# Applying Predictive Analytic Capabilities for CardioMEMS Patients

Mark E. Riley

Department of Data Science, University of Wisconsin – La Crosse

DS785: Capstone

Dr. Jennifer Garland

August 5, 2020

## Acknowledgements

I would like to thank the faculty, staff, and my fellow students in the University of Wisconsin Master’s in Data Science program for their outstanding support throughout my journey.

Thank you to The University of Kansas Health System, especially Mr. Dongsul Kim and Mr. Casey Bryson for their partnership on this capstone project.

A huge thanks to my family, especially my wife Christine and my son Liam for their support, encouragement, and understanding as I worked many nights and weekends for two years to complete this degree.

## Abstract

Heart failure is a chronic disease in which the heart is unable to supply a sufficient volume of blood to the body. Patients with heart failure experience frequent hospitalizations and trips to the emergency room. Although heart failure hospitalizations have decreased in the Kansas City metro area in recent years, the disease is still a significant societal burden and the leading cause of death in the State of Kansas. The University of Kansas Health System (TUKHS) is treating a number of advanced heart failure patients with CardioMEMS, a micro-electromechanical device implanted into the pulmonary artery to measure hemodynamic pressures and heart rates. This analysis aimed to determine if we could predict four adverse heart failure-related events by employing supervised classification algorithms with interpretable results using a combination of data from the CardioMEMS devices and patients' electronic health records. The adverse health events in scope for this analysis were all-cause hospitalizations, heart failure-related hospitalizations, use of intravenous diuretic therapy outside of a hospitalization, and changes to pulmonary artery diastolic (PAD) pressure thresholds.

Data was extracted from the TUKHS's Epic electronic health record (EHR) system, deidentified, and exported to comma separated value (CSV) files for analysis. The patient cohort consisted of CardioMEMS patients (n = 88) who were enrolled in the TUKHS program between April 2019 and June 2020. After fixing errors, imputing some missing data, and performing feature engineering, we aggregated and combined the timeseries device data with the patient level risk factors for predictive modeling. Based on our goals and dataset characteristics we chose logistic regression, random forest, and boosted decision trees for our classification algorithms.

Our predictive models were successful against test data for three out of the four targeted adverse events. The all-cause hospitalizations model had an AUC of 0.944 using boosting with all predictor variables and scaled training data. The heart failure hospitalizations model had an AUC of 0.88 also using boosting with all predictor variables and scaled training data. The IV diuretics model had an AUC of 0.987 using boosted trees with select predictor variables and scaled training data. Our predictive model for PAD threshold changes was not successful. The best algorithm produced an AUC of only 0.767 using post-stepwise logistic regression and scaled training data.

Our analysis also produced a number of significant features related to our four targeted events that can be monitored by the clinical team at TUKHS. If and when the successful models are put into use, our analysis could help focus the care of the clinical team towards patients with the greatest risk of adverse events, reduce healthcare costs, and improve the quality of life for CardioMEMS heart failure patients.

## Contents

# List of Figures

## List of Tables

## Applying Predictive Analytic Capabilities for CardioMEMS Patients

Frequent trips to the emergency room (ER) and hospitalizations are common for people with heart failure, a chronic disease in which the heart doesn't pump enough blood to its organs. Due to both the prevalence of heart failure and the frequent ER visits and hospitalizations associated with heart failure, it is the single largest item in the Medicare budget at $50B annually (CardioMEMS, Inc., 2014). According to Kansas Health Matters, congestive heart failure hospital admission rates have fallen over the last 10 years in the Kansas City metro area, down to 10.1 per 10,000 in population (KansasHealthMatters, 2020). However, it is still the leading cause of death in the State of Kansas and a significant burden to the healthcare system and society (Kansas Department of Health and Environment Division of Public Health Bureau of Epidemiology and Public Health Informatics, 2018).

The ability to predict adverse health events in advance could provide a chance for care providers to proactively administer treatments that reduce the risks to patients. Reducing avoidable health events could contribute to reductions in healthcare costs and improvements to the quality of life for those with heart failure.

CardioMEMS is a product for heart failure patients developed by Abbott.
A MEMS (micro-electromechanical system) is a miniature machine that has both mechanical and electronic components. CardioMEMS devices are implanted into the pulmonary arteries (PA) of patients with advanced heart failure to measure pulmonary artery pressures. Cardiologists can remotely monitor their patients who have the device implanted and adjust treatments to try to prevent adverse health events.

The initial clinical trial of the CardioMEMS device, also known at the CHAMPION (CardioMEMS Heart Sensor Allows Monitoring of Pressure to Improve Outcomes in NYHA Class III Heart Failure Patients) trial showed that use of the CardioMEMS device reduced hospitalization rates by 28% (Givertz, et al., 2017). Additional post-approval study in a commercial setting found even further reduction in hospitalizations for CardioMEMS patients than the CHAMPION trial (Raval, et al., 2017).

The University of Kansas Health System (TUKHS) has treated approximately 88 patients with the CardioMEMS device in the last 15 months. Using data from the CardioMEMS database and the health system's electronic medical record (EMR) system, we will attempt to predict which patients are at risk for targeted adverse events.

To do this we will use the following variables from (or derived from) the CardioMEMS devices:

- pulmonary artery diastolic (PAD) pressure
- pulmonary artery systolic (PAS) pressure
- pulsatility, PAS - PAD
- pulmonary artery mean pressure, 2/3 PAD + 1/3 PAS
- heart rate
- PAD upper and lower threshold values
- PAD flag, indicating if a patient's PAD was outside of the threshold range
- threshold change indicator, if a patient's thresholds were changed since the previous device measurement

In addition to the data from the CardioMEMS devices, we will also include over 50 further variables from patients' electronic medical records. Examples include comorbidities such as hypertension, diabetes, and obesity, demographic data such as age and sex, and quantitative measurements such as the number of heart related hospitalizations.

We will develop four different models that will attempt to identify if a patient is at risk for one of the following adverse events using various classification algorithms and evaluating which perform better and which are more practical.

- Model 1: All-cause hospitalizations
- Model 2: Heart failure-related hospitalizations
- Model 3: IV diuretics outside of hospitalization
- Model 4: Pulmonary artery diastolic threshold adjustments

Hospitalizations for heart failure or any other reason are very costly and signal deterioration in the patient's health and quality of life. Physicians typically use medications, such as diuretics, to remove fluids that build up behind the heart. Patients requiring intravenous (IV) diuretics outside of a hospitalization signal deteriorating health and involve strict monitoring of the patient. Finally, if a patient's PAD threshold is adjusted by his cardiologist, that could be a sign of a poor prognosis. Cardiologists treating patients with advanced heart failure establish thresholds for PAD pressure. The thresholds are adjusted if the patient's heart failure condition is improving or worsening.

This analysis will also further the data-driven decision-making process in treating advanced heart failure. Prior to the use of the CardioMEMS device, cardiologists had very limited data on which to treat their heart failure patients. The CardioMEMS device now provides valuable data on which cardiologists can assess their patients' conditions and adjust treatments as necessary. This project will provide additional valuable information about our ability to predict who has higher or lower risk of targeted events to guide us on where to focus our limited resources by paying more attention on high risk patients. This information will ideally be useful to identify increased risk for adverse events earlier and lead to more effective patient care. If we are successful it would contribute lowering healthcare costs by decreasing the number of avoidable hospitalizations and improving the quality of life for heart failure patients.

## Literature Review

Our main objective in this analysis was to determine if we can predict which patients are likely to have a heart failure-related adverse health event in the future using data from the CardioMEMS devices combined with selected risk factors from the patients' medical records. We start by presenting a review of results of using telehealth to monitor heart failure patients to reduce hospitalizations and other adverse events. We will also review studies that have tried to predict adverse events in heart failure patients to analyze that variables that were most valuable for predictive ability. Our review of the literature revealed gaps including a lack of studies using CardioMEMS data to predict adverse events, and the scope of the adverse events studied limited to hospitalizations and mortality.

### Telehealth Monitoring of Heart Failure Patients

There have been several studies conducted on the relationship between the remote monitoring of heart failure patients using telehealth and the risk of hospitalization and mortality. A systems-based analysis showed that CardioMEMS patients who transmit data more frequently, and have providers who review that data more frequently, have fewer heart failure hospitalization days than those who do not (Tran, Wolfson, O'Brien, Yousefian, & Shavelle, 2019). Koehler, et al. (2018) also showed that using telehealth to monitor heart failure patients could reduce unplanned cardiac-related hospitalizations and all-cause mortality when combined with patient-tailored risk profiles and a 24-hour, seven-day telemedical center staffed with physicians and nurses. Heywood, et al. (2017) demonstrated that 2,000 patients with CardioMEMS devices remained compliant with their transmission of device data (average upload every 1.2 days) in a general practice setting even six months or more after device implantation. Those patients also continued to see significantly reduced pulmonary artery pressures through remote monitoring via the CardioMEMS device, which lowered the risks of hospitalizations and mortality. An intervention review of randomized controlled trials by Inglis, Clark, Dierckx, Prieto-Merino, & Cleland (2015) found that heart failure patient telemonitoring via non-invasive devices reduced both heart failure-related hospitalizations and all-cause mortality, but was not effective for reducing all-cause hospitalizations. Lastly, a review by Veenis & Brugts (2020) found great promise in the CardioMEMS device's ability to enable timely interventions for increases in patients' pulmonary artery pressures.

The literature shows that telehealth monitoring of heart failure patients using either invasive devices such as the CardioMEMS, or non-invasive is effective in reducing mortality and heart failure hospitalizations, backing up the findings of the CHAMPION trial. Adherence to frequently transmitting device data and frequent reviews by the care team are also significant to reducing adverse events.

However, these studies only examined outcomes of either hospitalizations, or hospitalizations and mortality. There were no studies showing to examine the effects of telehealth monitoring on other adverse outcomes such as altering PAD thresholds or the use of diuretic therapy outside of a hospitalization. These studies also did not attempt to predict hospitalizations or mortality. Heywood, et al. (2017) also only looked at data from the CardioMEMS device available in the manufacturer's Merlin.net Patient Care Network (PCN) system and did not combine other risk factors from the patients' health records.

## Health Record Variables and Predictive Ability for Heart Failure

A number of studies have examined the ability to predict hospitalizations and mortality for heart failure patients to determine what factors are most highly correlated with those outcomes. Jing, et al. (2018) found that although patient age and left ventricular ejection fraction were useful for predicting hospitalizations within six month and mortality in one year, the four most important variables were hemoglobin, lymphocytes, NT-proBNP, and creatinine. A systematic review by Calvillo–King, et al. (2013) also found that age was a factor in heart failure readmissions and mortality along with a number of social factors such as low socioeconomic status, living situation, lack of social support, unmarried status, and smoking status demonstrating predictive value. A third study looked at clinical factors for heart failure patients with left ventricular ejection fraction (LVEF) values below 0.35 and found that peak oxygen consumption ($VO_2$), KCCQ symptom stability score, blood urea nitrogen (BUN) region (USA vs. non-USA), LV ejection fraction, and sex were the top six significant predictors for all-cause hospitalizations and mortality (O'Connor, et al., 2010). Zame, Yoon, Asselbergs, & van der Schaar (2018) used interpretable machine learning to identify significant predictors of mortality for heart failure patients, finding that conditions including stroke, rales, shortness of breath at rest, angina, myocardial infarction (MI), oxidation enhanced diffusion (OED), and chronic obstructive pulmonary disease (COPD) are significant risk predictors. Finally, Nichols, Pesa, & Patel (2018) used patient electronic health records to determine significant factors for predicting cardiac-related hospitalizations in the next year. That study found that African American race, current smoker status, obesity, hypertension, coronary artery disease, diabetes, atrial fibrillation, and valvular disorder were the most meaningful predictors.

Together the literature shows that there are a wide variety of factors that yield predictive power when it comes to hospitalizations and mortality in heart failure patients, with some overlap in variables such as age, race, and smoking status. However, none of these studies used data from remote monitoring devices such as the CardioMEMS when making their predictions. Like the heart failure telehealth studies above, these studies also only looked at hospitalizations and mortality while ignoring other adverse outcomes such as PAD threshold changes and IV diuretic use outside of a hospitalization.

This paper will address the gaps in the literature by combining data from CardioMEMS devices with risk factors (some of which were identified as predictive in the studies above) to make projections of adverse events. This paper will expand upon the types of adverse outcomes examined to include PAD threshold changes and use of diuretics outside of a hospitalization, while also looking at both heart failure-related and all-cause hospitalizations.

## Data Collection/Methodology

### Research Question

The purpose of this analysis was to determine if we can we predict the risk for adverse health events for heart failure patients using quantitative data from a combination of CardioMEMS devices and risk factors from the patients' medical records, while balancing accuracy and interpretability of the models. The adverse health events in scope for this analysis included all-cause hospitalizations, heart failure-related hospitalizations, use of IV diuretic therapy outside of a hospitalization, and changes to PAD pressure thresholds.

### Analysis Design

When selecting methods for our predictive models, TUKHS specifically requested to balance the accuracy and interpretability of the results. Our four response variables were all binary, meaning 1 for presence of the adverse event in a patient's history or 0 for absence of the adverse event. This meant that our analysis required the use of classification algorithms. Because we were attempting to model a specific outcome, rather than to find hidden patterns or grouping within the data, our analysis required supervised learning algorithms. A number of algorithms including logistic regression (LR), decision trees (specifically random forests and boosting), and support vector machines (SVM) show up repeatedly in healthcare industry studies using supervised classification.

### ***Supervised Classification Algorithm Performance***

Algorithm performance is key to end user adoption. If an algorithm cannot perform better than random chance, its potential consumers will not see the effort of using the algorithm as worthwhile. A number of studies have shown that random forest (RF) and boosting outperform SVMs in healthcare-related predictions. In predicting 30-day and 180-day all-cause and heart failure-related hospital admissions Mortazavi, et al. (2016) found that RF and boosting algorithms outperformed both LR/SVM and RF/SVM combination models. A study to predict heart failure up to six months prior to diagnosis found that LR and boosting outperformed SVMs (Wu, Roy, & Stewart, 2010). Robinson, Palczewska, Palczewski, & Kidley (2017) found that RF classifiers outperformed SVMs on a number of benchmark datasets.

### ***Supervised Classification Algorithm Interpretability***

In healthcare, and many other highly regulated industries, interpretability can outweigh accuracy in algorithm importance. Interpretability is synonymous with transparency, and algorithms that are interpretable, like decision trees, are essential for user trust (Lee & Shin, 2020). Models labeled as 'black box' that only communicate which patients are at risk without telling the user why, are less actionable than interpretable models (Wiens & Shenoy, 2018).

Interpretable models can be used not just to indicate risks but provide understanding of the causes of adverse events (Jovanovica, Radovanovica, Vukicevica, Pouckeb, & Delibasic, 2016).

Our analysis of the targeted events used LR, RF, and boosting to predict our response variables. We chose LR because it has been used frequently in clinical research to predict binary outcomes due to its interpretability and also has been used as a benchmark against more sophisticated machine learning algorithms (Lorenzoni, et al., 2019). As a supervised classification algorithm, LR models the probability that a given observation falls into one of two response variable levels. In our models this will be the presence (1) or absence (0) of a targeted event for each patient, given the values in their predictor variables. We interpret LR models by checking the coefficient values of the predictor variables to assess if they have a positive or negative influence on the log odds of the response variable. We can also observe the p-value for each predictor variable coefficient to determine its correlation with the response value. A small p-value (<= 0.1) infers that there is a non-zero correlation between the variables.

Our second supervised classification method is Random Forest. RF is an ensemble method, meaning the algorithm creates a number of weaker models, then combines them into a stronger composite model. RF uses an ensemble of decision trees to perform either regression or classification. Each time RF builds a decision tree it only uses a small, random subset of all of the available predictor variables at each split in a tree (James, Witten, Hastie, & Tibshirani, 2013). This approach results in an algorithm that can deal well with imbalanced datasets, outliers, and correlated predictor variables as we have in our dataset. RF models are interpreted by a variable importance metric that is calculated by gauging which variable would lower the overall error the most if it were used as the next decision point in the decision tree. After running thousands of decision trees, the variables that were most often selected, and were selected earliest, are considered the most important.

Boosting is our third and final supervised classification algorithm. Like RF, boosting is an ensemble method that uses decision trees. Boosting builds decision trees sequentially by focusing on lowering the largest errors resulting from previous decision trees (James, Witten, Hastie, & Tibshirani, 2013). While boosting is good at handling unbalanced classes and a large number of predictor variables as we have in our dataset, it can be sensitive to outliers. Like RF, the boosting algorithm computes the importance of each predictor variable for easy interpretability.

We considered using the SVM algorithm due to its popularity in clinical research and its flexibility in applying numerous and various types of predictor variables. However, SVMs are considered complex and difficult to interpret (Martin-Barragan, Lillo, & Romo, 2014; NGUYEN & LE, 2014; Dai, et al., 2015), so we eliminated it from consideration for this analysis.

### *Algorithm Performance Assessment Metric*

We have a few options to assess how well our models are able to predict each of our targeted events. The first metric is accuracy, which is computed by dividing the number correctly classified observations by the total number of observations in our testing dataset. For example, if a model run against test data results in 45 true positives, 35 true negatives, and 100 total observations, the accuracy would be (45 + 35)/100 = 0.8. The higher the accuracy metric, the better the model.

