

# Building Computational Models to Predict One-Year Mortality in ICU Patients with Acute Myocardial Infarction and Post Myocardial Infarction Syndrome

Laura A. Barrett, BA<sup>1</sup>, Seyedeh Neelufar Payrovnaziri, MS<sup>1</sup>,  
Jiang Bian, PhD<sup>2</sup>, Zhe He, PhD<sup>1</sup>

<sup>1</sup>School of Information, Florida State University, Tallahassee, Florida, USA;

<sup>2</sup>Department of Health Outcomes and Biomedical Informatics, University of Florida, Gainesville, Florida, USA

## Abstract

*Heart disease remains the leading cause of death in the United States. Compared with risk assessment guidelines that require manual calculation of scores, machine learning-based prediction for disease outcomes such as mortality can be utilized to save time and improve prediction accuracy. This study built and evaluated various machine learning models to predict one-year mortality in patients diagnosed with acute myocardial infarction or post myocardial infarction syndrome in the MIMIC-III database. The results of the best performing shallow prediction models were compared to a deep feedforward neural network (Deep FNN) with back propagation. We included a cohort of 5436 admissions. Six datasets were developed and compared. The models applying Logistic Model Trees (LMT) and Simple Logistic algorithms to the combined dataset resulted in the highest prediction accuracy at 85.12% and the highest AUC at .901. In addition, other factors were observed to have an impact on outcomes as well.*

## Introduction

Heart disease is the leading cause of death in the United States. About half of Americans have at least one of the three most common risk factors for heart disease: high blood pressure, high low-density lipoprotein (LDL) cholesterol, and smoking. Other factors that contribute to cardiovascular health are physical activity, diet, weight, and glucose control. By 2035, it is estimated that almost half of adults in the United States will have some form of heart disease, with the cost exceeding \$1.1 trillion.<sup>1</sup> Around 720,000 Americans will have their first hospitalization due to acute myocardial infarction (AMI) or coronary heart disease, and around 335,000 will have a recurrent event, in 2018.<sup>1</sup> It was suggested that one in every 7.4 Americans will die of an AMI.<sup>1</sup> Moreover, 170,000 AMI incidences, out of approximately 805,000 every year, are silent or without the classic symptoms such as chest pain, shortness of breath, and digestive issues.<sup>1</sup> Most treatment plans involve lifestyle modification and medications that require a high level of patient participation, compliance, and agreement with the treatment plan. Non-compliance has been studied and the results remain consistent that compliance is key to reduced cost-benefit, lower readmission rates, and reduced mortality.<sup>2-4</sup> An accurate prediction model for AMI mortality can inform the design of treatment plans and promise to improve patient compliance with the plans, thereby reducing costs, readmission, and mortality rates. In addition, it can set realistic expectations for medical staff, patients, and families at a time when emotion and judgment conflict. An acute medical illness that requires hospitalization often results in a high mortality rate during the year after discharge, especially for elderly patients. This is due to progressive physical decline.<sup>5</sup> Considering severely ill patients, who are usually prone to unnecessary interventions, communication about end-of-life planning (EOLp) becomes critical to the hospitals. A timely and reasonable discussion on EOLp is possible when there is an accurate estimation of patient's risk of death.<sup>6</sup>

Currently there are a variety of score-based risk assessment guidelines and systems associated with mortality prediction for outcomes of AMI and other heart diseases. CADILLAC (Controlled Abciximab and Device Investigation to Lower Late Angioplasty Complications) risk score is used to predict mortality for patients that suffered from an ST-segment-elevation myocardial infarction (STEMI) with percutaneous coronary intervention (PCI) as the primary treatment.<sup>7</sup> CRUSADE (Can Rapid risk stratification of Unstable angina patients Suppress ADverse outcomes with Early implementation of the ACC/AHA guidelines) is a risk assessment scoring system that measures bleeding risk in patients with non-ST-segment-elevation myocardial infarction (NSTEMI).<sup>8</sup> Global Registry of Acute Coronary Events (GRACE) is used to predict in-hospital and six-month mortality risk following an AMI.<sup>7</sup> Korea Acute Myocardial Infarction Registry is a prediction model for long term outcomes related to NSTEMI.<sup>9</sup> The Ontario mortality prediction scores can be used to predict 30-day and one-year mortality rates

following an AMI.<sup>10</sup> The Primary Angioplasty in Myocardial Infarction (PAMI) mortality risk score, which is similar to CADILLAC but uses different variables, is specifically used when primary PCI was the treatment of choice for patients who suffered from a STEMI.<sup>11</sup> The Soroka Acute Myocardial Infarction (SAMI) score predicts one- and five- year mortality following an AMI.<sup>12</sup> The Seattle Post Myocardial Infarction Model (SPIM) can predict survival following an AMI with left ventricular dysfunction.<sup>13</sup> Thrombolytic In Myocardial Infarction (TIMI) risk score is to predict in-hospital and one-year mortality in STEMI patients.<sup>7</sup> The Zwolle risk score is used for early discharge of low-risk patients following primary PCI for patients that suffered from a STEMI.<sup>14</sup>

