0.402	0.403	-	-	-	-	0.418	0.414
3	RC	1.102	0.102	0.079	0.076	-	-	0.103	0.102	0.083	0.077	-	-
3	LOS	0.389	0.382	0.302	0.298	-	-	0.391	0.383	0.305	0.300	-	-
4	R	-	-	-	-	0.406	0.411	-	-	-	-	0.417	0.426
4	RC	0.100	0.102	0.077	0.079	-	-	0.103	0.104	0.080	0.082	-	-
4	LOS	0.388	0.390	0.302	0.306	-	-	0.390	0.391	0.304	0.309	-	-
5	R	-	-	-	-	0.403	0.398	-	-	-	-	0.419	0.412
5	RC	0.102	0.098	0.078	0.077	-	-	0.103	0.099	0.080	0.078	-	-
5	LOS	0.390	0.370	0.304	0.286	-	-	0.392	0.372	0.306	0.288	-	-
6	R	-	-	-	-	0.406	0.410	-	-	-	-	0.421	0.410
6	RC	0.101	0.102	0.078	0.079	-	-	0.103	0.104	0.080	0.082	-	-
6	LOS	0.388	0.394	0.301	0.307	-	-	0.389	0.396	0.304	0.310	-	-
7	R	-	-	-	-	0.405	0.423	-	-	-	-	0.416	0.429
7	RC	0.102	0.099	0.079	0.078	-	-	0.103	0.102	0.080	0.079	-	-
7	LOS	0.390	0.373	0.303	0.291	-	-	0.392	0.373	0.306	0.292	-	-
8	R	-	-	-	-	0.400	0.415	-	-	-	-	0.415	0.436
8	RC	0.102	0.101	0.079	0.076	-	-	0.103	0.102	0.080	0.078	-	-
8	LOS	0.387	0.400	0.300	0.313	-	-	0.388	0.401	0.303	0.315	-	-
9	R	-	-	-	-	0.406	0.401	-	-	-	-	0.418	0.412
9	RC	0.102	0.103	0.078	0.080	-	-	0.103	0.105	0.080	0.082	-	-
9	LOS	0.389	0.381	0.303	0.293	-	-	0.391	0.383	0.306	0.296	-	-
10	R	-	-	-	-	0.408	0.379	-	-	-	-	0.422	0.375
10	RC	0.102	0.101	0.079	0.078	-	-	0.103	0.101	0.080	0.080	-	-
10	LOS	0.387	0.402	0.301	0.309	-	-	0.388	0.403	0.304	0.313	-	-

R – Readmissions; RC – Readmission Costs; LOS – Readmitted Length of Stay

CURRICULUM VITAE

DEEPIKA GOPUKUMAR

EDUCATION --

2015-2021 ***Ph.D.:*** ***Management Science***
University of Wisconsin-Milwaukee
Dissertation Title: Two Essays on Leveraging Analytics to Improve Healthcare
Dissertation Committee: Dr. Huimin Zhao (Chair), Dr. Abhijeet Ghoshal, Dr. Sanjoy Ghose, and Dr. Yang Wang

2010-2012 ***Master of Technology (MTech.): Software Development and Management***
Vellore Institute of Technology, India

2006-2010 ***Bachelor of Engineering: Computer Science and Engineering***
Anna University, India

ACADEMIC WORK EXPERIENCE ---

Sep 2015 - Aug 2020 ***Graduate Assistant***, Sheldon B. Lubar School of Business, UWM

- **Instructor**
  -> ***BUS ADM 530: Introduction to e-Business*** *for undergraduate ITM students*
  ->***BUS ADM 533: Information Technology Infrastructure for Business*** *for undergraduate ITM students*
  -> ***BUS ADM 746: Artificial Intelligence for Business*** (Topics in ITM) *for graduate students*
- **Lead Teaching Assistant** for ***BUS ADM 230: Introduction to Information Technology & Management***
- **Teaching Assistant** for ***BUS ADM 230: Introduction to Information Technology & Management***
- **Grader** for ***BUS ADM 535: Global Information Technology Management***
- **Research Assistant**

INDUSTRY WORK EXPERIENCE ---

July 2010- August 2013 ***Junior Product Specialist / Programmer Analyst (Microsoft Business Intelligence Developer)***, Cognizant Technology Solutions, India

- Worked on different projects for US Healthcare Insurance using Healthcare Products like Facets and QNXT.
- Strong knowledge in Software Development Life Cycle (SDLC) with extensive experience in Requirement Analysis, Design, Development, and Testing.
- Comprehensive experience in writing complex stored procedures using SQL queries, SSIS package development, and SSRS reports development by using SQL Server 2008.
- Mentored and assisted team members to understand client business requirements and provided technical support for writing SQL queries for reports development.

AWARDS AND HONORS ---

2018 –2019 ***Sheldon B. Lubar Doctoral Scholarship*** from Sheldon B. Lubar School of Business

2015 –2019 ***Chancellor Graduate Student Award*** from the University of Wisconsin-–Milwaukee

2012 ***Awarded Associate of the Month*** for the Month of November by Cognizant Technology Solutions for exceptional performance

2010 ***Certificate of appreciation*** from Cognizant Technology Solutions for exemplary academic performance during training period on .Net Track

2010 ***Certificate of appreciation*** from Cognizant Technology Solutions for being overall topper in Dovetail Exams which test the skills of employees related to the technologies they work in their project for their business unit.