The second option is the area under the curve (AUC) metric. The ‘curve’ in AUC is the receiver operating characteristics (ROC) curve. The ROC curve is a plot of the true positive rate, calculated as true positives/(true positives + false negatives) on the y-axis, against the false positive rate, calculated as the false positives/(false positives + true negative) on the x-axis for a binary classification model. The area under the ROC curve ranges between 0 and 1 and represents how well the model performs when classifying a response variable. The closer the AUC is to 1, the better the model performance against the testing dataset. An AUC of 0.5 implies that the model has no predictive ability and a value of at least 0.7 is considered a minimum acceptable AUC value (Mandrekar, 2010).

Although accuracy is appealing from a conceptual standpoint, research done by Ling & Zhang (2002) showed that when algorithms are optimized for AUC, they produce not only better AUC but also produce better accuracy than models optimized for accuracy. Bradley (1997) illustrated that, when compared to accuracy, AUC has more attractive characteristics and should be favored for model selection.

## Participants

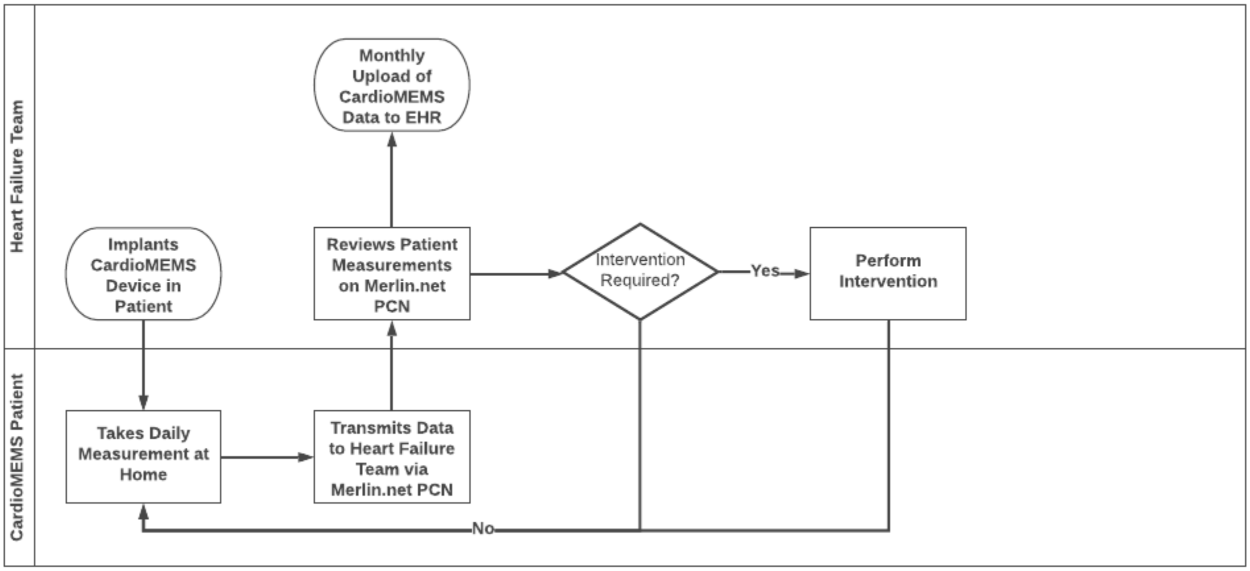
The participants in the cohort were heart failure patients with an implanted CardioMEMS device who were being monitored by the TUKHS Heart Failure Program. The dataset included 88 patients, of which 53% were male. The patients' ages ranged between 43 and 90 years with a median age of 74 years. The patients' length of participation in the program ranged from 1 to 15 months, with a median length of 8 months.

## Data Collection

All patients who has a CardioMEMS device and were being monitored by TUKHS since April 2019 were included in this cohort. Data for the analysis was collected from TUKHS's Epic electronic health record system. The CardioMEMS device data was originally taken from Abbott's Merlin.net PCN system and pasted into Epic once a month by a Heart Failure Clinical Program Coordinator as shown in the workflow depicted in Figure 1.

**Figure 1**

*The CardioMEMS Data Collection Workflow*


Heart Failure Team
CardioMEMS Patient
Monthly Upload of CardioMEMS Data to EHR
Implants CardioMEMS Device in Patient
Reviews Patient Measurements on Merlin.net PCN
Intervention Required?
Yes
Perform Intervention
Takes Daily Measurement at Home
Transmits Data to Heart Failure Team via Merlin.net PCN
No

A TUKHS Data Engineer performed a one-time extract of the analysis data from the Epic Clarity reporting database into a Microsoft SQL Server database. From the SQL Server database, the data was extracted into comma separated value (CSV) file format by the TUKHS data team and protected health information (PHI) was removed from the data before being provided to us for analysis. The variables incorporated in our two datasets, CardioMEMS Device and Risk Factors, are included in the tables in Appendix A.

The scope of the CardioMEMS Device data is for all measurements taken from the device for from the patient's enrollment in the program through July 2, 2020. The scope of the Risk Factors extends backward one year from each patient's first device measurement reading through July 26, 2020.

## Data Analysis

We conducted the analysis was using RStudio software, including a number of software libraries that are common for predictive analytics. RStudio and the R language were requested by the TUKHS team because they are familiar with the toolset and so that they may be able to reuse the code and rerun analysis again in the future. Python and its common data science libraries such as Pandas, NumPy, and scikit-learn were also considered but ruled out due to unfamiliarity with the toolsets by the TUKHS data team.

### ***CardioMEMS Device Dataset***

The CardioMEMS dataset contains 16,503 readings from our patient cohort between April 4, 2019 and July 2, 2020. The distribution of when patients completed device readings is seen in Figure 2, which shows a growing volume of measurements each month through June 2020. There are some residual measurements from early July 2020 when the TUKHS Heart Failure Program Coordinator migrated the June 2020 measurements to Epic. The minimum number of device measurements for a patient was 7 and the maximum was 452, with a median of 177 device readings. Summary statistics for the continuous columns are available in Table 1 where we can see that none of the continuous variables followed a normal gaussian distribution as determined by the Shapiro-Wilk test, all columns had a number of high outliers, and more than half the variables have low outliers. For purposes of this analysis an outlier is defined as a value that is greater than the third quartile value plus 1.5 times the interquartile range (IQR), or less than the first quartile minus 1.5 times the IQR. The upper and lower PAD pressure thresholds both show 216 missing values.

One interesting metric from Table 1 was a minimum PA Diastolic reading of 0. After some investigating this patient had many previous low measures for this variable prior to the observation with the 0, and the following PAD measurement was 1. We determined that this measurement was not a mistake and left it as-is in the dataset.

**Figure 2**

*CardioMEMS Device Measurements by Month*

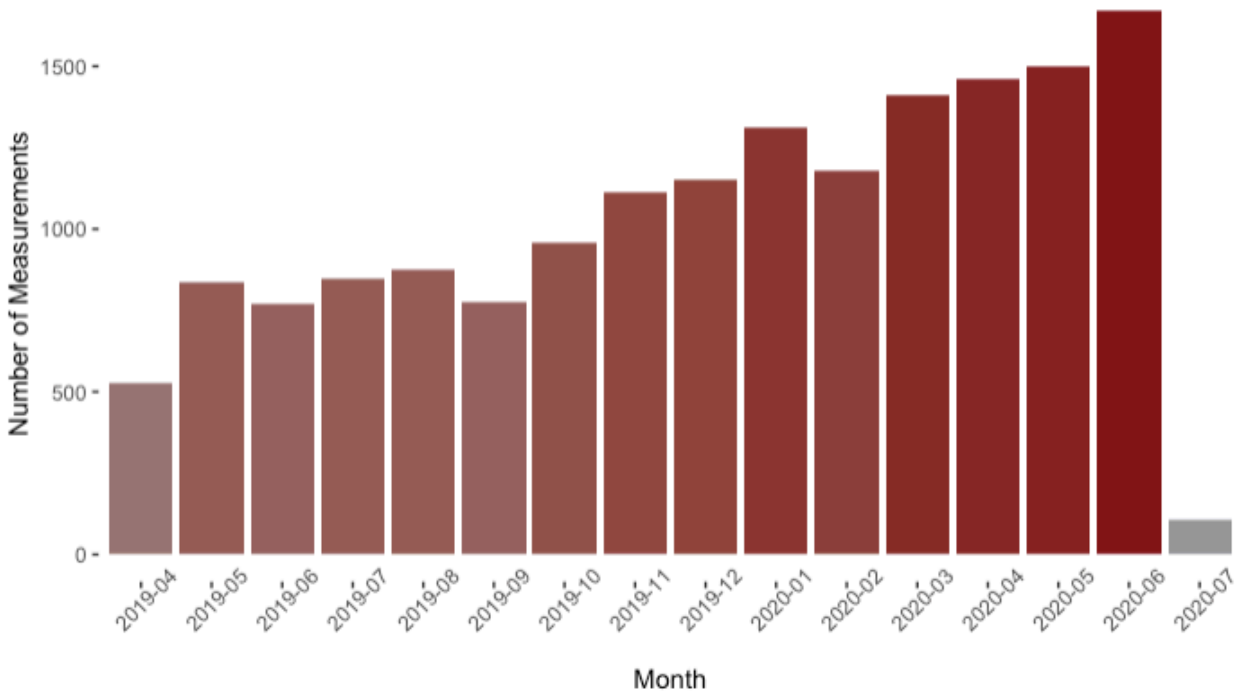


**Table 1**

*CardioMEMS Device Dataset Continuous Variables Summary Statistics*

| Field Name | Max | Min | IQR | Median | Std. Dev. | Num. NA | Normal Dist. | Num. High Out-liers | Num. Low Out-liers |
|---|---|---|---|---|---|---|---|---|---|
| PA Systolic | 111.00 | 11.00 | 15.00 | 38.00 | 13.01 | 0.00 | FALSE | 583.00 | 0.00 |
| PA Diastolic | 69.00 | 0.00 | 8.00 | 17.00 | 6.46 | 0.00 | FALSE | 401.00 | 1.00 |
| PA Pulsatility | 71.00 | 2.00 | 9.00 | 21.00 | 8.83 | 0.00 | FALSE | 699.00 | 11.00 |
| PA Mean | 72.00 | 6.00 | 11.00 | 26.00 | 8.69 | 0.00 | FALSE | 447.00 | 0.00 |
| Heart Rate | 172.00 | 0.00 | 18.00 | 79.00 | 13.26 | 0.00 | FALSE | 162.00 | 5.00 |
| PA Diastolic Threshold Lower | 36.00 | 5.00 | 6.00 | 12.00 | 5.13 | 216.00 | FALSE | 238.00 | 0.00 |

| Field Name | Max | Min | IQR | Median | Std. Dev. | Num. NA | Normal Dist. | Num. High Outliers | Num. Low Outliers |
|---|---|---|---|---|---|---|---|---|---|
| PA Diastolic Threshold Upper | 120.00 | 10.00 | 4.00 | 20.00 | 6.70 | 216.00 | FALSE | 908.00 | 442.00 |

### *CardioMEMS Device Continuous Data Feature Engineering*

To address the missing PAD threshold values, we first performed some exploration for patterns of missing data. We identified 8 patients with missing values. Of those, 6 were missing the threshold values from the beginning of their measurement periods. One patient was missing data from the middle of his/her measurements, and one patient was missing all PAD threshold values. We consulted with the TUKHS team and they agreed that patients missing data from the beginning of their measurements should have missing values replaced with their first measured values. The two other patients should have their missing thresholds replaced with default values of 8 for the lower PAD threshold and 20 for the upper PAD threshold.

One patient also had some outliers in the PA Diastolic Threshold Upper field that we determined to be typos. All of this patient’s other upper threshold values were 20, but some observations had values of 120. After confirming with TUKHS, we set the outlier values to match the patient’s other observations.

The CardioMEMS device dataset has two variables that are dependent on the values in the upper and lower PAD threshold variables, PAD Flag and Threshold Change Indicator. The value in PAD Flag equals 1 if the PAD pressure is outside of the threshold boundaries and 0 otherwise. We recalculated the PAD Flag values for the patients whose threshold values had been imputed using the guidance from the TUKHS team. The recalculation resulted in updates to 103 observations from 0 to 1.

The Threshold Change Indicator variable denotes when a patient's PAD thresholds (upper or lower) have been changed by the clinical team since the previous reading. Our analysis showed that 38 of the 88 patients (43%) had a threshold change during the analysis period. One patient had a series of 15 threshold changes in consecutive device measurements. We consulted with the TUKHS team who agreed that this was noise in the data and that a change of at least 2 mmHg in the PAD thresholds was needed to trigger a positive Threshold Change Indicator. After we made the update the number of patients with PAD threshold changes dropped to 36 (41%) and the total number of PAD threshold changes in the data dropped from 89 to 65.

As was shown in Tran, Wolfson, O'Brien, Yousefian, & Shavelle (2019), patients who transmit their CardioMEMS measurements more frequently experience lower heart failure hospitalization days. After consultation with the TUKHS team we agreed that calculating the number of missed measurements was worthwhile. We completed the calculation by first sorting the device dataset by ID and measurement date. We then subtracted the previous observation's measurement date from each current observation's measurement date. If the difference was greater than 2 then we considered that as missing days. We stored the difference between the dates, minus 1, in a new column named Missed Days. For example, if one measurement were taken on 04-26-2019, and the next measurement for the same patient was taken on 04-29-2019, we calculated that as a gap of three days, and we would store a value of two, representing 04-27-2019 and 04-28-2019, in the new column for the more recent of the two observations. The results showed that the median number of missed days was 28 with a maximum of 313 missed days. All but two of the 88 patients in the cohort had missed days.

The final feature engineering step for the CardioMEMS Device dataset was to address an observation with a heart rate of zero. We consulted with the TUKHS team and determined this was an error in the device reading and this value was not available in any of the patient's records. To ensure a complete dataset we decided to set the heart rate to NA and impute the value using the mice package in RStudio. The mice (multivariate imputation by chained equation) package is an effective algorithm for imputing missing quantitative data in dataset where less than 15% of the dataset is missing values (Putri, Notobroto, & Wibowo, 2018; Ferguson, et al., 2018). We chose to use the classification and regression trees (CART) method of imputing because there was still some level of collinearity in the variables and the CART method works well in this situation. Once we completed the imputation, we stored the imputed value into the CardioMEMS Device dataset.

### *Risk Factors Dataset*

The Risk Factors dataset contained one row for each patient in the cohort (n = 88) with 58 different predictor variables that were identified by the TUKHS team for possible inclusion in the models. Table 2 shows summary statistics for the continuous variables. We see that there were limited high and low outliers, with four variables having missing values, and four variables being normally distributed.

**Table 2**

*CardioMEMS Patient Risk Factors Continuous Variables Summary Statistics*

| Field Name | Max | Min | IQR | Median | Std. Dev. | Num NA | Normal Dist. | Num. High Out-liers | Num. Low Out-liers |
|---|---|---|---|---|---|---|---|---|---|
| Total Hospital Admission Count | 16.00 | 0.00 | 4.00 | 1.00 | 3.39 | 0.00 | FALSE | 3.00 | 0.00 |

| Field Name | Max | Min | IQR | Median | Std. Dev. | Num NA | Normal Dist. | Num. High Out-liers | Num. Low Out-liers |
|---|---|---|---|---|---|---|---|---|---|
| Heart Failure Hospital Admission Count | 9.00 | 0.00 | 1.25 | 0.00 | 1.99 | 0.00 | FALSE | 8.00 | 0.00 |
| Secondary HF Admission Count | 2.00 | 0.00 | 0.00 | 0.00 | 0.44 | 0.00 | FALSE | 10.00 | 0.00 |
| Avg Systolic | 153.00 | 97.00 | 17.75 | 123.00 | 11.79 | 2.00 | TRUE | 0.00 | 0.00 |
| Avg Diastolic | 86.00 | 46.00 | 9.00 | 66.50 | 7.45 | 2.00 | TRUE | 2.00 | 1.00 |
| Ejection Fraction | 65.00 | 20.00 | 30.00 | 55.00 | 16.25 | 27.00 | FALSE | 0.00 | 0.00 |
| Calculated BMI | 61.79 | 19.33 | 11.28 | 34.65 | 8.39 | 2.00 | TRUE | 1.00 | 0.00 |
| Age | 90.00 | 43.00 | 14.00 | 74.00 | 9.61 | 0.00 | TRUE | 0.00 | 1.00 |

### ***Risk Factors Continuous Variable Feature Engineering***

To address the missing values in the risk factors data, we consulted with the team at TUKHS. For Avg Systolic, Avg Diastolic, and Calculated BMI we were advised to drop the continuous variables from the dataset and instead use their associated categorical variables, Average Daily BP Category and BMI Category respectively. To address missing values in those variables we created new category of 'Missing.'

The Ejection Fraction variable had 28 missing values and did not have an associated categorical variable. To address the missing values, we created a new variable, Ejection Fraction Cat. We used established guidelines (Rekha Mankad, 2019) to determine to which category we should transform each observation. An Ejection Fraction value greater than or equal to 55 is considered 'Normal,' under 50 is considered 'Reduced,' and the gap between in considered 'Borderline.' We set the value to 'Missing' for any missing observations and dropped the original, continuous columns from the dataset.