There is not one go-to risk scale for all AMI patients, since it depends on the specific situation of the patient. In addition, different kinds of information are required for different risk scores. Though calculators do exist for some of these scales, the required information may not always be readily available. For example, Killip class is not readily available in the electronic health records (EHR), but is needed for calculating a majority of aforementioned scores. Killip class, is a major factor considered in many scores, where classes are based on physical examination of the patient to gauge the level of heart failure.<sup>15</sup> Other factors used in these different scoring systems include demographic information such as age, gender, weight, ethnics; chart values such as heart rate and systolic blood pressure; lab values such as hematocrit, creatinine clearance, serum and creatinine; test results such as echocardiography results, TIMI flow rate, and left ventricular ejection fraction; type, timing, and location of AMI; treatment decision; cardiac health information including presence of triple vessel disease, cerebrovascular disease, pulmonary edema, cardiac dysrhythmias, angina, hypertension; and information regarding comorbidities and overall health such as diabetes status, anemia, renal health, cancer status, smoking status, and obesity.<sup>7-14</sup>

With the wide adoption of EHR systems in the US, the rich EHR data have been recognized as a promising resource for comparative effectiveness research, outcomes research, epidemiology, drug surveillance, and public health research.<sup>16</sup> On the other hand, machine learning has become a popular approach for building models using EHR data to predict mortality rates, as well as a variety of other disease outcomes such as progression to Type 2 diabetes,<sup>17</sup> intensive care readmission,<sup>18</sup> and development of Alzheimer's disease.<sup>19</sup> There are other examples of predicting outcomes for radiation therapy<sup>20</sup> and acute ischemic stroke.<sup>21</sup> Lately, multiple studies have used machine learning and EHR data to predict mortality for patients diagnosed with diabetes,<sup>22</sup> breast cancer,<sup>23</sup> burn injury,<sup>24</sup> and sepsis or septic shock.<sup>25-26</sup> The information availability and ease of access to information within EHR has allowed for this newer technology to move to the forefront. When previous predictive models were built, they were done so when paper files were still the standard. Since EHRs have now become the *de facto* data source for health-related research, it follows that utilizing machine learning to build prediction models would also become the new standard.

This is especially true, in cardiology, as practitioners become required to utilize large amounts of data from different sources and practice more personalized medicine instead of a one-size fits all approach.<sup>27,28</sup> In addition, the complexity of these patients can create an environment well suited for machine learning techniques. Recently, a few studies have used machine learning to predict cardiovascular issues. Goldstein, Navar and Carter<sup>29</sup> worked specifically in the pros and cons of using machine learning versus regression models in cardiovascular risk prediction. They studied 1,944 patients with AMI admitted through the emergency room and utilized lab tests, demographics, and comorbidity information. Wallert et al.<sup>30</sup> compared four different machine learning algorithms on 51,943 patients to predict two-year survival following an AMI. They evaluated boosted C5.0 trees/rule sets, Random Forests, Support Vector Machines (SVM), and Logistic Regression. They utilized 39 different predictors, such as established risk indicators, survival factors, demographics, symptoms, echocardiography results, lab values, reperfusion, and pharmacology. Mansoor et al.<sup>31</sup> studied in-hospital mortality in females with STEMI, using Logistic Regression and Random Forest algorithms. Their study consisted of 12,047 patients and utilized a different number of variables, including demographic information, comorbidities, treatment, health status, zip code, primary payer, admission information, and hospital characteristics. Miao et al.<sup>32</sup> predicted one-year mortality in cardiac arrhythmias. They studied 10,488 patients from MIMIC-II using 40 variables and Random Survival Forest classification. Their variables included demographics, clinical variables, lab values, and pharmacology. Deep learning and artificial neural networks, have recently been widely used in prediction tasks relating to healthcare due to their superior prediction accuracy.<sup>33,34</sup> Auto-encoders have successfully been utilized to predict diagnosis. Recurrent neural networks (RNN) have been applied to handle temporal sequence of events.<sup>33</sup>

The current study built and evaluated various machine learning models to predict one-year mortality in patients that were diagnosed with an AMI or post myocardial syndrome (PMS). One-year mortality was selected for this study as a starting point because it would allow for comparison to other studies and consider that there are patients that had multiple AMI admissions within a two-year period. In our study, the results of the best performing machine learning model are compared to a deep feedforward neural network (Deep FNN) with back propagation. There are some

overlapping characteristics compared to previous cardiac associated prediction models and some differences. For example, this study does not differentiate between NSTEMI and STEMI as many previous models do. In addition, this study does not use any interpreted data such as echocardiography, location of AMI, or Killip class. The goal is to build prediction models using different machine learning algorithms and datasets to determine which classification model(s) performs the best. As shown by Johnson et al. there are pros and cons in each of the algorithms.<sup>27</sup> Based on the results from many recent studies, it is proved that machine learning models are more accurate in mortality prediction tasks compared to score-based systems. Contrary to score-based guidelines, which require physicians to manually calculate scores, machine learning-based models for predicting disease outcomes are based on complex algorithms with powerful analysis capability on large datasets.<sup>32</sup>

