Predicting Survival of Cardiac ICU Patients

Josue N Rivera, Nadia Khalil, Alvin Thamrin

December 15, 2019

1 Abstract

Patient survival and mortality has long been studied from the context of medical and non-medical factors both within and outside of hospital settings. Cardiac conditions are often the cause of admission to a hospital’s intensive care unit. Survival after an ICU stay with a cardiac diagnosis may vary depending on socioeconomic status and other social and demographic factors. This study aims to predict patient survival of an ICU stay on the basis of social and demographic factors among those who have frequently occurring cardiac diagnoses, namely congestive heart failure and atrial fibrillation. This study determines that factors such as gender, age, admission type, insurance type, marital status, diagnoses, and previous and current ICU unit admissions have predictive power in determining the survival of a patient. The proposed model utilizes neural networks and generates predictions with 88.29% accuracy, thereby confirming the predictive power of social and demographic factors.

2 Background and Introduction

A patient’s stay in an intensive care units (ICU) generates large volumes of real-time data. Surviving an ICU stay can be affected by a number of factors that are both medical (i.e. existing diagnoses) and non-medical (i.e. demographic, social, and environmental). Extant research has shown that non-medical factors do in fact have predictive power in determining which patients are likely to survive an ICU stay and which are not.

Cardiac-related diagnoses are among the most frequently occurring diagnoses in intensive care unit admissions. Many cases have poor prognoses that result in high mortality rates. Two of the common heart-related diagnoses that patients who are admitted to an ICU suffer from are congestive heart failure (CHF) and atrial fibrillation (AFIB). Computer-assisted decision support for diagnoses that are time sensitive such as CHF and AFIB could potentially aid doctors in making faster and more informed decisions that can improve the likelihood of patient survival. Smaller hospitals with generalized ICUs may provide a different level of care quality than will a larger hospital (perhaps in a more affluent area) with a specialized cardiac-ICU. This exploratory study aims to evaluate the predictive power of social factors in survival from two potentially fatal heart conditions during an ICU stay.

3 Related Work

Existing research has discovered relationships between social factors and survival after a cardiac-related event. Gerward et. al.’s (2006) research studied survival rate 28 days after hospital admission with a patient’s first myocardial infarction. Logistical regression and the Kaplan-Meier method resulted in the finding that survival rate is inversely related to socioeconomic satisfaction.

Farmer et. al. (2013)’s research examined predictors of mortality five to ten years after heart transplantation. Using a self-reported “quality of life” instrument and a sample from four different hospitals, it was found that education level and higher levels of socioeconomic satisfaction are predictive of improved survival rates. Furthermore, poorer survival rates were associated with being married, poor adherence to medical care, and the presence of blood-related disorders. Using statistical analysis software, the results were generated through the calculation of frequencies, means, Pearson correlation coefficients, and Cox proportional hazard modeling.

Using machine learning techniques is a newer approach towards modeling predictions for patient survival. Computer-assisted decision-making using machine learning for cardiac-related diagnoses was studied by Donal et. al. (2017). The study developed a bedside prediction score (RISK-E) to aid physicians and their assistants in decision making for cardiac surgery. The time sensitivity of the acute stages of a patient’s infective endocarditis diagnoses necessitates real-time decision making, but the model was not found sufficiently accurate to replace the expertise of an endocarditis team. Nanayakkara et. al. (2018) studied in-hospital mortality following cardiac arrest. Patient demographic, physiological, and biochemical data used for logistic regression and machine learning methods to develop a model with 80% accuracy.

4 Problem Definition

The objective of this study is to build a model to predict survival after an ICU stay among patients with cardiac diagnoses (AFIB and/or CFH) will survive their ICU stay. This study will base the predictive model on the following:

* Patient gender (GEN): male or female.
* Age (AGE): age when admitted to the ICU.
* Admission type (ADMIT): the nature of the admission (e.g. urgent, emergency, or elective).
* Insurance type (INS): government-provided, Medicare, Medicaid, or private insurance.
* Marital status (MAR): married, single, or widowed, among others.
* Diagnoses (DIAG): all other diagnoses excluding CHF and AFIB that the patient is diagnosed with upon admission to the ICU as well as diagnoses that arise during the patient’s stay.
* Previously administered (PREV): medical or surgical cardiac services that the patient has undergone in the past.
* Currently administered (CURR): medical or surgical cardiac services that the patient is currently undergoing.

5 Methodology

The framework that this study implements is original and is based on the available attributes in the dataset that is used to train the model. Using eight attributes to represent the independent variables, the model takes into account demographic, socioeconomic, and hospital stay-related information in order to predict the dependent variable, survival. Figure 1 shows the eight independent variables (GEN, AGE, ADMIT, INS, MAR, DIAG, PREV, and CURR) and their relationship with the dependent variable (survival/mortality). The grey dotted lines represent the effects of the independent variables on one another thereby indirectly affecting the dependent variable whereas the black solid lines represent the direct effect of each independent variable on the dependent variable.