We then added a new continuous variable for length of participation in the CardioMEMS program in months, as calculated by the difference between the date of each patient's first and last measurements in the CardioMEMS device dataset. This variable, Length Participation, was created for use in averaging continuous variables, such as number of missed measurements, by the length of participation in the program.

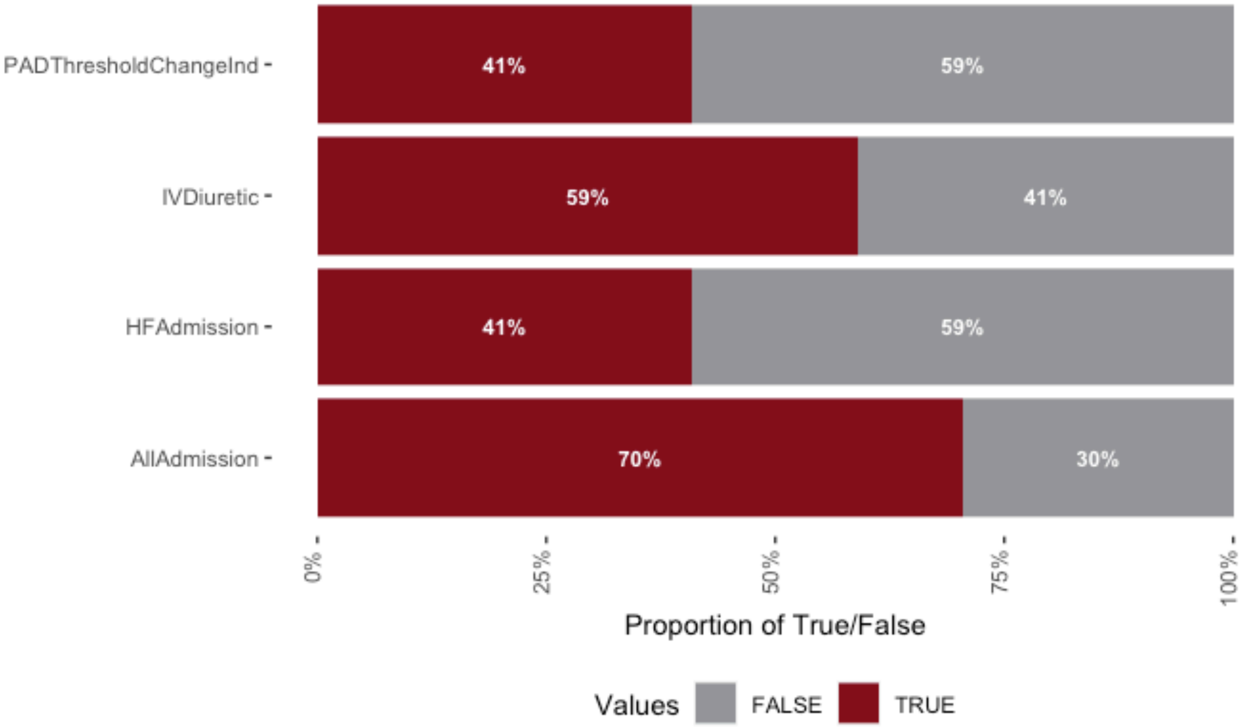
### *Risk Factors Boolean Variables Summary and Feature Engineering*

This dataset included many Boolean columns representing the presence or absence of patient risk factors as identified by the TUKHS Heart Failure Program clinical team, including three of the four response variables for our planned models. All Admission showed that 70% of all patients had a hospital admission during the analysis period and 41% had a heart failure hospitalization during the analysis period as indicated by the HF Admission variable. IV Diuretic therapy in an ambulatory setting was needed by 59% of patients during the analysis period.

Our risk factors dataset was missing the fourth and final response variable of PAD threshold changes. We were able to calculate this value for each patient by referencing the CardioMEMS device dataset and summing the Threshold Change Indicator values by patient. If a patient had a total greater than zero, we set the PAD Threshold Change Ind to 1, otherwise the variable was 0. The result was 41% of patients with a PAD threshold change. We added the new column to the risk factors dataset. The four response variables proportions are seen in Figure 3.

**Figure 3**

*Response Variable Proportions*



The only Boolean variable in the Risk Factors dataset with missing values was BP Uncontrolled. The patients who were missing this value were also missing Avg Systolic, Avg Diastolic, and Avg Daily BP Category. After discussion with TUKHS we decided to drop BP Uncontrolled from the dataset because there was already a blood pressure-related variable, Average Daily BP Category. We added a new category value of 'Missing' for any NA values in this variable.

The Risk Factors dataset had two columns for diabetes representing complicated and uncomplicated diagnoses. Exploration of these variables revealed that 10 patients had positive indicators in both. Because we were not able to distinguish which condition was most recent, we combined the variables into a single Diabetes variable and dropped the originals.

In Table 3 we see the proportions of True/Yes and False/No for all of the Boolean predictor variables in the Risk Factors dataset, sorted by higher proportion of True/Yes. We see that 74 of the 88 patients enrolled in the program were white. Unsurprisingly many of the CardioMEMS program patients have risk factors associated with cardiovascular disease including hypertension, congestive heart failure, coronary artery disease, and high cholesterol. Just over half have diabetes. The dataset also shows some of the risk factors are sparse, such as specific diabetes medications DPPV4 and SGLT2 associated with fewer than 10 patients each. Alcohol abuse and drug abuse affect only one patient each in the dataset.

**Table 3**

*CardioMEMS Patient Risk Factors Boolean Variable Proportions*

| Field Name | Percent True | Percent False |
|---|---:|---:|
| RaceWhite | 84.09% | 15.91% |
| Hypertension | 82.95% | 17.05% |
| CongestiveHF | 68.18% | 31.82% |
| CoronaryArteryDisease | 62.50% | 37.50% |
| HighCholesterol | 59.09% | 40.91% |
| Insulin | 57.95% | 42.05% |
| Diabetes | 54.55% | 45.45% |
| Male | 53.41% | 46.59% |
| CPAPUsage | 45.45% | 54.55% |
| ChronicPulmonary | 45.45% | 54.55% |
| Obesity | 39.77% | 60.23% |
| PulmonaryCirculation | 27.27% | 72.73% |
| Hypothyroidism | 27.27% | 72.73% |
| Depression | 27.27% | 72.73% |
| DeficiencyAnemia | 21.59% | 78.41% |
| PeripheralVascular | 19.32% | 80.68% |
| ValvularHeartDisease | 18.18% | 81.82% |
| GLP1 | 15.91% | 84.09% |
| Cancer | 15.91% | 84.09% |
| Metformin | 14.77% | 85.23% |
| Tumor | 12.50% | 87.50% |
| Arthritis | 11.36% | 88.64% |
| RenalFailure | 10.23% | 89.77% |
| LiverDisease | 9.09% | 90.91% |
| DPPV4 | 7.95% | 92.05% |

| Field Name | Percent True | Percent False |
|---|---:|---:|
| NeurologicalDisorder | 7.95% | 92.05% |
| PepticUlcer | 7.95% | 92.05% |
| SGLT2 | 6.82% | 93.18% |
| SleepApnea | 5.68% | 94.32% |
| ElectrolyteDisorder | 5.68% | 94.32% |
| Psychoses | 3.41% | 96.59% |
| Lymphoma | 2.27% | 97.73% |
| Coagulopathy | 2.27% | 97.73% |
| AlcoholAbuse | 1.14% | 98.86% |
| DrugAbuse | 1.14% | 98.86% |

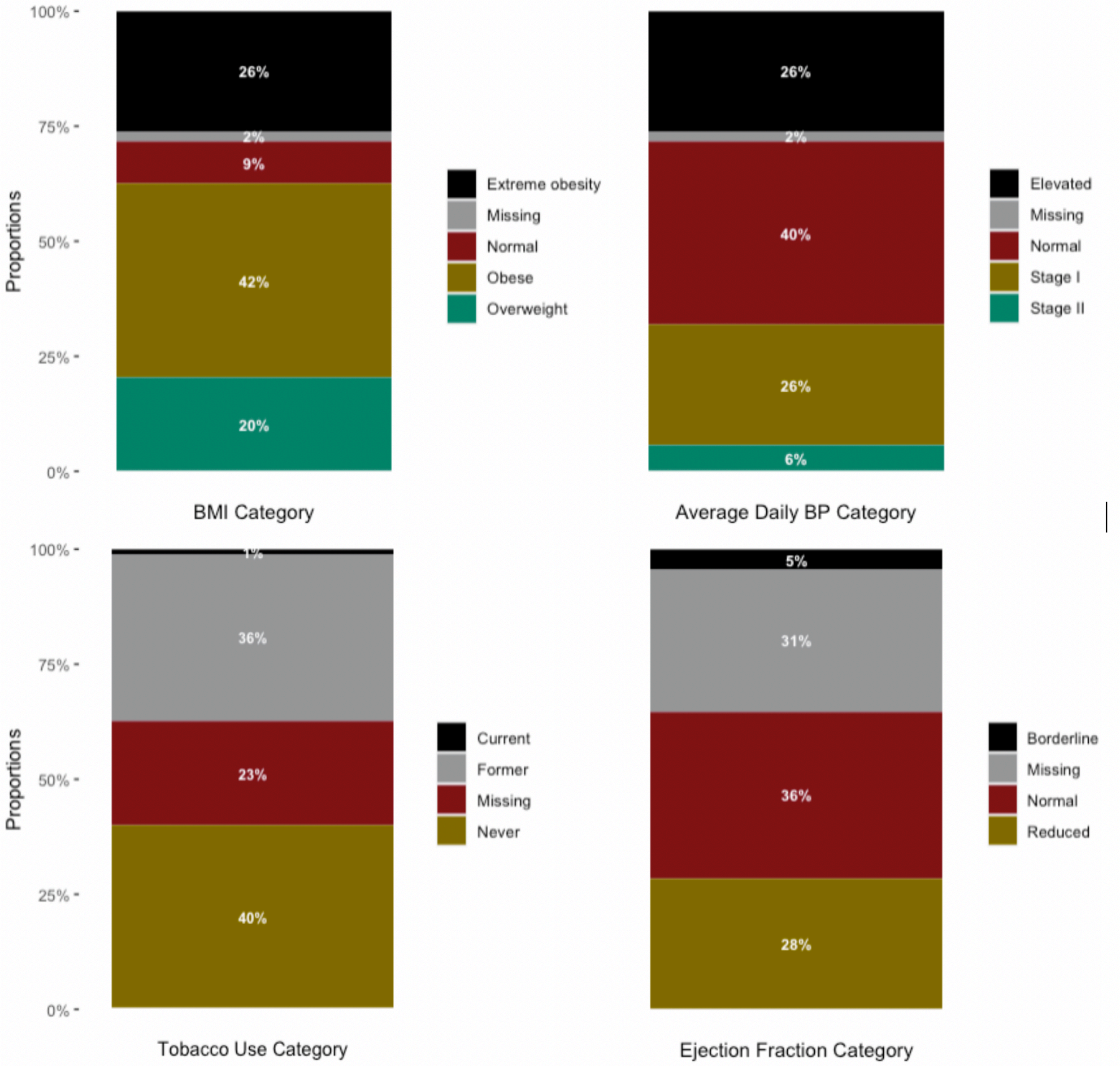
### ***Risk Factors Categorical Variables Summary and Feature Engineering***

The Risk Factors dataset has four categorical variables. Three of the variables, Average Daily BP Category, BMI Category, and Tobacco Use were original to the dataset and we added Ejection Fraction Cat as a replacement for a continuous variable. The three original variables all had some missing observations, which, after consulting with the TUKHS team, we decided to replace with a value of ‘Missing.’

After addressing missing values, we explored the proportions of the categorical variables, as seen in Figure 4. For BMI, only 9% of patients had a normal BMI with 2% missing values. The remaining 88% of patients overweight, obese, or had extreme obesity (BMI >= 40). For average daily blood pressure, 39% of patients had normal averages, 11% were missing data, and the remaining 50% had high average pressures. For tobacco use, 1% were current users, 36% were former users, 40% never used, with 23% missing data. For ejection fraction, 36% had normal function, 27% had reduced function, 5% had borderline function, and 32% were missing data.

**Figure 4**

*Risk Factors Categorical Variable Proportions*



## Combining CardioMEMS and Risk Factors Data

Prior to creating our predictive models, we needed to combine the data from the two datasets. We considered two approaches for the task. The first approach considered was to merge the Risk Factors data into the CardioMEMS Device data, which would multiply each patient's risk factors across each device reading. The authors and the TUKHS team were concerned about what effect that would have on the model results given that some patients had as few as 7 device readings while others had many hundreds of readings. This is due to the fact that the patients' lengths of participation in the program during our analysis window varied between 15 months and 1 month. I also consulted with J. Kraker, an Associate Professor of data mining and machine learning at the University of Wisconsin – Eau Claire, who advised that if each patient had the same length of participation in the CardioMEMS program, this approach would be more feasible for our research questions (personal communication, July 23, 2020).

The second approach we considered was to aggregate the CardioMEMS Device data by patient and then merge the results with the Risk Factors data. This would eliminate bias towards the risk factors for patients who had been participating in the CardioMEMS program for longer than other patients. The trade-off was the loss of some detail in the device time series information that might have had predictive ability. Given that the response variables were at the patient level, it was agreed to by the authors, TUKHS, and Professor J. Kraker (personal communication, July 23, 2020), that this approach was the better option.

Using the selected approach, we created a plan for each of the predictor variables in the CardioMEMS Device dataset, which were summarized in Table 4. For the quantitative measurements that came from the CardioMEMS device we decided that, for each patient we would measure the maximum value, minimum value, average (mean) value, and the difference between the first measurement and last measurement to get a sense of the direction and magnitude of change in values. For indicator (Boolean) variables that were derived from CardioMEMS device data we summed each patient's values and then divided (scaled) each total by the patient's length of participation in the program to reduce bias against patients with longer participation lengths. This resulted in 27 new continuous variables for the risk factors dataset.

**Table 4**

*CardioMEMS Device Data Aggregation Summary*

| Field Name | Min | Max | Mean | Difference (First-Last) | Summed and Scaled by Length of Participation |
|---|---|---|---|---|---|
| PA Systolic | ✓ | ✓ | ✓ | ✓ | |
| PA Diastolic | ✓ | ✓ | ✓ | ✓ | |
| PA Pulsatility | ✓ | ✓ | ✓ | ✓ | |
| PA Mean | ✓ | ✓ | ✓ | ✓ | |
| Heart Rate | ✓ | ✓ | ✓ | ✓ | |
| PAD Threshold Range | ✓ | ✓ | ✓ | ✓ | |
| PAD Flag | | | | | ✓ |
| Threshold Change Indicator | | | | | ✓ |
| Missed Days | | | | | ✓ |

### ***Combined Dataset***

Summary statistics for the new columns derived from the CardioMEMS Device dataset are available in Table 5. Only two of the columns, Heart Rate Min and Heart Rate Mean were normally distributed. Outliers were somewhat limited when compared to the number of patients in the program, with PAD Threshold Range Diff. (between first and last device readings) having the most outliers with 12 high and 13 low observations.

The 'Diff' columns also revealed some interesting information by showing the first measured value minus the last measured value for each patient. Over the course of enrollment in the program the median patient Pulsatility and PAD Threshold Range values stayed the same. The PA Systolic, PA Diastolic, and PA Mean median values each decreased by 1 mmHg. The median patient heartbeat increased by 1.5 beats per minute. One final interesting statistic is that the median number of missed measurements per month was 4.29 with a maximum of 24.5 missed days in a month.