## **Methods**

We obtained patient data from the Medical Information Mart for Intensive Care III (MIMIC-III) database.<sup>36</sup> MIMIC-III is a freely available database developed by the MIT Lab for Computational Physiology. It integrates deidentified, comprehensive clinical data of patients admitted to the Beth Israel Deaconess Medical Center in Boston, Massachusetts. It is released as a relational database comprised of 26 data tables of which 7 were used in this study. This database covers over 58,000 hospital admissions of approximately 40,000 patients spanning an eleven-year period.<sup>36</sup> Waikato Environment for Knowledge Analysis (WEKA) is a freely available Java based software developed at the University of Waikato, New Zealand. It provides user-friendly and easy-to-use implementations of many machine learning algorithms.<sup>37</sup> WEKA was used in this study to explore the performance of various machine learning algorithms on the obtained datasets. In WEKA, we used the default configuration of each algorithms with 10-fold-cross-validation for model validation. TensorFlow is an open-source machine learning system released by Google Brain.<sup>38</sup> It is used in this study to implement a Deep FNN, utilizing packages from SciPy,<sup>39</sup> including Scikit-learn,<sup>40</sup> NumPy<sup>41</sup> and Pandas<sup>42</sup> in Python,<sup>43</sup> on a system with 2.2 GHz Intel Core i7 CPU. The implemented Deep FNN in this study has 2 hidden layers and 1 prediction layer for the classification task, using softmax cross entropy with logits for cost estimation, and gradient descent algorithm to optimize the cost. The size of each hidden layer was set as twice the input size. Rectified linear unit (ReLU) was used as the linear activation function. 10-fold-cross-validation was considered for model evaluation since data is not very large.<sup>44</sup>

## **Data and Variable Selection**

Data selection was initially based on features used in other similar studies and previous risk scales, and then further refined and limited by the data available in MIMIC-III. Interpreted data such as echocardiography and Killip class, certain specific comorbidity diagnoses, and pharmacology were not readily available in MIMIC-III and excluded because of that. Features related to admission, demographic, treatment, and comorbidity, were considered in this study. Regarding lab values, those evaluating both long- and short-term overall health, were taken into consideration. In addition, kidney and/or liver function have proven to be predictors of mortality,<sup>7-9,13,22,25,26</sup> so values related to these were included. Lastly, since preferred cardiac markers changed over the eleven-year period, a variety of them were included as well. Any data value with multiple entries such as lab values were averaged and the average value was used. In this study, we created different real-world datasets (i.e., admission information, demographics, treatment information, diagnostic information, and lab & chart values) from MIMIC-III, and investigated which set of features would contribute to the most reliable prediction models. In addition, a combination of all datasets was created to see if it improves model performance. Details regarding the variables employed in this study are given in Table 1.

Initially, there were 5037 subjects that met the criteria to be included in this study. These were patients who had an ICD-9 code of 410.0-411.0 (used for claims with a date of service on or before September 30, 2015) in MIMIC-III. Upon further review, it was determined that these subjects accounted for a total of 7590 admissions over an eleven-year period, out of which only 5436 admissions had a diagnosis of AMI or PMS, because there were patients that had multiple admissions for AMI or PMS while others had multiple admissions with only one being for AMI or PMS. It was decided to look at the data based on admission, rather than the individuals, since there were cases in which an individual survived a year in one admission, but did not survive a year in another admission. We treated each admission as a separate instance because each admission presented set of circumstances. Though certain demographic features would remain the same if the same patient was readmitted, all of the other factors could be different.

**Table 1.** Variables included in each dataset

Dataset	Variables
Admission Information	Total days of admission, month of admission, discharge location, whether or not initial ER diagnosis was AMI or rule out AMI
Demographics	Age at admission, gender, religion, ethnicity, marital status
Treatment Information	Cardiac catheterization, cardiac defibrillator and heart assist anomaly, cardiac defibrillator implant with cardiac catheterization, cardiac defibrillator implant without cardiac catheterization, cardiac valve and other major cardiothoracic procedures with cardiac catheterization, cardiac valve and other major cardiothoracic procedures without cardiac catheterization, cardiac valve procedures with cardiac catheterization, cardiac valve procedures without cardiac catheterization, coronary bypass with cardiac catheterization, coronary bypass with cardiac catheterization or percutaneous cardiac procedure, coronary bypass with PTCA, coronary bypass without cardiac catheterization, coronary bypass without cardiac catheterization or percutaneous cardiac procedure, other cardiac pacemaker implantation, other major cardiovascular procedures, other permanent cardiac pacemaker implant or PTCA with coronary artery stent implant, percutaneous cardiac procedure with drug-eluting stent, percutaneous cardiac procedure with non-drug-eluting stent, percutaneous cardiac procedure without coronary artery stent, percutaneous cardiovascular procedure, and permanent cardiac pacemaker implant
Diagnostic Information	Cancer, endocrinology, gastroenterology, genitourinary, hematological disorder, infection, liver or kidney issues, neurological disorder, orthopaedic disorder, other cardiovascular disease, other comorbidities (DRG codes), respiratory disorder, and toxicity issues
Lab & Chart Values	Average lab values included a lipid profile (cholesterol ratio, LDL cholesterol, HDL cholesterol, total cholesterol, and triglycerides), liver and kidney function tests (alanine transaminase, aspartate transaminase, alkaline phosphatase, albumin, bilirubin, blood urea nitrogen, creatinine, gamma-glutamyltransferase, L-lactate dehydrogenase, and total protein), cardiac function tests (N-terminal prohormone of B-type natriuretic peptide, C-reactive protein, creatine kinase, creatine phosphokinase-MB, cortisol, homocysteine, troponin I and troponin T), and electrolytes (bicarbonate, calcium, chloride, potassium, and sodium), glucose, hematocrit, hemoglobin, and white blood count; Average chart values include diastolic and systolic blood pressure, heart rate, and respiratory rate
Combined	All of the Above