**Table 5**

*New Continuous Variables Summary Statistics*

| Field Name | Max | Min | IQR | Median | Std. Dev. | Normal Dist. | Num. High Outliers | Num. Low Outliers |
|---|---|---|---|---|---|---|---|---|
| PA Systolic Min | 80.00 | 11.00 | 11.25 | 28.00 | 11.61 | FALSE | 4.00 | 0.00 |
| PA Systolic Max | 111.00 | 25.00 | 19.50 | 54.00 | 16.36 | FALSE | 4.00 | 0.00 |
| PA Systolic Mean | 93.81 | 18.38 | 14.47 | 38.22 | 13.44 | FALSE | 6.00 | 0.00 |
| PA Diastolic Min | 33.00 | 0.00 | 6.50 | 12.00 | 6.06 | FALSE | 3.00 | 0.00 |
| PA Diastolic Max | 69.00 | 11.00 | 11.25 | 26.00 | 9.69 | FALSE | 3.00 | 0.00 |
| PA Diastolic Mean | 39.19 | 6.92 | 6.62 | 17.81 | 6.23 | FALSE | 4.00 | 0.00 |
| PA Pulsatility Min | 47.00 | 2.00 | 7.00 | 12.00 | 8.11 | FALSE | 5.00 | 0.00 |
| PA Pulsatility Max | 71.00 | 13.00 | 13.50 | 28.50 | 11.12 | FALSE | 2.00 | 0.00 |
| PA Pulsatility Mean | 54.62 | 7.66 | 8.95 | 20.18 | 9.04 | FALSE | 5.00 | 0.00 |
| PA Mean Min | 52.00 | 6.00 | 8.25 | 19.00 | 7.89 | FALSE | 4.00 | 0.00 |
| PA Mean Max | 72.00 | 17.00 | 12.25 | 37.00 | 11.52 | FALSE | 4.00 | 0.00 |
| PA Mean Mean | 60.94 | 13.33 | 8.67 | 26.70 | 8.70 | FALSE | 5.00 | 0.00 |
| Heart Rate Min | 87.00 | 39.00 | 13.25 | 65.50 | 10.91 | TRUE | 0.00 | 1.00 |
| Heart Rate Max | 172.00 | 71.00 | 21.25 | 104.00 | 17.89 | FALSE | 2.00 | 0.00 |
| Heart Rate Mean | 112.01 | 59.54 | 13.95 | 79.95 | 10.97 | TRUE | 1.00 | 0.00 |
| PAD Threshold Range Min | 12.00 | 0.00 | 3.00 | 6.00 | 1.86 | FALSE | 0.00 | 1.00 |
| PAD Threshold Range Max | 16.00 | 4.00 | 2.00 | 7.00 | 1.86 | FALSE | 2.00 | 0.00 |
| PAD Threshold Range Mean | 12.00 | 3.71 | 2.27 | 6.66 | 1.59 | FALSE | 1.00 | 0.00 |
| PA Systolic Diff | 37.00 | -26.00 | 8.25 | 1.00 | 9.13 | FALSE | 1.00 | 4.00 |
| PA Diastolic Diff | 15.00 | -14.00 | 4.25 | 1.00 | 5.77 | FALSE | 5.00 | 8.00 |
| PA Pulsatility Diff | 22.00 | -16.00 | 6.00 | 0.00 | 4.92 | FALSE | 1.00 | 1.00 |
| PA Mean Diff | 24.00 | -17.00 | 6.00 | 1.00 | 7.07 | FALSE | 3.00 | 5.00 |
| Heart Rate Diff | 53.00 | -38.00 | 16.00 | -1.50 | 13.85 | FALSE | 2.00 | 1.00 |
| PAD Threshold Range Diff | 9.00 | -4.00 | 0.00 | 0.00 | 1.48 | FALSE | 12.00 | 13.00 |
| PAD Flag Scaled | 13.23 | 0.00 | 0.00 | 0.00 | 2.78 | FALSE | 9.00 | 0.00 |
| ThresholdChangeInd Scaled | 0.33 | 0.00 | 0.14 | 0.00 | 0.11 | FALSE | 0.00 | 0.00 |
| MissedDays Scaled | 24.50 | 0.00 | 6.63 | 4.29 | 5.66 | FALSE | 4.00 | 0.00 |

## Modeling Preparations

We would like to describe some of the steps taken to prepare the consolidated dataset prior to running each of the different models, regardless of which model we ran. Our first step was to load the full dataset and then drop any variables that had either too much, or no, predictive value. For example, the ID variable was a random number that uniquely identified each patient in the dataset and had no predictive value. Other columns may have had too much predictive value because they were derived from our response variable or were directly correlated with the response variable. In the case of all-cause hospital admissions, the HF Admission variable was a subset of the response variable, and the Total Hospital Admission Count was directly correlated because it counted the number of times the patient was admitted to a hospital.

For our first model, all-cause hospitalizations, we removed the HF Admission, Total Hospital Admission Count, Heart Failure Hospital Admission Count, Secondary HF Admission Count, and ID variables. We removed the Heart Failure Hospital Admission Count, All Admission, Total Hospital Admission Count, Secondary HF Admission Count, and ID variables from our second model, heart failure hospitalizations. The only variable removed from model 3, IV diuretic therapy was ID. Finally, we removed the Threshold Change Ind Scaled, PAD Threshold Range Diff, and ID variables from model 4, PAD threshold changes.

The next step in preparing the data was to set all of the categorical and Boolean variables as factors because that was required by some of the predictive algorithms we used. Setting a variable as a factor told the algorithm that data was limited to a certain number of discrete values, such as {0,1} in the case of the Boolean variables. We then changed the response variable in each model to use levels (0 = ‘No’, 1 = ‘Yes’). This was required by the R caret package to calculate the performance metrics for each of our models.

We then used the caret package to divide the data into training and testing datasets with 70% of the data used for training the algorithm and 30% of the data used for testing the algorithms’ predictive ability using new data. Because we used LR as one of our algorithms, we created a copy of the training data and then scaled it according to the response variable. Scaling the dataset made the number of observations for each response value (‘Yes’/‘No’) more balanced. For example, in the all-cause hospitalization model the training dataset had 44 observations with the response variable equal to ‘Yes’ (patient experienced a hospitalization) and only 19 observations with the response variable equal to ‘No’ (patient was not hospitalized). That was a ratio of 2.3 times as many records in favor of a hospitalization. The scaled dataset had a ratio of 1.25 patients with hospitalization to patients without hospitalizations. Had we left the data unbalanced in favor of patients with hospitalizations that could have led LR to downplay the value of the data for patients without hospitalizations (Ramyachitra & Manikandan, 2014).

## Modeling Approach

In order to determine which variables to include in our logistic models we then checked each predictor variable against the response variable to gauge its predictive power. For our factor variables (Boolean and categorical) we used the chi-squared test of independence, which determined if two categorical variables, one being the response variable and the other being a predictor variable, are associated or not associated. For these tests we used a 0.1 level of significance to choose the predictor variables that were associated with the response variable. For the continuous predictor variables, we used the Wilcoxon-Mann Whitney test to determine the association with the response variables, again with a significance level of 0.1 as a basis for inclusion in the models.

When training each of our predictive models we used five-fold cross-validation. This method divided the training data into five equally sized random groups/folds, then performed five iterative fits leaving one fold out each time and fitting on the four remaining folds. The excluded fold was then used as a validation set to measure the error rate. Five-fold cross validation is a type of k-fold cross-validation. We used this method in all of our models because it helped estimate variable coefficients that balance reduced error rates without overfitting the model to our training data (James, Witten, Hastie, & Tibshirani, 2013). We chose five folds due to the size of our training datasets. With 88 observations in our full dataset a 70/30 split of our data resulted in 63 observations in our training datasets. We determined that folding 63 observations ten ways would results in only six or seven observations in each validation set during the cross-validation process. Dividing the 63 training observations into five folds would result in 12 or 13 validation records in each fold.

Once we had identified the predictor variables most associated with our response variables, we trained an initial LR using the scaled training dataset with all associated predictor variables and gauged the models' performances against the test data. We then used the initial model to determine if a more parsimonious model was available with equal or better performance than the initial LR model. Including too many predictor variables in a model may result in overfitting the model to the training data with insignificant predictor variables classified as significant (Portet, 2020). Our chosen method was backwards stepwise variable elimination. This method eliminated one predictor variable at a time from the initial model and measured each new model's Akaike Information Criterion (AIC) value. AIC is a measure of information loss from a statistical model and it favors models with fewer predictor variables. Lower AIC values are better. The stepwise algorithm determined the best combination of predictor variables once it could no longer improve the AIC value by removing additional variables. We then trained a second LR model using the predictor variables identified by the stepwise analysis and measured its performance against the testing dataset.

The second algorithm we ran in each of our models was the RF using five-fold cross-validation. To tune our RF models, we varied the number of randomly selected predictor variables used at each split, known as mtry, to find the number of predictors that produced the best model accuracy. Because the RF algorithm had very few assumptions and it was able to deal with dataset issues such as multicollinearity (James, Witten, Hastie, & Tibshirani, 2013), we kept all predictor variables in the models and used the unscaled training datasets. Once the models were trained, we compared them against the testing dataset and evaluated their performances.

The third and final algorithm we used was boosting with five-fold cross-validation. Boosting had a few more tuning parameters for us to select than RF. The first was the number of trees that would be fit. We chose between 500 and 6,000 trees in increments of 500. We also evaluated three shrinkage parameters of 0.001, 0.01, and 0.1. The final tuning parameter was the interaction depth. We chose integer values between 1 and 5. One final parameter we used with boosting was the minimum number of observations used per node, where we chose the commonly used value of 10. Boosting has been shown to do well using unbalanced classes and using all predictive variables as well as with scaled training data a selected predictor variables (Hasanin, Khoshgoftaar1, Leevy, & Seliya, 2019). We attempted four different versions of boosting. The first model used all predictors and the unscaled training dataset. The second model used all predictors and the scaled training dataset. The third model used selected predictors matching the initial LR model predictors with the unscaled training dataset. The fourth model used selected predictors and the scaled training dataset. As with LR and RF, the trained models were tested against the testing dataset to determine their predictive ability.

### Ethical Considerations

The University of Kansas Health System (TUKHS) and the University of Wisconsin – La Crosse jointly completed a non-clinical Affiliation Agreement ("the agreement") to authorize this analysis, which was completed in partnership with TUKHS Health Information Technology Services in accordance with the agreement.

### Assumptions, Delimitations, and Limitations

We assume that the data included in the analysis, including transformations made to the data during the extraction process, are accurate to the specifications requested by the clinical team at TUKHS.

The analysis was undertaken within a timeline defined by the establishment of the Affiliation Agreement between the University of Kansas Health System and the University of Wisconsin – La Crosse and the end of the Summer 2020 University of Wisconsin Master's in Data Science Capstone semester.

## Results

### Model 1 – All-cause Hospitalizations

This model aimed to predict CardioMEMS patients who are at risk for being admitted to the hospital for any reason. When we tested each of the predictor variables against the response variable using the chi-squared test or Wilcoxon-Mann Whitney test the following predictor variables were identified as associated with the response variable: Missed Days Scaled, PAD

Threshold Range Max, Congestive HF, CPAP Usage, IV Diuretic, Insulin, Diabetes, Ejection Fraction Cat, and Tobacco Use.

### *Logistic Regression*

The initial logistic regression against the scaled training dataset and using the predictor variables identified above had an AUC value of 0.913. By examining the coefficient estimates in Table 6 we see that increases/positives in PAD Threshold Range Max, Missed Days Scaled, Diabetes, Tobacco User Former, and Tobacco User Missing all increase the odds of all-cause hospitalization, while the remaining coefficient estimates decrease the odds. Although, as we can see from the p-values, none of the predictor variables' coefficients were significantly different from zero.

**Table 6**

*All-cause Hospitalizations Initial Logistic Regression Results*

| Term | Estimate | Std. Error | Statistic | p-value |
|---|---|---|---|---|
| (Intercept) | -5.9703 | 6.1224 | -0.9752 | 0.3295 |
| PAD_Threshold_Range_Max | 0.7011 | 0.7305 | 0.9598 | 0.3372 |
| MissedDays_Scaled | 0.1596 | 0.2677 | 0.5963 | 0.5510 |
| EjectionFractionCatNormal | -1.4435 | 3.5332 | -0.4085 | 0.6829 |
| Diabetes1 | 2.4796 | 7.1591 | 0.3464 | 0.7291 |
| CPAPUsage1 | -0.5437 | 2.6588 | -0.2045 | 0.8380 |
| TobaccoUseFormer | 0.9997 | 6.2201 | 0.1607 | 0.8723 |
| CongestiveHF1 | -0.4210 | 4.4143 | -0.0954 | 0.9240 |
| EjectionFractionCatMissing | -0.6747 | 7.8993 | -0.0854 | 0.9319 |
| Insulin1 | -0.2503 | 9.3615 | -0.0267 | 0.9787 |
| TobaccoUseMissing | 23.7574 | 5,889.4293 | 0.0040 | 0.9968 |
| IVDiuretic1 | -22.3704 | 5,896.7745 | -0.0038 | 0.9970 |
| EjectionFractionCatReduced | -2.2209 | 7,660.0542 | -0.0003 | 0.9998 |
| TobaccoUseNever | NA | NA | NA | NA |

When we used the backwards stepwise variable selection, the result was a model with an AIC of 27.12, down from 42.62, and only two variables to include in the next model. The more parsimonious model with an AUC of 0.869 did not outperform the original model. The coefficient estimates as seen in Table 7 did not change significantly from the initial LR model. Again, in this model none of the predictor variable coefficient p-values were significantly different from zero.

**Table 7**

*All-cause Hospitalizations Parsimonious Logistic Regression Results*

| Term | Estimate | Std. Error | Statistic | p-value |
|---|---|---|---|---|
| TobaccoUseFormer | 0.5754 | 1.0801 | 0.5327 | 0.5943 |
| (Intercept) | -0.2877 | 0.7638 | -0.3767 | 0.7064 |
| IVDiuretic1 | -21.5547 | 5,492.9730 | -0.0039 | 0.9969 |
| TobaccoUseMissing | 21.8538 | 6,379.0410 | 0.0034 | 0.9973 |
| TobaccoUseNever | NA | NA | NA | NA |

### ***Random Forest***

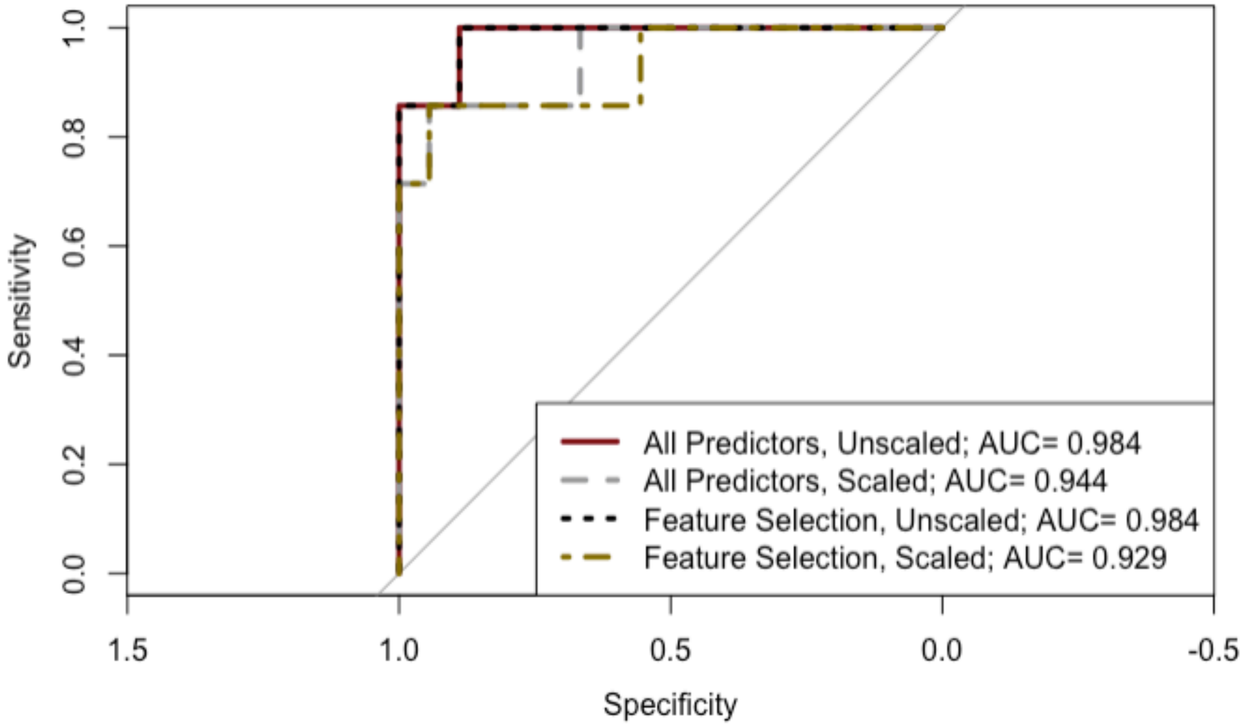
Our RF algorithm with all predictors and unscaled training dataset had an AUC of 0.925, which outperformed both the initial and parsimonious LR. The best performing model from our tuning parameters had 8 predictor variables per split. The top five most important predictor variables for the RF model and their importance scores were Tobacco Use Missing (3.670), IV Diuretic 1 (3.226), Insulin (1.193), CPAP Usage (1.132), and Heart Rate Max (0.708).

### ***Boosting***

All of the boosting models outperformed the RF model and both LR models. As Figure 5 shows, the order of best to worst performing models were a tie between all predictors with unscaled training dataset and selected predictors with unscaled training dataset (AUC = 0.984), followed by all predictor variables with scaled training dataset (AUC = 0.944), and then selected predictors with scaled training dataset (AUC = 0.929).

**Figure 5**

*All-cause Hospitalizations Boosting AUC Comparison*



Two boosting models tied for best performance. The model with all predictors and unscaled training data' top five most important variables and their relative influences were IV Diuretic (79.053), Tobacco Use Missing (9.042), Insulin (5.548), CPAP Usage (4.272), and Missed Days Scaled (0.339). This model's parameters were number of trees = 500, interaction depth = 5, and shrinkage = 0.001.

For the model with selected features and unscaled training data the top five most important variables and their relative influences were IV Diuretic (55.361), Tobacco Use Missing (43.118), PAD Threshold Range Max (0.555), Missed Days Scaled (0.415), and Insulin (0.2). This model's parameters were number of trees = 500, interaction depth = 1, and shrinkage = 0.001.

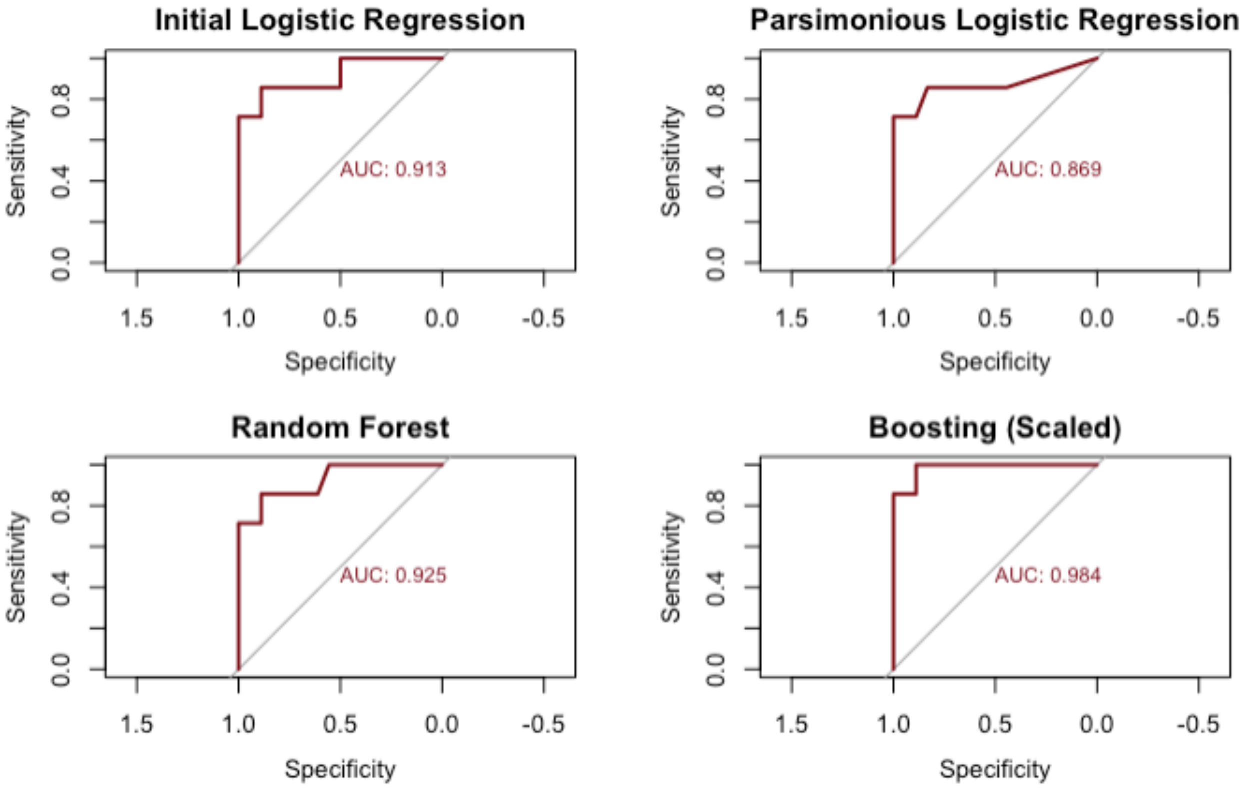
### *All-cause Hospitalizations Discussion*

As Figure 6 shows, both boosting methods with unscaled training datasets were the best performing models for predicting all-cause hospitalizations. When we compare the top ten most important variables across the best performing models for each algorithm there is a fair amount of overlap for predicting all-cause hospitalizations. IV Diuretic, Tobacco Use Missing, CPAP Usage, and Insulin factored into all models, typically at the top of the list of each algorithm's predictor variables. It is also interesting that IV Diuretic, one of our targeted adverse events, was a top predictor of all-cause hospitalizations across algorithms. It's negative coefficient estimate in the LR models suggests that a patient using IV diuretics reduces the odds of a hospitalization for any cause, though neither of our LR models had any predictor variable coefficients that were significantly different from zero, implying that those variables are not good predictors of the response variable value. We had tried to avoid this issue by scaling the training dataset to balance the observations between patients with and without all-cause hospitalizations. Research has shown that in generalized linear models such as LR, the p-value has fundamental shortcomings, especially on smaller datasets as in our CardioMEMS dataset (VakhitovaI & Alston-Knox, 2018). This may explain why we saw such high p-values in our LR models.

Given the generally high AUC values for all models when compared to established guidelines (Mandrekar, 2010), LR (initial), RF, and Boosting are all considered outstanding predictors of all-cause hospitalizations. Boosting using predictors selected by chi-squared or Wilcoxon-Mann Whitney tests for association and unscaled data had a nearly perfect AUC of 0.984.

**Figure 6**

*All-cause Hospitalizations Predictive Models AUC Comparison*



**Model 2 – Heart Failure Hospitalizations**

This model sought to predict patients who are at risk for being admitted to the hospital for a heart failure-related reason as the primary diagnosis. When we tested each of the predictor variables against the response variable using the chi-squared test or Wilcoxon-Mann Whitney test the following predictor variables were identified as associated with the response variable: PA Diastolic Max, PAD Flag Scaled, PA Mean Max, PA Diastolic Mean, Heart Rate Mean, IV Diuretic, Insulin, Congestive HF, CPAP Usage, Metformin, and Tobacco Use.

***Logistic Regression***

The initial LR against the scaled training dataset and using the predictor variables identified above had an AUC value of 0.813. By examining the coefficient estimates in Table 8 we see that increases/positives in IV Diuretic, Congestive HF, Tobacco Use Former, PAD Flag Scaled, Tobacco Use Never, and Insulin all increase the odds of heart failure hospitalization while the remaining coefficient estimates decrease the odds. Although, as we can see from the p-values none of the predictor variables' coefficients were significantly different from zero.

**Table 8**

*Heart Failure Hospitalizations Initial Logistic Regression Results*

| Term | Estimate | Std. Error | Statistic | p-value |
|---|---:|---:|---:|---:|
| IVDiuretic1 | 75.0723 | 134,997.3799 | 0.0006 | 0.9996 |
| CongestiveHF1 | 38.4832 | 121,960.5885 | 0.0003 | 0.9997 |
| Metformin1 | -60.5424 | 225,750.7362 | -0.0003 | 0.9998 |
| (Intercept) | -110.2292 | 590,897.6398 | -0.0002 | 0.9999 |
| PA_Diastolic_Mean | -1.0567 | 6,899.6803 | -0.0002 | 0.9999 |
| PA_Diastolic_Max | 1.8387 | 15,266.7515 | 0.0001 | 0.9999 |
| Heart_Rate_Mean | -0.2667 | 3,961.5284 | -0.0001 | 0.9999 |
| TobaccoUseFormer | 71.3837 | 1,147,295.5379 | 0.0001 | 1.0000 |
| TobaccoUseMissing | -48.4055 | 969,162.8342 | -0.0000 | 1.0000 |
| CPAPUsage1 | -17.1776 | 442,104.3006 | -0.0000 | 1.0000 |
| PAD_Flag_Scaled | 1.5086 | 47,726.5280 | 0.0000 | 1.0000 |
| TobaccoUseNever | 29.4496 | 979,280.3030 | 0.0000 | 1.0000 |
| PA_Mean_Max | -0.0080 | 3,669.4837 | -0.0000 | 1.0000 |
| Insulin1 | 0.2027 | 393,422.8315 | 0.0000 | 1.0000 |

When we used the backwards stepwise variable selection, the result was a model with an AIC of 14, down from the initial AIC of 28, and four variables to include in the next model. The more parsimonious model with an AUC of 0.850 outperformed the original LR model. The coefficient estimates in Table 9 changed in magnitude from the initial LR with Tobacco Use Never switching from negative to positive. Again, none of the predictor variable coefficient p-values were significantly different from zero in this model.

**Table 9**

*Heart Failure Hospitalizations Parsimonious Logistic Regression Results*

| Term | Estimate | Std. Error | Statistic | p-value |
|---|---:|---:|---:|---:|
| IVDiuretic1 | 134.0270 | 77,814.5009 | 0.0017 | 0.9986 |
| Metformin1 | -89.0332 | 54,513.5511 | -0.0016 | 0.9987 |
| CongestiveHF1 | 44.1913 | 31,347.2018 | 0.0014 | 0.9989 |
| Insulin1 | 43.9990 | 34,828.4983 | 0.0013 | 0.9990 |
| (Intercept) | -151.6523 | 369,637.7271 | -0.0004 | 0.9997 |
| TobaccoUseMissing | -94.1206 | 362,162.5401 | -0.0003 | 0.9998 |
| TobaccoUseFormer | 40.3663 | 359,479.9730 | 0.0001 | 0.9999 |
| TobaccoUseNever | -4.9977 | 357,063.1210 | -0.0000 | 1.0000 |

### *Random Forest*

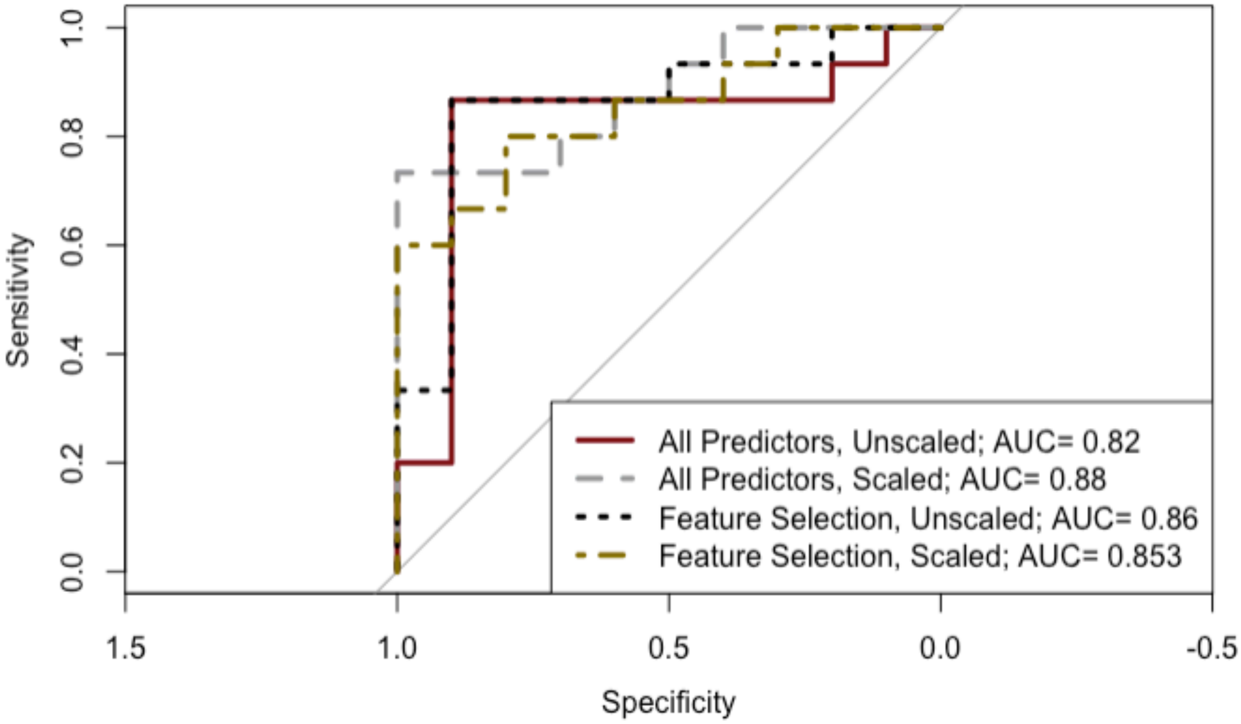
Our RF algorithm with all predictors and unscaled training dataset had an AUC of 0.83, which outperformed the initial LR model but could not match the parsimonious LR model. The best performing model from our tuning parameters had 34 predictor variables per split. The top five most important predictor variables and their importance scores were IV Diuretic (6.161), Tobacco Use Missing (1.978), Heart Rate Diff (1.78), Congestive HF (1.139), and Heart Rate Mean (1.033).

### *Boosting*

Three of the boosting models beat both LR models and the RF model for heart failure hospitalizations. The boosting model with all predictor variables and an unscaled dataset (AUC = 0.820) finished next-to-last between the initial LR and RF. As Figure 7 shows, the best performing boosting models were all predictors and scaled training dataset (AUC = 0.880), select predictors and unscaled training dataset (AUC = 0.860), and selected features and scaled training dataset (AUC = 0.853).

**Figure 7**

*Heart Failure Hospitalizations Boosting AUC Comparison*



For the best performing boosting model, the top five most important variables and their relative influences were IV Diuretic (24.506), Congestive HF (8.475), Heart Rate Diff (8.289), PA Systolic Max (8.058), PA Pulsatility Diff (5.922). This model's parameters were number of trees = 1,000, interaction depth = 1, and shrinkage = 0.01.

### *Heart Failure Hospitalizations Discussion*

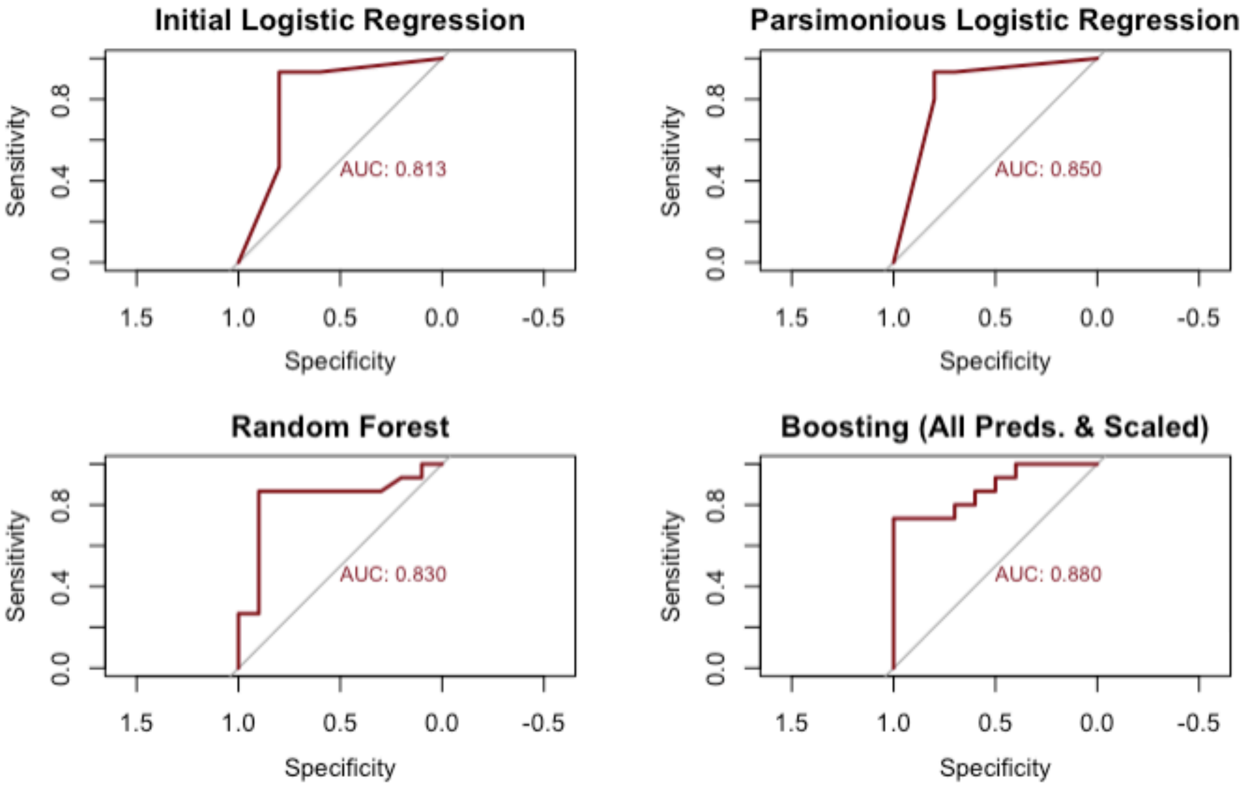
As Figure 8 shows, the best performing model for predicting heart failure hospitalizations was the boosting method using all predictors and scaled training dataset. When we compare the top ten most important variables across the best performing models for each algorithm there was a good deal of agreement. All three models established that IV Diuretics was the most influential variable. As we saw with all-cause hospitalizations, a patient's use of diuretics demonstrated a strong correlation. All three models also used Congestive HF and Tobacco Use Missing predictors. It is not surprising that Congestive HF would be common across models for predicting heart failure hospitalizations. Its coefficient estimate in both LR models suggests presence of congestive HF increases the odds of a heart failure hospitalization, but a very high p-value means this should be taken with a grain of salt.

We found it interesting that Missed Days Scaled (the number of missed device readings per month) only factored into the RF model given the previous research has shown it to be an important predictor of hospitalizations (Tran, Wolfson, O'Brien, Yousefian, & Shavelle, 2019). Once again neither of our LR models had any predictor variable coefficients that were significantly different from zero, potentially due to our small training dataset with 63 observations.

All of our predictive models had AUC scores between 0.8 and 0.9, making them excellent (Mandrekar, 2010) predictors of heart failure hospitalizations. Boosting with all predictor variables and scaled training dataset came closest to an 'outstanding' rating with an AUC of 0.880.