### ***Data Preprocessing***

From the initial list of hospital admissions, there were admissions that had multiple diagnostic codes for AMI or AMI and PMS. Duplicates were removed to ensure that there was one admission per actual instance. Duplicate entries also existed in the treatment data and the comorbidities. In both cases, it was either due to multiple codes for the same treatment (due to multiple entries in different parts of the EHR) or due to multiple treatments or comorbidities for the same admission. In the latter case, both items were counted. For the demographics, one data item that needed to be corrected was the age and death age of anyone over the age of 89. These fields were changed in MIMIC-III to mask the age per HIPAA rules. The ages were changed back by subtracting 211 years from the values presented. The lab and chart values were reviewed to remove or correct entry errors. There were a few instances in the blood pressures where the numbers had been entered in reverse order. It was also realized that a '0' value in the lab values occur when there is a 'see comment' entered at any time with that lab value, thus all 0 values were removed from the lab values. Any lab or chart values that seemed biologically invalid were also removed. Outliers in lab values were removed. The first and third quartile were calculated and any values that fell outside of the first quartile minus 1.5 times the interquartile range or the third quartile plus 1.5 times the interquartile range were removed. Classifications were run with both the outliers in place and without the outliers. Results were better with the outliers. This was potentially due to more abnormal lab results.

### ***Modeling and Evaluation***

The outcome of the prediction models is the mortality that occurred within one year after admission. Admissions in which the patient died within a year were considered as positive instances and those in which the patient did not die

within a year were considered as negative instances. The classification algorithms that were run for each dataset include AdaBoost, Attribute Selected Classifier, Bayes Net, Classification Via Regression, Decision Stump, Decision Table, Iterative Classifier Optimizer, Hoeffding Tree, J48, JRip, Logistic, Logistic Model Trees (LMT), LogitBoost, NaiveBayes, OneR, PART, Random Forest, Random SubSpace, Random Tree, Randomizable Filtered Classifier, REP Tree, Stochastic Gradient Descent (SGD), Simple Logistic, Sequential Minimum Optimization (SMO), and Voted Perceptron. As some algorithms performed poorly, not all results are reported here. The classification algorithms were compared to determine which models yielded the best results based on overall accuracy, precision, recall, and area under the curve (AUC) of the receiver operator curve (SROC). Precision measures positive versus negative results, while recall measures the percent of positive instances identified out of the total amount of positive instances in the dataset. F-measure is the harmonic mean of precision and recall. Machine learning algorithms were first applied on each dataset individually, and then on the combined dataset with predicted outcome as one-year mortality. Deep FNN was applied on the combined dataset and compared with the best performing shallow machine learning algorithm.

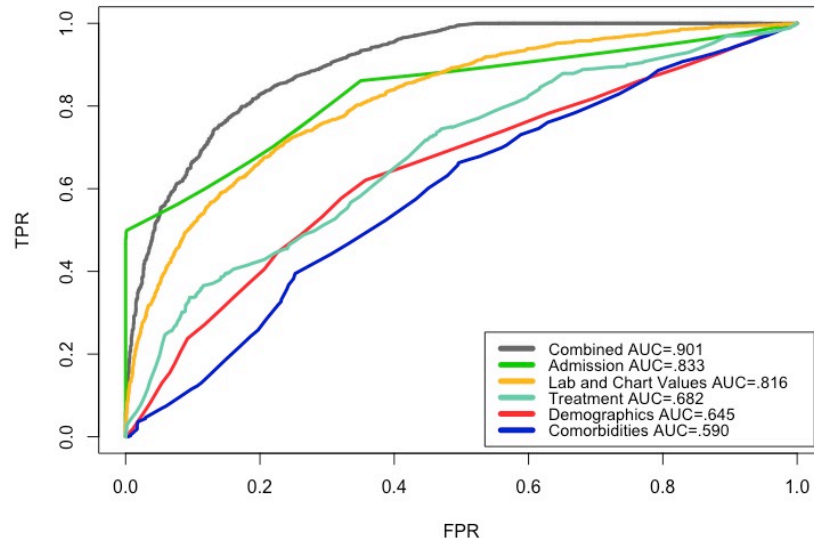
## Results

Of the 5436 admissions included in this study, 1629 (30%) in which the patient died within one year of admission are positive instances and the remaining 3807 (70%) in which the patient survived past the one-year mark are classified as negative instances. Summary statistics for these instances can be found in Table 2.

**Table 2.** Summary statistics of the cohort of patients

Characteristics		Number of Instances	Number and Percentage of Positive Instances		Number and Percentage of Negative Instances		P-value
Gender	Male	3308	877	26.5%	2431	73.5%	<.00001
	Female	2128	752	35.3%	1376	64.7%	<.00001
Age when first admitted	Under 30	13	4	30.8%	9	69.2%	.949588
	30 to 49.9	420	58	13.8%	362	86.2%	<.00001
	50 to 59.9	784	92	11.7%	692	88.3%	<.00001
	60 to 69.9	1209	273	22.6%	936	77.4%	<.00001
	70 to 79.9	1428	465	32.6%	963	67.4%	.012632
	80 to 90	1524	708	46.5%	816	53.5%	<.00001
	Over 90	58	29	50.0%	29	50.0%	.000813
Ethnicity	Asian	87	23	26.4%	64	73.6%	.468718
	Black	291	103	35.4%	188	64.6%	.037736
	Hispanic/Latino	97	19	19.6%	78	80.4%	.024348
	Other	108	25	23.1%	83	76.9%	.118186
	White	3774	1144	30.3%	2630	69.7%	.401699
Initial ER Diagnosis MI?	Yes	2061	470	22.8%	1591	77.2%	<.00001
	No	3375	1159	34.3%	2216	65.7%	<.00001
With Comorbidities	Total	1614	798	49.4%	816	50.6%	<.00001

Figure 1 shows the ROC for the datasets based on the classification that produced the highest accuracy for each dataset. The best performing models, in terms of AUC, were generated by applying Logistic Model Trees (LMT) and Simple Logistic algorithms on the combined dataset. Applying J48 and Logistic Regression algorithms, on the admission dataset, and the lab & chart values dataset, respectively, resulted in models with .80 AUC. Models that resulted from applying Random Forest, Attribute Select Classifier, and Decision Table algorithms on the treatment, demographics, and comorbidities datasets, respectively, did not perform well, all scoring an AUC under .70.



**Figure 1.** ROC of shallow models based on the highest accuracy with each dataset.

Table 3 shows the results based on highest accuracy for each dataset, along with the shallow classification algorithm used. The model based on the combined dataset showed the highest accuracy of 85.12% applying LMT and Simple Logistic algorithms. The model developed from J48 algorithm on the admission dataset was closely behind at 84.88% accuracy. These two models were not drastically different in other measurements as well. The model developed from J48 algorithm on the admission dataset, resulted in the highest precision, meaning it returned the most positive versus negative instances. The recall rates were low across the board meaning that none of these models were especially great at returning a high percentage of positive instances. In the case of the comorbidities dataset, modeled using Decision Table algorithm, the sensitivity is slightly higher than the specificity meaning that this dataset and model were slightly better at detecting the positive instances than the negative instances.

**Table 3.** Summary of WEKA results based on the highest accuracy

Dataset	Best Performing Classification Algorithm	Highest Percent Correctly Classified	AUC	Precision	Recall	F-Measure
Combined	LMT and Simple Logistic	<b>85.12%</b>	<b>0.901</b>	0.867	0.594	0.705
Admission	J48	84.88%	0.833	<b>0.993</b>	0.499	0.664
Treatment	Random Forest	83.73%	0.682	0.789	0.028	0.055
Lab & Chart Values	Logistic	78.46%	0.816	0.708	0.478	0.571
Demographics	Attribute Selected Classifier	70.70%	0.645	0.524	0.238	0.327
Comorbidities	Decision Table	58.18%	0.590	0.566	0.660	0.610

Table 4 shows the results of applying various shallow machine learning algorithms in WEKA and the deep learning algorithm implemented in TensorFlow, on the combined dataset. For the most part, the models based on machine learning algorithms performed similarly well with a few exceptions. LMT and Simple Logistic showed the highest accuracy of 85.12%. They also showed the highest AUC at .901. Decision Stump had the highest precision at 1 meaning 100% of instances classified as positive were indeed positive. None of the machine learning based models showed recall rates that would be considered stellar, the highest rate being .744 produced by Bayes Net. This shows that none of these models were great at detecting a high percentage of positive instances in the dataset. This is where we examined Deep FNN. The model using Deep FNN, drastically outperformed all machine learning algorithms in terms of recall at .820, while maintaining precision at 0.831 and F-measure at .813. This means the Deep FNN

model outperformed the shallow machine learning algorithms in identifying a higher percentage of positive instances. Since its AUC is .751, this model does not have a great sensitivity or specificity and its ROC curve would fall somewhere between the ones for lab & chart values and treatment in Figure 1.

**Table 4.** Comparison of the performance of different models using the combined dataset

Classification Algorithms	Percent Correctly Classified	AUC	Precision	Recall	F-Measure
AdaBoost	84.28%	0.849	0.981	0.484	0.649
Bayes Net	81.55%	0.889	0.674	<b>0.744</b>	0.707
Decision Stump	84.31%	0.730	<b>1</b>	0.476	0.645
Decision Table	84.29%	0.865	0.923	0.519	0.665
J48 <sup>a</sup>	84.64%	0.843	0.97	0.503	0.662
JRip	84.27%	0.737	0.985	0.483	0.648
LMT	<b>85.12%</b>	<b>0.901</b>	0.867	0.594	0.705
Logistic	85.06%	0.899	0.835	0.626	<b>0.715</b>
Naïve Bayes	82.63%	0.862	0.768	0.603	0.675
OneR	84.88%	0.749	0.993	0.499	0.664
PART	84.58%	0.869	0.973	0.500	0.660
Random Forest	84.60%	0.893	0.949	0.514	0.667
Random Tree	76.77%	0.776	0.628	0.552	0.588
REP Tree	83.61%	0.845	0.884	0.522	0.656
SGD	84.90%	0.765	0.904	0.555	0.688
Simple Logistic	<b>85.12%</b>	<b>0.901</b>	0.867	0.594	0.705
SMO <sup>b</sup>	84.84%	0.751	0.972	0.509	0.668
Voted Perceptron	70.70%	0.519	0.773	0.031	0.06
Deep FNN	82.02%	0.751	0.831	<b>0.820</b>	<b>0.813</b>

<sup>a</sup>: The implementation of C4.5 decision tree in WEKA

<sup>b</sup>: WEKA's implementation of John Platt's sequential minimal optimization algorithm for training a support vector classifier.