**Figure 8**

*Heart Failure Hospitalizations Predictive Models AUC Comparison*



## Model 3 – IV Diuretics Outside of Hospitalizations

This model sought to predict patients who are at risk for requiring IV diuretic therapy at home to treat symptoms of congestive heart failure. When we tested each of the predictor variables against the response variable using the chi-squared test or Wilcoxon-Mann Whitney test the following predictor variables were identified as associated with the response variable: Total Hospital Admission Count, Heart Failure Hospital Admission Count, Secondary HF Admission Count, PAD Threshold Range Max, PA Diastolic Max, PAD Flag Scaled, Heart Rate Mean, All Admission, HF Admission, Insulin, CPAP Usage, Congestive HF, Tobacco Use, and Ejection Fraction Cat.

### *Logistic Regression*

The initial logistic regression against the scaled training dataset and using the predictor variables identified above had an AUC value of 0.843. By examining the coefficient estimates in Table 10 we see that increases/positives in Total Hospital Admission Count, All Admission, PAD Flag Scaled, and Heart Failure Hospital Admission Count all increase the odds of IV diuretics while the remaining coefficient estimates decrease the odds. As we can see from the p-values none of the predictor variables' coefficients were significantly different from zero.

**Table 10**

*IV Diuretics Initial Logistic Regression Results*

| Term | Estimate | Std. Error | Statistic | p-value |
|---|---|---|---|---|
| HFAdmission1 | -226.0360 | 253,422.5716 | -0.0009 | 0.9993 |
| TotalHospitalAdmissionCount | 12.1708 | 20,611.3920 | 0.0006 | 0.9995 |
| EjectionFractionCatNormal | -181.0653 | 311,835.2673 | -0.0006 | 0.9995 |
| (Intercept) | 492.0922 | 848,834.8779 | 0.0006 | 0.9995 |
| EjectionFractionCatReduced | -167.3969 | 293,922.3996 | -0.0006 | 0.9995 |
| AllAdmission1 | 47.8533 | 140,300.9873 | 0.0003 | 0.9997 |
| CPAPUsage1 | -36.3879 | 116,139.3172 | -0.0003 | 0.9998 |
| Heart_Rate_Mean | -0.8891 | 3,657.5330 | -0.0002 | 0.9998 |
| TobaccoUseNever | -173.2243 | 798,606.8493 | -0.0002 | 0.9998 |
| EjectionFractionCatMissing | -63.6127 | 329,503.6304 | -0.0002 | 0.9998 |
| SecondaryHFAdmissionCount | -40.4141 | 209,393.6265 | -0.0002 | 0.9998 |
| TobaccoUseFormer | -223.4751 | 1,265,793.6530 | -0.0002 | 0.9999 |
| PA_Diastolic_Max | -1.5868 | 10,809.0514 | -0.0001 | 0.9999 |
| PAD_Threshold_Range_Max | -3.6034 | 26,837.8764 | -0.0001 | 0.9999 |
| CongestiveHF1 | -15.4575 | 139,852.2254 | -0.0001 | 0.9999 |
| PAD_Flag_Scaled | 1.5075 | 14,234.9203 | 0.0001 | 0.9999 |
| TobaccoUseMissing | -110.8909 | 1,152,668.3197 | -0.0001 | 0.9999 |
| HeartFailureHospitalAdmissionCount | 6.2357 | 70,629.9302 | 0.0001 | 0.9999 |
| Insulin1 | -4.2248 | 155,638.5766 | -0.0000 | 1.0000 |

When we used the backwards stepwise variable selection, the result was a model with an AIC of 18, down from the initial AIC of 38, and six variables to include in the next model. The more parsimonious model with an AUC of 0.877 outperformed the original LR model. The coefficient estimates increased by several magnitudes and Insulin switched signs. Again, none of the predictor variable coefficient p-values were significantly different from zero in this model.

**Table 11**

*IV Diuretics Parsimonious Logistic Regression Results*

| Term | Estimate | Std. Error | Statistic | p-value |
|---|---|---|---|---|
| HFAdmission1 | -2,073.4582 | 257,963.3500 | -0.0080 | 0.9936 |
| CPAPUsage1 | -795.4410 | 99,279.6733 | -0.0080 | 0.9936 |
| TotalHospitalAdmissionCount | 148.5427 | 18,578.3711 | 0.0080 | 0.9936 |
| SecondaryHFAdmissionCount | -296.9782 | 38,007.4005 | -0.0078 | 0.9938 |
| Insulin1 | 264.0025 | 36,296.2795 | 0.0073 | 0.9942 |
| EjectionFractionCatReduced | -2,141.0127 | 47,453,880.8768 | -0.0000 | 1.0000 |
| (Intercept) | 2,117.7076 | 47,453,863.9106 | 0.0000 | 1.0000 |
| EjectionFractionCatNormal | -1,647.7821 | 47,453,574.4820 | -0.0000 | 1.0000 |
| EjectionFractionCatMissing | -1,265.0453 | 47,453,671.0128 | -0.0000 | 1.0000 |

### *Random Forest*

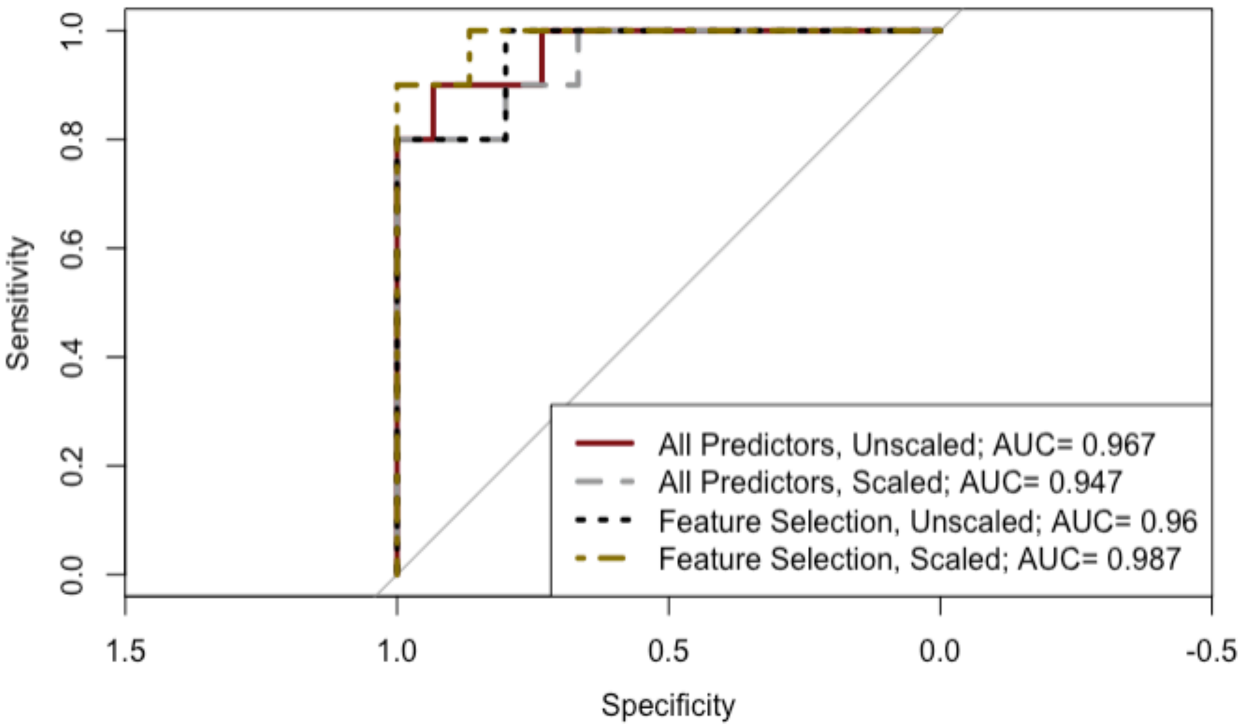
Our RF algorithm with all predictors and unscaled training dataset had an AUC of 0.953, which outperformed both the initial LR and the parsimonious LR. The best performing model from our tuning parameters had 84 predictor variables per split. The top five most important predictor variables for the RF model and their importance scores were Total Hospital Admission Count (7.157), All Admission (5.435), Tobacco Use Missing (1.594), HF Admission (1.518), and Heart Failure Hospital Admission Count (1.416).

### *Boosting*

All of the boosting models outperformed the RF model and both LR models. As Figure 9 shows, the order of best to worst performing models were selected predictors and scaled training dataset (AUC = 0.987), all predictors with unscaled training dataset (AUC = 0.967), selected predictors and unscaled training dataset (AUC = 0.960), and all predictors and scaled training dataset (AUC = 0.947).

**Figure 9**

*IV Diuretics Boosting AUC Comparison*



For the best performing model, the top five most important variables and their relative influence values were Heart Failure Hospital Admission Count (22.202), PAD Flag Scaled (20.077), Secondary HF Admission Count (16.127), HF Admission (15.243), and Total Hospital Admission Count (10.263). This model's parameters were number of trees = 1,500, interaction depth = 3, and shrinkage = 0.001.

### *IV Diuretic Discussion*

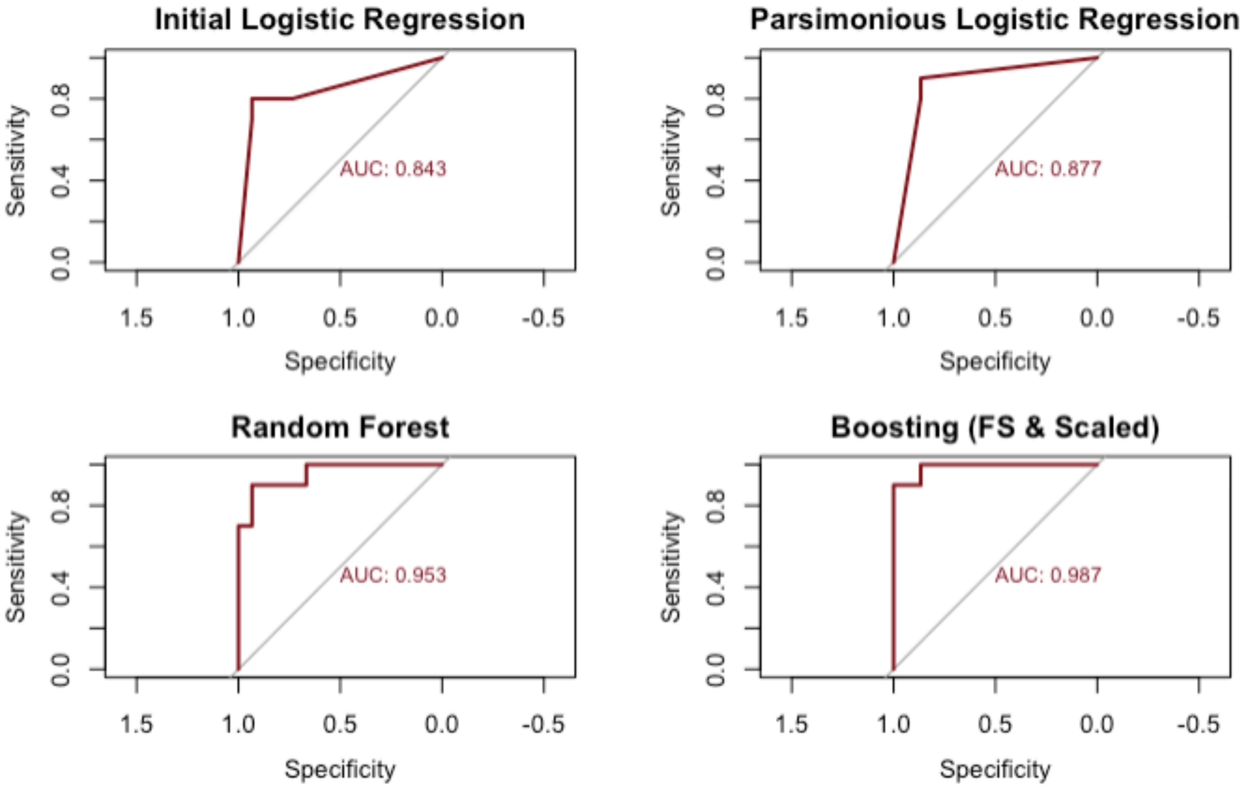
As Figure 10 shows, the boosting method with selected predictors and scaled training dataset was the best performing model for predicting IV diuretic therapy. When we compare the top ten most important variables across the best performing models for each algorithm there was some commonality. All three models agreed that HF Admission, Total Hospital Admission Count, and Secondary HF Admission Count were significant predictors. Random Forest and

Boosting both used additional predictors of All Admission and Heart Failure Admission Count. From these results we can conclude that hospitalizations are highly correlated with IV diuretics. For the third time our LR models failed to have any predictor variable coefficients significantly different from zero. This could again be due to the training dataset having only 63 observations.

All of our predictive methods are effective at forecasting use of IV diuretics with LR classified as excellent and RF and boosting classified as outstanding according to the scale from Mandrekar (2010). Boosting using predictor variables selected by chi-squared or Wilcoxon-Mann Whitney tests for association and scaled training data had a nearly perfect AUC of 0.987.

**Figure 10**

*IV Diuretics Predictive Models AUC Comparison*



## Model 4 – PAD Threshold Changes

This model sought to predict patients who are at risk for their PAD pressure thresholds being changed by their care team. When we tested each of the predictor variables against the response variable using the chi-squared test or Wilcoxon-Mann Whitney test the following predictor variables were identified as associated with the response variable: PAD Threshold Range Min, PAD Flag Scaled, PA Diastolic Diff, PAD Threshold Range Max, Length Participation, PA Mean Diff, PA Pulsatility Min, Hypertension, GLP1, and High Cholesterol.

### *Logistic Regression*

The initial logistic regression against the scaled training dataset and using the predictor variables identified above had an AUC value of 0.733. By examining the coefficient estimates in Table 12 we see that increases/positives in GLP11, Length Participation, PAD Flag Scaled, PA Mean Diff, PAD Threshold Range Min, and PAD Threshold Range Max all increase the odds of PAD threshold changes while the remaining coefficient estimates decrease the odds. Four of the top five predictor variables had coefficients significantly different from zero using a 0.1 level of significance.

**Table 12**

*PAD Threshold Changes Initial Logistic Regression Results*

| Term | Estimate | Std. Error | Statistic | p-value |
|---|---:|---:|---:|---:|
| PA_Pulsatility_Min | -0.2029 | 0.0779 | -2.6040 | 0.0092 |
| GLP11 | 4.6548 | 1.8614 | 2.5007 | 0.0124 |
| HighCholesterol1 | -2.5175 | 1.1503 | -2.1886 | 0.0286 |
| LengthParticipation | 0.1883 | 0.0865 | 2.1785 | 0.0294 |
| Hypertension1 | -1.8433 | 1.2355 | -1.4920 | 0.1357 |
| PAD_Flag_Scaled | 0.1366 | 0.1162 | 1.1763 | 0.2395 |
| PA_Mean_Diff | 0.0740 | 0.0653 | 1.1343 | 0.2566 |
| PAD_Threshold_Range_Min | 0.2645 | 0.3186 | 0.8303 | 0.4064 |
| PAD_Threshold_Range_Max | 0.0642 | 0.2313 | 0.2777 | 0.7813 |
| (Intercept) | 0.6240 | 2.7198 | 0.2294 | 0.8185 |
| PA_Diastolic_Diff | -0.0038 | 0.1137 | -0.0338 | 0.9730 |

When we used the backwards stepwise variable selection, the result was a model with an AIC of 56.11, down from the initial AIC of 61.64, and six variables to include in the next model. The more parsimonious model with an AUC of 0.767 outperformed the original LR model. The coefficient estimates as seen in Table 13 did not change significantly from the initial LR model. Again, four of the top five predictor variables had coefficients significantly different from zero using a 0.1 level of significance.