Table 5 shows a comparison of the results of this study to similar cardiovascular studies. As shown, this study resulted in the highest AUC. It is interesting that no study utilized the same machine learning algorithm to obtain the highest AUC in their best model. This reflects a common understanding in data science that there is no universally applicable algorithm that outperforms all the other algorithms all the time.

**Table 5.** Comparison of AUC to other Cardiovascular Studies

Study	Study Topic	Classification Algorithm	Highest AUC	# of Variables
Barrett et al. (This paper)	1-year mortality AMI	Simple Logistic/LMT	0.901	79
Goldstein et al. <sup>29</sup>	Cardiovascular Risk	Boosting Tree Based	0.763	43
Mansoor et al. <sup>31</sup>	In-hospital mortality STEMI	Logistic Regression	0.840	11 (with backward elimination)
Miao et al. <sup>32</sup>	1-year mortality arrhythmia	Random Survival Forest	0.810	40
Wallert et al. <sup>30</sup>	2-year survival AMI	Support Vector Machine	0.845	39

## Discussion

There are many factors that can affect mortality rates following a myocardial infarction. Figuring out a way to utilize the information regarding these factors will assist in accurately predicting possible outcomes. As seen from this dataset, there is not one specific factor that provides the needed predictability information, while being able to

include all relevant criteria leads to improved predictions. The various factors including comorbidities, treatment options, patient history, short-term and long-term overall health, and demographics can all impact outcomes in cardiac health, which makes it important to consider all of them when predicting cardiac outcomes.

One interesting finding in this study, was the importance of initial diagnosis, as a “myocardial infarction” or “rule out myocardial infarction”. Of the 5436 admissions, 3375 of these were not tagged as either of these as the primary reason for admission. Of these, the percentage in the positive instances was higher than the percentage in the negative instances (34.3% versus 30%). Chi-square statistics was applied to compare the expected outcome with the observed outcome. The p-value was  $< .00001$  and the result was significant at  $p < .05$ . This indicates the importance of recognizing myocardial infarction initially so that appropriate treatment can be used as soon as possible. Another possible explanation for this is when patient was hospitalized for a comorbidity, a significant number of AMI occurred. As there is no way to time stamp when the AMI occurred for patients who had AMI before admission, it is not possible to differentiate between those that occurred prior to admission and after admission. If that is the case, it indicates that an AMI that occurs during an admission may lead to a higher mortality rate.

Two elements of the demographic dataset also appeared to be statistically significant. First off was gender. The percentage of female was higher than the percentage of male in the positive class. Chi-square statistics, considering p-value  $< .00001$  and significance at  $p < .05$ , indicate that gender does play a role in one-year mortality following a myocardial infarction. This may be attributed to differing symptoms in male and female patients. Because of the difference of symptoms in males and females, perceptual differences about these symptoms can play a role. This may cause female to seek treatment less proactively than male due to downplaying the symptoms or attributing the cause of the symptoms to something else. Another factor could be age, the average age of male in the positive class was 75.08 versus the average age of female in the same class at 77.65. In the negative class the gap is greater with males at an average of 66.02 and females averaging 72.09. The second element was ethnicity. Those that identified as White and Black had a higher percentage in the positive class than the overall cohort did. Those that identified as Asian and Hispanic/Latino had a lower percentage in the positive class than the overall cohort did.

Admissions in the comorbidity dataset had a higher percentage in the positive class than the overall cohort did. Statistical analysis indicates that if patients have a comorbidity on the same admission, they are more prone to die within a year of a myocardial infarction than someone that does not have any treatment for a comorbidity on the same admission.

### ***Limitations and Future Opportunities***

A potential limitation of this study was the imbalanced dataset of positive and negative cases (30% vs. 70%). Future studies could be conducted in which there is closer to a 50/50 balance using some random sampling strategies. Another limitation was the availability of some lab values in MIMIC-III. Some patients did not have all lab values collected. In addition, some of the desired lab and chart values could not be easily obtained. Potentially, a different database would have different access to lab values and chart values allowing for more robust records to be used. Another limitation to this study is that none of the machine learning models and Deep FNN model were compared against the standard predictive scores to determine which were more effective and efficient. This would be an interesting further study if the variables needed for calculating the risk scores (e.g., Ontario, SAMI) are available in EHR. One last limitation to this study is the lack of information related to cause of death. We are unable to tell if cause of death was related to AMI, cardiac health, a comorbidity, or a completely unconnected occurrence. This additional information may provide helpful insight into improving mortality rates.

Other opportunities for additional studies include those related to mortality rates as they relate to correct identification of AMI in the ER versus rates of those that have an AMI while already admitted for an unrelated comorbidity. Based on the data system in MIMIC-III, there is no specific timeline associated with when the AMI occurred so it is not possible to note if the diagnosis was associated with the initial admission or occurred during an unrelated ICU stay. In addition, studying the role of gender feature in mortality related to AMI would help. This study noted that the mortality rate in female was higher than male, thus discovering the reasons would be helpful to lowering those rates. Further studies could also use differing timelines related to mortality rates, for example 30 days, six months, and two-years as done in other comparable studies. The deep learning architecture that was used in this study was a fully connected feedforward neural network. The non-linear activation function considered for hidden layers is usually logistic sigmoid, tanh, or ReLU. We chose ReLU which yielded the best performance results. During the training, the cross-entropy loss was optimized between the target output and prediction output. Another future study could investigate other deep learning algorithms’ performances on mortality prediction, like convolutional neural network (CNN), recurrent neural network (RNN), or long short-term memory (LSTM),



incorporating more dense features (e.g., longitudinal data) and records, including clinical notes and temporal information related to patients.

## Conclusions

This study examined multiple machine learning algorithms on multiple datasets obtained from MIMIC-III, to predict one-year mortality rate in patients diagnosed with AMI or PMS. In addition, the results of the best performing machine learning model on the Combined dataset, was compared to a Deep FNN model. Highest prediction accuracy and AUC was achieved by LMT and Simple Logistic algorithms, while Deep FNN had less accuracy and precision but higher recall. Correct diagnosis and treatment of AMI can have an effect on mortality as this study has shown. This emphasizes the importance of a few things. One, there needs to be more training to the public on reducing the risk factors associated with heart disease and AMI. Two, in order to lower the chances of a comorbidity, there needs to be more incentives to the public to maintain overall health when risk factors of heart disease exist. Three, there also needs to be more training to the public on recognizing the signs and symptoms of AMI, especially in women. Four, the improved predictability obtained by using machine learning can help at-risk patients strive for compliance to treatment plans to improve mortality risk.

## References

1. Benjamin EJ, Virani SS, Callaway CW, et al. Heart disease and stroke statistics—2018 update: a report from the American Heart Association. *Circulation*. 2018 Mar 20;137(12):e67–492.
2. Choudhry NK, Winkelmayr WC. Medication adherence after myocardial infarction: a long way left to go. *J Gen Intern Med*. 2008 Feb;23(2):216–8.
3. Arif H, Aijaz B, Islam M, Aftab U, Kumar S, Shafqat S. Drug compliance after stroke and myocardial infarction: a comparative study. *Neurol India*. 2007 Jun;55(2):130–5.
4. Kolandaivelu K, Leiden BB, O’Gara PT, Bhatt DL. Non-adherence to cardiovascular medications. *Eur Heart J*. 2014 Dec 7;35(46):3267–76.
5. Walter LC, Brand RJ, et al. Development and validation of a prognostic index for 1-year mortality in older adults after hospitalization. *JAMA*. 2001 Jun 20;285(23):2987–94.
6. Sahni N, Simon G, Arora R. Development and validation of machine learning models for prediction of 1-year mortality utilizing electronic medical record data available at the end of hospitalization in multicondition patients: a proof-of-concept study. *J Gen Intern Med*. 2018 Jun;33(6):921–8.
7. Scruth EA, Page K, Cheng E, Campbell M, Worrall-Carter L. Risk determination after an acute myocardial infarction: review of 3 clinical risk prediction tools. *Clin Nurse Spec*. 2012;26(1):35–41.
8. Subherwal S, et al. Baseline risk of major bleeding in non-ST-segment-elevation myocardial infarction: the CRUSADE (Can Rapid risk stratification of Unstable angina patients Suppress ADverse outcomes with Early implementation of the ACC/AHA guidelines) bleeding score. *Circulation*. 2009 Apr 14;119(14):1873–82.
9. Kim HK, Jeong MH, et al. A new risk score system for the assessment of clinical outcomes in patients with non-ST-segment elevation myocardial infarction. *Int J Cardiol*. 2010 Dec;145(3):450–4.
10. Tu JV, Austin PC, Walld R, Roos L, Agras J, McDonald KM. Development and validation of the Ontario acute myocardial infarction mortality prediction rules. *J Am Coll Cardiol*. 2001 Mar;37(4):992–7.
11. Addala S, Grines CL, et al. Predicting mortality in patients with ST-elevation myocardial infarction treated with primary percutaneous coronary intervention (PAMI risk score). *Am J Cardiol*. 2004 Mar;93(5):629–32.
12. Plakht Y, Shiyovich A, Weitzman S, Fraser D, Zahger D, Gilutz H. A new risk score predicting 1- and 5-year mortality following acute myocardial infarction. *Int J Cardiol*. 2012 Jan;154(2):173–9.
13. Ketchum ES, Dickstein K, Kjekshus J, Pitt B, Wong MF, Linker DT, et al. The Seattle Post Myocardial Infarction Model (SPIM): prediction of mortality after acute myocardial infarction with left ventricular dysfunction. *Eur Heart J Acute Cardiovasc Care*. 2014 Mar;3(1):46–55.
14. Tralhão A, Ferreira AM, Madeira S, Borges Santos M, Castro M, Rosário I, et al. Applicability of the Zwolle risk score for safe early discharge after primary percutaneous coronary intervention in ST-segment elevation myocardial infarction. *Rev Port Cardiol*. 2015 Sep;34(9):535–41.
15. Killip Classification Topic Review [Internet]. 2018. Available from: <https://www.healio.com/cardiology/learn-the-heart/cardiology-review/topic-reviews/killip-classification>.
16. Weiskopf NG, Hripcsak G, Swaminathan S, Weng C. Defining and measuring completeness of electronic health records for secondary use. *J Biomed Inform*. 2013 Oct;46(5).