**Table 13**

*PAD Threshold Changes Parsimonious Regression Results*

| Term | Estimate | Std. Error | Statistic | p-value |
|---|---:|---:|---:|---:|
| GLP11 | 3.7515 | 1.3437 | 2.7919 | 0.0052 |
| PA_Pulsatility_Min | -0.1840 | 0.0718 | -2.5632 | 0.0104 |
| LengthParticipation | 0.1863 | 0.0816 | 2.2837 | 0.0224 |
| HighCholesterol1 | -2.1188 | 1.0234 | -2.0703 | 0.0384 |
| Hypertension1 | -1.7344 | 1.1046 | -1.5703 | 0.1164 |
| (Intercept) | 2.3337 | 1.5553 | 1.5005 | 0.1335 |
| PA_Mean_Diff | 0.0725 | 0.0497 | 1.4599 | 0.1443 |

### *Random Forest*

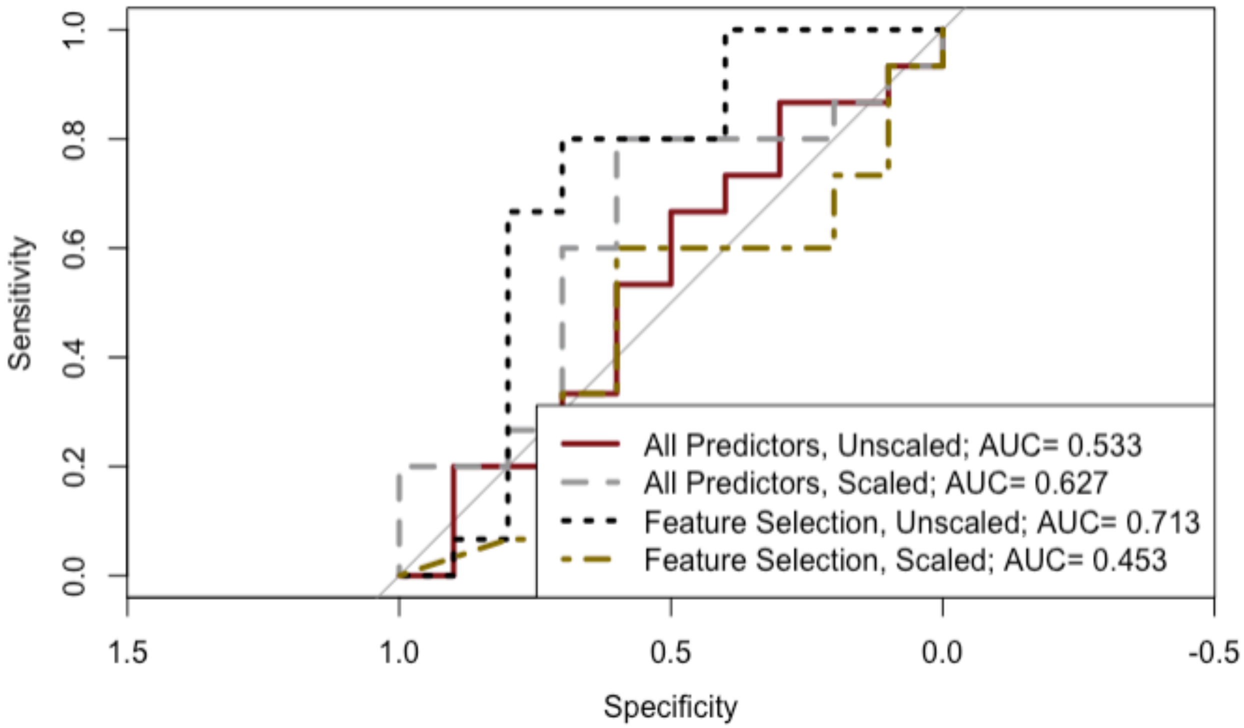
Our random forest algorithm with all predictors and unscaled training dataset had an AUC of 0.620, which failed to outperform either of the LR models. The best performing model from our tuning parameters had 40 predictor variables per split. The top five most important predictor variables for the random forest model and their importance scores were PA Pulsatility Min (1.926), PAD Threshold Range Max (1.824), Heart Rate Max (1.707), PA Diastolic Max (1.682), and Length Participation (1.642).

***Boosting***

The boosting models overall performed poorly when compared to LR and, to some extent, RF. Figure 11 shows that the top two boosting models, selected predictors and unscaled training dataset (AUC = 0.713) and all predictors and scaled training dataset (AUC = 0.627) outperformed RF. The bottom two boosting models, all predictors and unscaled training dataset (AUC = 0.533) and selected predictors and scaled training dataset (AUC = 0.453) failed to outperform any other model and had AUC values that indicated performances roughly equal to, or worse than, random guessing.

**Figure 11**

*PAD Threshold Changes Boosting AUC Comparison*



For the best performing model, the top five most important variables and their relative influence values were PA Pulsatility Min (25.824), Length Participation (14.676), PAD Threshold Range Min (13.869), PAD Flag Scaled (13.141), and PAD Threshold Range Max

(12.008). This model's parameters were number of trees = 2,000, interaction depth = 2, and shrinkage = 0.01.

### *PAD Threshold Changes Discussion*

As Figure 12 shows, the parsimonious LR model was the best performing model for predicting PAD threshold changes. When we compared the top ten most important variables across the best performing models for each algorithm, parsimonious LR and boosting had substantial overlap while more than half of the RF model's top predictors were unique. All of the predictors used in the parsimonious LR were used in the best boosting model. LR, RF, and boosting all showed consensus in using the predictor variables PA Pulsatility Min and Length Participation

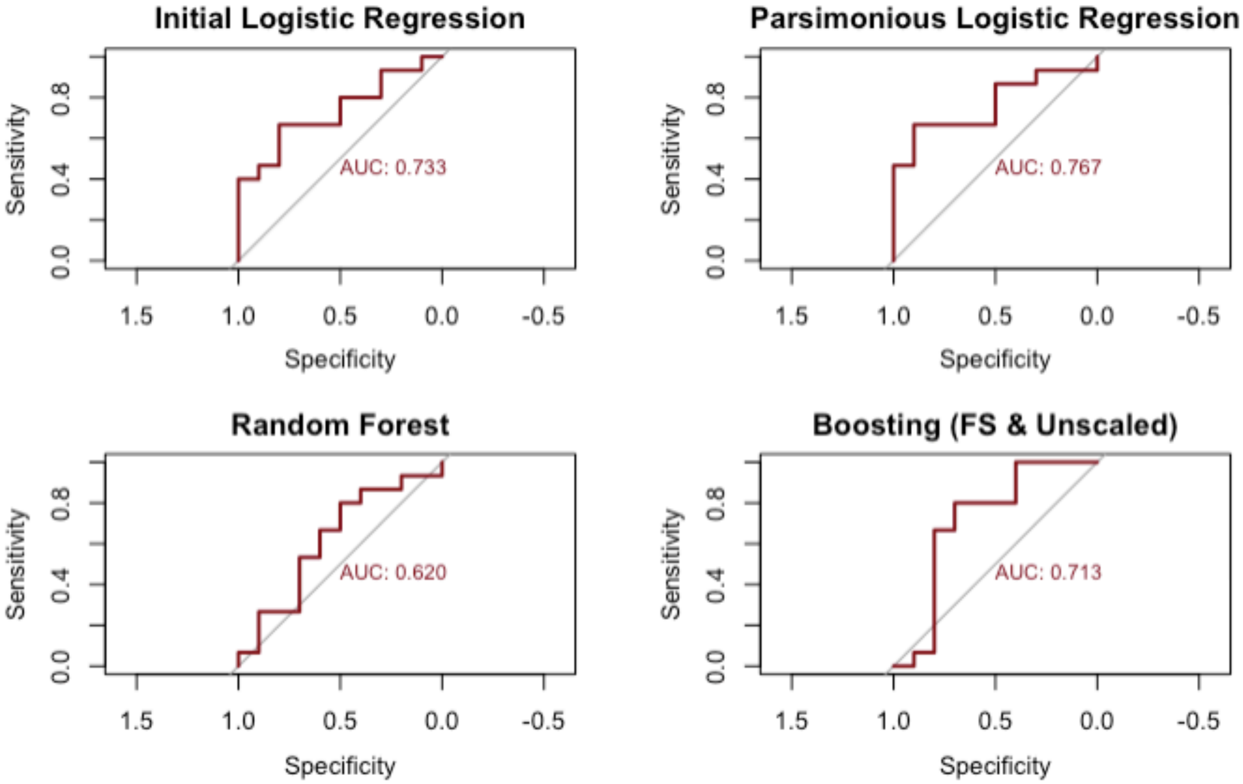
It is interesting that this is the only one of the four models that had logistic regression variable coefficients that were significantly different than zero at a 0.1 level of significance despite having the same training dataset set (n = 63), and the is the overall worst performing model in our analysis.

Our models are not good predictors of PAD threshold changes. RF, and three of the four boosting models, were not acceptable with AUCs below 0.7. LR and boosting with predictors selected by chi-squared or Wilcoxon-Mann Whitney tests for association and unscaled training data were just beyond the acceptable range.

The lack of predictive ability could be due to how sparse this indicator was in the original CardioMEMS Device dataset, with only 65 records out of 16,503 (0.39%). We attempted to counteract this sparsity by totaling the number of PAD threshold changes per patient and scaling by the length of participation, resulting in 41% of patients having a True/Yes indicator for this response variable. We also used scaled training datasets to further reduce bias toward the False response value.

**Figure 12**

*PAD Threshold Changes Predictive Models AUC Comparison*



### Overall Discussion

When we looked across the results from all models, we saw a few interesting trends. Random Forests was not the best predictor for any of our models. This is despite the fact that it fit well with our dataset's characteristics of many predictor variables, correlation between predictors, outliers, and imbalanced datasets. Because RF is such a flexible and powerful classification method, we did not attempt to compare results to a model trained on the scaled dataset.

Three of the four models had a scaled dataset as part of their best-performing models. Only all-cause hospitalizations did not. All-cause hospitalizations had the most unbalanced dataset out of all three models, with 70% of patients having a True/Yes value. In this case we may want to choose the best performing boosting model trained with scaled data (all predictors, AUC = 0.944).

The predictor variables that showed up most frequently in the top performing LR, RF, and boosting algorithms across all four models were Tobacco Use Missing (7), IV Diuretic (6), Congestive HF (5), and Insulin (5). It was not surprising to see that Congestive HF and IV Diuretics are influential in predicting our targeted events and adds a level of confirmation to our results.

## Recommendations

### Next Steps

TUKHS should review the results of this analysis and obtain approval from the IRB prior to using any of the models in patient care for all models except PAD threshold changes, whose AUC was barely within an acceptable range and should not be used until it is improved. When selecting the predictive algorithms for the remaining targeted events, we recommend using the top performing algorithm trained on a scaled dataset, even if it did not have the largest AUC for that event. This is because there may still be some bias in the algorithms toward the majority event in the unscaled training data. For both all-cause and heart failures hospitalizations that algorithm was boosting with all predictors. For IV diuretics that algorithm was boosting with selected predictors.

Through our analysis we identified the most significant predictors of each targeted event. The clinical team at TUKHS Heart Failure Program should provide special focus on these metrics for the patients with CardioMEMS devices as the models are implemented.

To validate this analysis and potentially improve algorithm performance, TUKHS should continue to collect additional data, both going forward with additional patient and device observations and retrospectively looking to fill in any missing values in the existing data. This is especially true with the categorical data fields for Tobacco Use and Ejection Fraction which had 20 and 27 missing values, respectively. Tobacco Use – Missing was one of the top predictors across all models so it would be worth investigating if there is a pattern to why that data is missing or what affect getting correct values has on the models' predictive performances.

In addition to gathering more observations, TUKHS should include additional clinical conditions and diagnostic test results from CardioMEMS patients' EHR records that may improve algorithm performance, especially the PAD Threshold Changes model. Examples of additional predictor variables to investigate include rales, shortness of breath at rest, peak VO2, hemoglobin, lymphocytes, and NT-proBNP (see, e.g. Jing, et al., 2018; Calvillo–King, et al., 2013; O'Connor, et al., 2010).

## Conclusion

Our analysis was an attempt to employ supervised classification algorithms with interpretable results to predict four heart failure-related targeted events using a combination of CardioMEMS device and patient EHR data. Based on previous research we selected logistic regression, random forest, and boosting with decision trees as the best algorithms given our data and objectives. Our results indicated that boosting, when combined with scaled training datasets, provided excellent to outstanding results for predicting all-cause hospitalizations, heart failure hospitalizations, and IV diuretic therapy. Our chosen algorithms were not effective at predicting threshold change events for pulmonary artery diastolic pressure. Our analysis also produced a number of significant features related to our four targeted events that can be monitored by the clinical team at TUKHS Heart Failure Program. If and when the three effective models are put into use, our analysis could help focus the care of the clinical team towards patients with the greatest risk of adverse events, reduce healthcare costs, and improve the quality of life for CardioMEMS heart failure patients.

# References

Bradley, A. E. (1997). THE USE OF THE AREA UNDER THE ROC CURVE IN THE EVALUATION OF MACHINE LEARNING ALGORITHMS. *Pattern Recognition*, 1145-1159.

Calvillo–King, L., Arnold, D., Eubank, K., Lo, M., Yunyongying, P., Stieglitz, H., & Halm, E. (2013). Impact of Social Factors on Risk of Readmission or Mortality in Pneumonia and Heart Failure: Systematic Review. *Journal of General Internal Medicine*, 269-282.

CardioMEMS, Inc. (2014, May 28). *CardioMEMS HF System [Video]*. Retrieved from YouTube: https://www.youtube.com/watch?v=OVwEpL6cT-A&feature=youtu.be&t=62

Dai, W., Brisimi, T. S., Adams, W. G., Mela, T., Saligrama, V., & Paschalidis, I. C. (2015). Prediction of hospitalization due to heart diseases by supervised learning methods. *International Journal of Medical Informatics*, 189-197.

Ferguson, K. K., Yu, Y., Cantonwine, D. E., McElrath, T. F., Meeker, J. D., & Mukherjee, B. (2018). Foetal ultrasound measurement imputations based on growth curves versus multiple imputation chained equation (MICE). *Paediatric and Perinatal Epidemiology*, 469-473.

Givertz, M. M., Stevenson, L. W., Costanzo, M. R., Bourge, R. C., Bauman, J. G., Ginn, G., . . . Investigators, C. T. (2017). Pulmonary Artery Pressure-Guided Management of Patients With Heart Failure and Reduced Ejection Fraction. *Journal of the American College of Cardiology*, 1875-1886.

Hasanin, T., Khoshgoftaar1, T. M., Leevy, J. L., & Seliya, N. (2019). Examining characteristics of predictive models with imbalanced big data. *Journal of Big Data*, 1-21.

Heywood, J. T., Jermyn, R., Shavelle, D., Abraham, W., Bhimaraj, A., Bhatt, K., . . . Stevenson, L. (2017). Impact of Practice-Based Management of Pulmonary Artery Pressures in 2000 Patients Implanted With the CardioMEMS Sensor. *Circulation*, 1509-1517.

Inglis, S. C., Clark, R. A., Dierckx, R., Prieto-Merino, D., & Cleland, J. G. (2015). Structured telephone support or non-invasive telemonitoring for patients with heart failure. *Cochrane Database Of Systematic Reviews*, CD007228.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R.* New York, NY: Springer New York.

Jing, L., Doddamani, S., Hartzel, D., Ulloa Cerna, A., Sauers, N., Good, C., . . . Fornwalt, B. (2018). Predicting Mortality and Hospitalization in 11,327 Patients With Heart Failure Using Machine Learning. *Circulation*, Vol.138 Suppl 1.

Jovanovica, M., Radovanovica, S., Vukicevica, M., Pouckeb, S. V., & Delibasic, B. (2016). Building interpretable predictive models for pediatric hospital readmission using Tree-Lasso logistic regression. *Artificial Intelligence in Medicine*, 12-21.

Kansas Department of Health and Environment Division of Public Health Bureau of Epidemiology and Public Health Informatics. (2018, November). *Annual Summary of Vital Statistics, 2017.* Retrieved from https://www.kdheks.gov/phi/as/2017/Annual_Summary_2017.pdf

KansasHealthMatters. (2020, January). *Indicators :: Congestive Heart Failure Hospital Admission Rate :: Public Health Preparedness Region : Kansas City Metro*. Retrieved from KansasHealthMatters: https://www.kansashealthmatters.org/?module=indicators&controller=index&action=view&comparisonId=&indicatorId=6747&localeTypeId=21&localeId=131164

Koehler, F., Koehler, K., Deckwart, O., Prescher, S., Wegscheider, K., Kirwan, B.-A., . . . Butte. (2018). Efficacy of telemedical interventional management in patients with heart failure (TIM-HF2): a randomised, controlled, parallel-group, unmasked trial. *Lancet*, 1047-1057.

Lee, I., & Shin, Y. J. (2020). Machine learning for enterprises: Applications, algorithm selection, and challenges. *Business Horizons*, 157-170.

Ling, C. X., & Zhang, H. (2002). Toward Bayesian classifiers with accurate probabilities. *Advances in Knowledge Discovery and Data Mining: 6th Pacific-Asia Conference, PAKDD 2002 Taipei, Taiwan, May 6–8, 2002 Proceedings* (pp. Vol. 2336, Lecture Notes in Computer Science, pp. 123-134). Berlin, Heidelberg: Springer Berlin Heidelberg.

Lorenzoni, G., Sabato, S. S., Lanera, C., Bottigliengo, D., Minto, C., Ocagli, H., . . . Pisanò, F. (2019). Comparison of Machine Learning Techniques for Prediction of Hospitalization in Heart Failure Patients. *Journal Of Clinical Medicine*, 1-13.

Mandrekar, J. N. (2010). Receiver Operating Characteristic Curve in Diagnostic Test Assessment. *Journal of Thoracic Oncology*, 1315-1316.

Martin-Barragan, B., Lillo, R., & Romo, J. (2014). Interpretable support vector machines for functional data. *European Journal of Operational Research*, 146-155.

Mortazavi, B. J., Downing, N. S., Bucholz, E. M., Dharmarajan, K., Manhapra, A., Li, S.-X., . . . Krumholz, H. M. (2016). Analysis of Machine Learning Techniques for Heart Failure Readmissions. *Circulation: Cardiovascular Quality and Outcomes*, 629-640.

NGUYEN, D.-H., & LE, M. T. (2014). Improving the Interpretability of Support Vector Machines-based Fuzzy Rules.

Nichols, G. A., Pesa, J., & Patel, A. (2018). Predicting New Heart Failure Hospitalization Among Patients with an Existing Heart Failure Diagnosis. *Journal of Cardiac Failure*, S68-S69.

O'Connor, C., Wojdyla, D., Leifer, E., Ellis, S., Lee, K., Clare, R., . . . Whellan, D. (2010). DETERMINANTS OF MORBIDITY AND MORTALITY IN CHRONIC HEART FAILURE (CHF) WITH SYSTOLIC DYSFUNCTION: RESULTS OF THE HF-ACTION PREDICTIVE MODEL. *Journal of the American College of Cardiology*, A28.E270.

Portet, S. (2020). A primer on model selection using the Akaike Information Criterion. *Infectious Disease Modelling*, 111-128.

Putri, B. D., Notobroto, H. B., & Wibowo, A. (2018). Comparison of MICE and Regression Imputation for Handling Missing Data. *Health Notions*, 183-186.

Ramyachitra, D., & Manikandan, P. (2014). IMBALANCED DATASET CLASSIFICATION AND SOLUTIONS: A REVIEW. *International Journal of Computing and Business Research*.

Raval, N. Y., Shavelle, D., Bourge, R. C., Costanzo, M. R., Shlofmitz, R., Heywood, J. T., . . . Stevenson, L. W. (2017). Significant Reductions in Heart Failure Hospitalizations with the Pulmonary Artery Pressure Guided HF System: Preliminary Observations From the CardioMEMS Post Approval Study. *Journal of Cardiac Failure*, S27.

Rekha Mankad, M. (2019, July 2). *Ejection fraction: What does it measure?* Retrieved from Mayo Clinic: https://www.mayoclinic.org/ejection-fraction/expert-answers/faq-20058286

Robinson, R. L., Palczewska, A., Palczewski, J., & Kidley, N. (2017). Comparison of the Predictive Performance and Interpretability of Random Forest and Linear Models on Benchmark Data Sets. *Journal of Chemical Information and Modeling*, 1773-1792.

Tran, J. S., Wolfson, A. M., O'Brien, D., Yousefian, O., & Shavelle, D. M. (2019). A Systems-Based Analysis of the CardioMEMS HF Sensor for Chronic Heart Failure Management. *Cardiology Research and Practice*, 7.

VakhitovaI, Z. I., & Alston-Knox, C. L. (2018). Non-significant p-values? Strategies to understand and better determine the importance of effects and interactions in logistic regression. *PLoS ONE*, E0205076.

Veenis, J. F., & Brugts, J. J. (2020). Remote monitoring of chronic heart failure patients: invasive versus non-invasive tools for optimising patient management. *Netherlands Heart Journal*, 3-13.

Wiens, J., & Shenoy, E. S. (2018). Machine Learning for Healthcare: On the Verge of a Major Shift in Healthcare Epidemiology. *Clinical Infectious Diseases*, 149-153.

Wu, J., Roy, J., & Stewart, W. F. (2010). Prediction Modeling Using EHR Data Challenges, Strategies, and a Comparison of Machine Learning Approaches. *Medical Care*, S106-S113.

Zame, W., Yoon, J., Asselbergs, F., & van der Schaar, M. (2018). Interpretable Machine Learning Identifies Risk Predictors in Patients With Heart Failure. *Circulation*, 138(Suppl_1), A14882.

## Appendix A

**Table 14**

*CardioMEMS Device Dataset*

| Variable Name | Description | Format |
|---|---|---|
| MeasureID | Unique identifier for each patient, month device measurement. | Numeric |
| ID | Unique patient identifier. | Numeric |
| MeasurementDT_UTC | Date of the device measurement in Coordinated Universal Time (UTC). | Date |
| PA_Systolic | Pulmonary artery systolic (PAS) pressure in mmHg. | Numeric |
| PA_Diastolic | Pulmonary artery diastolic (PAD) pressure in mmHg. | Numeric |
| PA_Pulsatility | PAS - PAD in mmHg. | Numeric |
| PA_Mean | Mean pulmonary artery pressure, calculated as (2/3)(PAD) + (1/3)(PAS) in mmHg. | Numeric |
| Heart_Rate | Heart rate in beats per minute. | Numeric |
| PA_Diastolic_Threshold_Lower | Lower end of target range for PAD in mmHg. | Numeric |
| PA_Diastolic_Threshold_Upper | Higher end of target range for PAD in mmHg. | Numeric |
| PAD_Flag | Indicator if the PAD is higher or lower than the threshold range. | Y/N (1/0) |
| ThresholdChangeIndicator | Indicator if the thresholds have changed since the previous measurement for this patient. | Y/N (1/0) |

**Table 15**

*Risk Factors Dataset*

| Variable Name | Description | Format | Associated ICD Codes |
|---|---|---|---|
| ID | Unique patient identifier. | Numeric | |
| Male | Patient gender. | Y/N (1/0) | |
| AllAdmission | History of all-cause hospital admission. | Y/N (1/0) | |
| TotalHospitalAdmissionCount | Number of Previous all-cause hospitalizetions. | Numeric | |
| HFAdmission | History of heart failure hospital admission. | Y/N (1/0) | |
| HeartFailureHospitalAdmission Count | Number of hospitalizetions with heart failure primary diagnosis. | Numeric | |
| 2ndHFAdmission | Number of hospitalizetions with heart failure secondary diagnosis. | Numeric | |
| IV Diuretic | Previous use of IV diuretics from outside of hospital. | Y/N (1/0) | |
| AvgSystolic | Average systolic blood pressure in mmHg. | Numeric | |
| AvgDiastolic | Average diastolic blood pressure in mmHg. | Numeric | |

| Variable Name | Description | Format | Associated ICD Codes |
|---|---|---|---|
| AverageDailyBPCategory | 0 - normal<br>1 - elevated<br>2 - stage I<br>3 - stage II | Factor | |
| BPUncontrolled | Controlled blood pressure with controlled defined as < 149/90 mmHg. | Y/N (1/0) | |
| HypertensionUncomplicated | | Y/N (1/0) | ICD 10: I10<br>ICD 9: 401.1, 401.9, 642.00-642.04 |
| HypertensionComplicated | | Y/N (1/0) | ICD 10: I11, I12, I13, I15<br>ICD 9: 401.0, 402.00, 402.10, 402.90, 403.00, 403.10, 403.90, 404.00, 404.10, 404.90, 405.01, 405.09, 405.11, 405.19, 405.91, 405.99 642.10-642.24, 642.70-642.94 |
| EjectionFraction | | Numeric | |
| WeightLoss | | Y/N (1/0) | ICD 10: E40, E41, E42, E43, E44, E45, E46, R64, R634<br>ICD 9: 260-263.9 |
| CalculatedBMI | Body mass index calculation as weight in kg divided by height in meters, squared | Numeric | |
| BMICategory | 0 – Underweight (<18.5)<br>1 – Normal (18.5-24.9)<br>2 – Normal (25-29.9)<br>3 – Obese (30-39.9)<br>4 - Extreme obesity (>=40) | Factor | |

| Variable Name | Description | Format | Associated ICD Codes |
|---|---|---|---|
| Obesity | | Y/N (1/0) | ICD 10: E66<br>ICD 9: 278.0, 278.00, 278.01 |
| DiabetesUncomplicated | | Y/N (1/0) | ICD 10: E100, E101, E109, E110, E111, E119, E120, E121, E129, E130, E131, E139, E140, E141, E149<br>ICD 9: 250.00-250.33 |
| DiabetesComplicated | | Y/N (1/0) | ICD 10: E102, E103, E104, E105, E106, E107, E108, E112, E113, E114, E115, E116, E117, E118, E122, E123, E124, E125, E126, E127, E128, E132, E133, E134, E135, E136, E137, E138, E142, E143, E144, E145, E146, E147, E148<br>ICD 9: 250.40-250.93 |
| Arrythmia | Cardiac arrythmia, atrial fibrillation, ventricular tachycardia, atrial flutter, supra-ventricular tachycardia | Y/N (1/0) | ICD10: I46.9, I47.1, I47.2, I48.0, I48.19, I48.20, I48.21, I48.91, I48.92, I49.01, Z86.79, Z98.890<br>ICD9: 427, 427.1, 427.31, 427.32, 427.41, 427.5, 427.89, V12.59, V45.89 |
| CoronaryArteryDisease | | Y/N (1/0) | ICD10: I25.10, I25.119, I25.810<br>ICD9: 413.9, 414, 414.01, 414.05 |
| ValvularHeartDisease | | Y/N (1/0) | ICD 10: A520, I091, I098, Q230, Q231, Q232, Q233, Z952, Z953, Z954, I05, I06, I07, I08, I34, I35, I36, I37, I38, I39<br>ICD 9: 093.20-093.24, 394.0-397.1, 424.0-424.91, 746.3-746.6, V42.2, V43.3 |
| HighCholesterol | History of Hyper-lipidemia/ Dyslipidemia diagnosis. | Y/N (1/0) | ICD 10: E78.5, E78.2<br>ICD 9: 272.4, 272.2 |

| Variable Name | Description | Format | Associated ICD Codes |
|---|---|---|---|
| TobaccoUse | 0 - Never<br>2 - Former<br>1 - Current<br>3 - No response | Factor | |
| AlcoholAbuse | | Y/N (1/0) | ICD 10: G621, I426, K292, K700, K703, K709, Z502, Z714, Z721, F10, E52, T51<br>ICD 9: 291.0-291.3, 291.5, 291.8, 291.81, 291.89, 291.9, 303.00-303.93, 305.00-305.03, V113 |
| SleepApnea | Suspected obstructive sleep apnea | Y/N (1/0) | |
| CPAPUsage | OSA current CPAP/BIPAP | Y/N (1/0) | |
| RenalFailure | | Y/N (1/0) | ICD 10: I120, I131, N250, Z490, Z491, Z492, Z940, Z992, N18, N19<br>ICD 9: 403.01, 403.11, 403.91, 404.02, 404.03, 404.12, 404.13, 404.92, 404.93, 585, 586, V42.0, V45.1, V56.0-V56.2, V56.8 |
| Insulin | | Y/N (1/0) | ICD 10: IMO0001, E11.9, Z79.4<br>ICD 9: IMO0002, 250.00, V58.67 |
| Age | | Numeric | |
| RaceWhite | Is the patient’s race Caucasian? | Y/N (1/0) | |
| SGLT2 | | Y/N (1/0) | |
| Metformin | History of diabetes medications. | Y/N (1/0) | |
| DPPV4 | History of diabetes medications. | Y/N (1/0) | |
| GLP1 | History of diabetes medications. | Y/N (1/0) | |

| Variable Name | Description | Format | Associated ICD Codes |
|---|---|---|---|
| Glitazones | History of diabetes medications. | Y/N (1/0) | |
| CongestiveHF | Congestive Heart Failure | Y/N (1/0) | ICD 10: I43, I50, I099, I110, I130, I132, I255, I420, I425, I426, I427, I428, I429, P290<br>ICD 9: 398.91, 402.01, 402.11, 402.91, 404.01, 404.11, 404.91, 428.0-428.9 |
| PulmonaryCirculation | Pulmonary circulation disorders. | Y/N (1/0) | ICD 10: I26, I27, I280, I288, I289<br>ICD 9: 416.0-416.9, 417.9 |
| PeripheralVascular | Peripheral vascular disorders | Y/N (1/0) | ICD 10: I70, I71, I731, I738, I739, I771, I790, I792, K551, K558, K559, Z958, Z959<br>ICD 9: 440.0-440.9, 441.00-441.9, 442.0-442.9, 443.1-443.9, 447.1, 557.1, 557.9, V43.4 |
| Paralysis | | Y/N (1/0) | ICD 10: G041, G114, G801, G802, G830, G831, G832, G833, G834, G839, G81, G82<br>ICD 9: 342.0-342.12, 342.9-344.9, 438.20-438.53 |
| NeurologicalDisorder | | Y/N (1/0) | ICD 10: G254, G255, G312, G318, G319, G931, G934, R470, G10, G11, G12, G13, G20, G21, G22, G32, G35, G36, G37, G40, G41, R56<br>ICD 9: 330.0-331.9, 332.0, 333.4, 333.5, 334.0-335.9, 340, 341.1-341.9, 345.00-345.11, 345.2-345.3, 345.40-345.91, 348.1, 348.3-348.39, 780.3, 780.39, 784.3 |
| ChronicPulmonary | | Y/N (1/0) | ICD 10: I278, I279, J684, J701, J703, J40, J41, J42, J43, J44, J45, J46, J47, J60, J61, J62, J63, J64, J65, J66, J67<br>ICD 9: 490-492.8, 493.00-493.92, 494-494.1, 495.0-505, 506.4 |

| Variable Name | Description | Format | Associated ICD Codes |
|---|---|---|---|
| Hypothyroidism | | Y/N (1/0) | ICD 10: E00, E01, E02, E03, E890<br>ICD 9: 243-244.2, 244.8, 244.9 |
| LiverDisease | | Y/N (1/0) | ICD 10: K711, K713, K714, K715, K717, K760, K762, K763, K764, K765, K766, K767, K768, K769, K70, K72, K73, K74, B18, I85, I864, I982, Z944<br>ICD 9: 070.22, 070.23, 070.32, 070.33, 070.44, 070.54, 456.0, 456.1, 456.20, 456.21, 571.0, 571.2, 571.3, 571.40-571.49, 571.5, 571.6, 571.8, 571.9, 572.3, 572.8, V42.7 |
| PepticUlcer | Peptic Ulcer disease excluding bleeding. | Y/N (1/0) | ICD 10: K257, K259, K267, K269, K277, K279, K287, K289<br>ICD 9: 531.41, 531.51, 531.61, 531.70, 531.71, 531.90, 531.91,532.41, 532.51, 532.61, 532.70, 531.71, 532.90, 532.91, 533.41, 533.51, 533.61, 533.70, 533.71, 533.90, 533.91, 534.41, 534.51, 534.61, 534.70, 534.71, 534.90, V12.71 |
| AIDS | | Y/N (1/0) | ICD 10: B20, B21, B22, B24<br>ICD 9: 042-044.9 |
| Lymphoma | | Y/N (1/0) | ICD 10: C81, C82, C83, C84, C85, C88, C96, C900, C902<br>ICD 9: 200.00-202.38, 202.50-203.01, 203.8-203.81, 238.6, 273.3, V10.71, V10.72, V10.79 |
| Cancer | Metastatic cancer. | Y/N (1/0) | ICD 10: C77, C78, C79, C80<br>ICD 9: 196.0-199.1 |

| Variable Name | Description | Format | Associated ICD Codes |
|---|---|---|---|
| Tumor | Solid tumor without metastasis. | Y/N (1/0) | ICD 10: C00, C01, C02, C03, C04, C05, C06, C07, C08, C09, C10, C11, C12, C13, C14, C15, C16, C17, C18, C19, C20, C21, C22, C23, C24, C25, C26, C30, C31, C32, C33, C34, C37, C38, C39, C40, C41, C43, C45, C46, C47, C48, C49, C50, C51, C52, C53, C54, C55, C56, C57, C58, C60, C61, C62, C63, C64, C65, C66, C67, C68, C69, C70, C71, C72, C73, C74, C75, C76, C97<br>ICD 9: 140.0-172.9, 174.0-175.9, 179-195.8, V10.00-V10.59, V10.81-V10.9 |
| Arthritis | Rheumatoid arthritis/ collagen vascular disease. | Y/N (1/0) | ICD 10: M05, M06, M08, M120, M123, M310, M311, M312, M313, M30, M32, M33, M34, M35, M45, M461, M468, M469, L940, L941, L943<br>ICD9: 701.0, 710.0-710.9, 714.0-714.9, 720.0-720.9, 725 |
| Coagulopathy | | Y/N (1/0) | ICD 10: D65, D66, D67, D68, D691, D693, D694, D695, D696<br>ICD 9: 2860-2869, 287.1, 287.3-287.5, 289.81-289.82 |
| ElectrolyteDisorder | | Y/N (1/0) | ICD 10: E86, E87, E222<br>ICD 9: 276.0-276.9 |
| LossAnemia | | Y/N (1/0) | ICD 10: D500<br>ICD 9: 2800, 648.20-648.24 |
| DeficiencyAnemia | | Y/N (1/0) | ICD 10: D508, D509, D51, D52, D53<br>ICD 9: 280.1-281.9, 285.21-285.29, 285.9 |
| DrugAbuse | | Y/N (1/0) | ICD 10: F11, F12, F13, F14, F15, F16, F18, F19, Z715, Z722<br>ICD 9: 292.0, 292.82-292.89, 292.9, 304.00-304.93, 305.20-305.93, 648.30-648.34 |

| Variable Name | Description | Format | Associated ICD Codes |
|---|---|---|---|
| Psychoses | | Y/N (1/0) | ICD 10: F20, F22, F23, F24, F25, F28, F29, F302, F312, F315<br>ICD 9: 295.00-298.9, 299.10, 299.11 |
| Depression | | Y/N (1/0) | ICD 10: F32, F33, F204, F313, F314, F315, F341, F412, F432<br>ICD 9: 300.4, 301.12, 309.0, 309.1, 311 |

# Appendix B

The code related to this paper is stored in a GitHub repository located at: https://github.com/yelirkram/Applying-Predictive-Analytic-Capabilities-for-CardioMEMS-Patients.