17. Anderson JP, Parikh JR, Shenfeld DK, Ivanov V, Marks C, Church BW, et al. Reverse engineering and evaluation of prediction models for progression to Type 2 diabetes: an application of machine learning using electronic health records. *J Diabetes Sci Technol*. 2016 Jan;10(1):6–18.
18. Rojas JC, Carey KA, Edelson DP, Venable LR, Howell MD, Churpek MM. Predicting intensive care unit readmission with machine learning using electronic health record data. *Ann Am Thorac Soc*. 2018 Jul;15(7):846–53.
19. Zhu F, Panwar B, Dodge HH, et al. COMPASS: A computational model to predict changes in MMSE scores 24-months after initial assessment of Alzheimer’s disease. *Sci Rep*. 2016 Dec;6(1).
20. Kang J, Schwartz R, Flickinger J, Beriwal S. Machine learning approaches for predicting radiation therapy outcomes: a clinician’s perspective. *Int J Radiat Oncol*. 2015 Dec;93(5):1127–35.
21. Asadi H, Dowling R, Yan B, Mitchell P. Machine learning for outcome prediction of acute ischemic stroke post intra-arterial therapy. Gómez S, editor. *PLoS ONE*. 2014 Feb 10;9(2):e88225.
22. Anand RS, Stey P, et al. Predicting mortality in diabetic ICU patients using machine learning and severity indices. *AMIA Jt Summits Transl Sci Proc AMIA Jt Summits Transl Sci*. 2018;2017:310–9.
23. Montazeri M, Montazeri M, Montazeri M, Beigzadeh A. Machine learning models in breast cancer survival prediction. *Technol Health Care*. 2016 Jan 27;24(1):31–42.
24. Stylianou N, Akbarov A, Kontopantelis E, Buchan I, Dunn KW. Mortality risk prediction in burn injury: comparison of logistic regression with machine learning approaches. *Burns*. 2015 Aug;41(5):925–34.
25. Carrara M, Baselli G, Ferrario M. Mortality prediction model of septic shock patients based on routinely recorded data. *Comput Math Methods Med*. 2015;2015:1–7.
26. Zhang Z, Hong Y. Development of a novel score for the prediction of hospital mortality in patients with severe sepsis: the use of electronic healthcare records with LASSO regression. *Oncotarget*. 2017 Jul 25;8(30):49637–45.
27. Johnson KW et al. Artificial intelligence in cardiology. *J Am Coll Cardiol*. 2018 Jun 12;71(23):2668–79.
28. Shameer K, Johnson KW, Glicksberg BS, Dudley JT, Sengupta PP. Machine learning in cardiovascular medicine: are we there yet? *Heart*. 2018 Jul 1;104(14):1156.
29. Goldstein BA, Navar AM, Carter RE. Moving beyond regression techniques in cardiovascular risk prediction: applying machine learning to address analytic challenges. *Eur Heart J*. 2016 Jul 19;ehw302.
30. Wallert J, Tomasoni M, Madison G, Held C. Predicting two-year survival versus non-survival after first myocardial infarction using machine learning and Swedish national register data. *BMC Med Inform Decis Mak*. 2017 Dec;17(1).
31. Mansoor H, Elgendy IY, et al. Risk prediction model for in-hospital mortality in women with ST-elevation myocardial infarction: a machine learning approach. *Heart Lung J Acute Crit Care*. 2017 Nov;46(6):405–11.
32. Miao F, Cai Y-P, Zhang Y-X, Li Y, Zhang Y-T. Risk prediction of one-year mortality in patients with cardiac arrhythmias using random survival forest. *Comput Math Methods Med*. 2015;2015.
33. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. *Npj Digit Med*. 2018 May 8;1(1):18.
34. Miotto R, et al. Deep learning for healthcare: review, opportunities and challenges. *Brief Bioinform*. 2017;
35. Allyn J, Allou N, Augustin P, et al. A comparison of a machine learning model with EuroSCORE II in predicting mortality after elective cardiac surgery: a decision curve analysis. *PLoS ONE*. 2017 Jan 6;12(1).
36. Johnson AEW, Pollard TJ, Shen L, Lehman LH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. *Sci Data*. 2016 May 24;3:160035.
37. Witten IH, Frank E, Hall MA. The WEKA workbench. Online appendix for “Data mining: practical machine learning tools and techniques.” 4th ed. Morgan Kaufmann; 2016.
38. TensorFlow [Internet]. TensorFlow. [cited 2018 Aug 8]. Available from: <https://www.tensorflow.org/>
39. SciPy.org [Internet]. SciPy.org. [cited 2018 Aug 7]. Available from: <https://www.scipy.org/>
40. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011 Nov;12:2825–2830.
41. Oliphant TE. Guide to NumPy [Internet]. 2006 [cited 2018 Aug 5]. Available from: <http://ns.ael.ru/ports/distfiles/numpybook.pdf>
42. McKinney W. Data structures for statistical computing in Python. In: *Proc of the 9th Python in Science Conf* [Internet]. [cited 2018 Aug 5]. p. 51–6. Available from: <https://pdfs.semanticscholar.org/f6da/c1c52d3b07c993fe52513b8964f86e8fe381.pdf>
43. Welcome to Python.org [Internet]. Python.org. [cited 2018 Aug 5]. Available from: <https://www.python.org/>
44. Che Z, Purushotham S, Khemani R, Liu Y. Interpretable deep models for ICU outcome prediction. *AMIA Annu Symp Proc*. 2017 Feb 10;2016:371–80.