A close up of a map

Description automatically generated

**Figure 1**: Relationships between the independent and dependent variables.

The steps for modeling the prediction is as follows and is also visualized in Figure 2:

1. Merge tables containing significant attributes
2. Normalize the data
3. Reformat for neural network
4. Neural network construction and training
5. Evaluation
6. Analysis

Merge tables

Normalize data

Reformat

for neural network

Construct and train

neural network

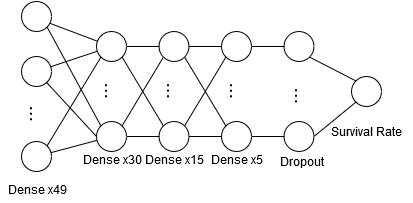
Evaluation

Analysis

**Figure 2**: Model steps.

The model that this study proposes to predict that survival of a patient with a cardiac condition is a neural network. The model consists of an input layer with 48 input neurons. These neurons represent the eight attributes his study uses to predict the survival of a patient during an ICU stay. Categorical data such as INS or MAR is converted into a series of neurons that represent each of the unique values of the data type. The neuron associated with a certain category is set to “1” while the other neurons are set to “0” when analyzing an entry in the database. For binary data, such as the GEN of a patient, a single neuron represents the information. A value of “1” refers to an extreme of the spectrum while a value of “-1” represents the opposite end. Following a similar method used for binary information, numerical data, such as AGE, is mapped to a range of values between -1 and +1 to represent the lower and upper ends, respectively. Numerical data is standardized into a z-score and is represented by a single neuron. One distinction that exists between the standardization and normalization of the various data types is the use of only zeros (0) over negative ones (-1) to represent categorical data that is not selected. Zeros (0) are used in this method to represent a neutral value. Unlike the other data types, categorical data does not imply any inherent ranking or order to the values. The proposed model concludes with a single output neuron that contains an inclusive continuous value between zero (0) and one (1). This value represents that probability of survival of a patient with CHF and/or AFIB in an ICU.

The architecture of the model that this study proposes is a multi-perceptron. A multiple perceptron is a class of feedforward neural network that takes some input and propagates the values through the network. In this type of model, the input of a layer is the previous layer in a sequence of computational layers. The structure of the proposed neural network consists of a single input layer with forty-eight neurons, three hidden dense layers in a decreasing number of neurons (thirty, fifteen, and five, respectively, a dropout layer with 20% dropout rate, and lastly, an output layer with a single neuron to represent the survival rate. The goal of the model is for the neural network to learn to recognize different complex patterns at different stages of the layers that can help provide a final survival rate. A diagram is presented in Figure 3 that showcases the structure of the model.



**Figure 3**: Model structure.

The inclusion of the dropout layer is due to an attempt to mitigate overfitting during training. Overfitting may occur due to disproportions in the distribution of the data that the model encounters. When this happens, the model recognizes the patterns that make the common data what it is, but the model will not learn to recognize the patterns that differentiate different data types. Due to the disproportion in the data distribution between those who survived and those who expire, overfitting needs to be mitigated. A dropout layer reduces overfitting by deactivating certain neurons during training and forcing other neurons to contribute to the classification of the data. The inclusion of a dropout layer is vital to building a model that is reliable. Furthermore, the activation function used at the different layers were selected in coherence with the values of the reformatted inputs to the model. The hidden layers consisted of neurons with a hyperbolic tangent activation function (tanh). The activation function results in values that range between negative one (-1) and positive one (+1) with symmetry around the origin. Due to the values that the function provides and the range of the derivative of the equation, the tanh activation function converges faster than other standard logistic function like sigmoid (LeCun, et. al., 2012). The output layer of the model consists of a single neuron with sigmoid activation function. Sigmoid activation functions are suited for proving normalized value in the range of continues values from zero (0) to one (1) (LeCun, et. al., 2012)

6 Experimental Setting

A database that accurately reflects the reality of an ICU stay is needed to build a model that can reliably predict the survival rate of a patient with a cardiac condition. The MIMIC-III (Medical Information Mart for Intensive Care III) is a large and comprehensive database comprising of deidentified health-related data associated with over sixty thousand admissions. The admissions are composed of entries in critical care units at the Beth Israel Deaconess Medical Center between 2001 and 2012. The database includes data on the demographics, vital signs, laboratory tests, services and other socioeconomic and health-related information of over forty thousand patients (Johnson, et. al., 2016). In the context of this study, patients that were admitted to the hospital and diagnosed with CHF and/or AFIB conditions became the focus of the research. Filtering for patients that were diagnosed with either one of these two conditions resulted in a total number 1,174 admissions entries. 1,557 of these entries were of admissions in which the patients survived the ICU stay. The remaining 217 entries were of admission in which the patient expired during the ICU stay. The distribution of the data concluded with a 7.11-to-1 ratio between patients who survived to patients who expired.

To train the model, the database is split into three sets: a training set, a validation set, and a testing set. After filtering for the dependent variables and joining various tables together, there is 1,774 admissions entries. Following the convention used to split a database for training, the database is initially split into two components. One set is the training set with eighty percent (80%) of all the data entries, and the other set is the testing set with twenty percent (20%). The testing set is used as a benchmark to test the efficiency and performance of the model. To measure the performance of the model during training, the training dataset is split once again. For the current split, the original training data is split into two components. One component consists of 66.67% of the data, and the other set consists of the remaining 33.33% The two new sets represent the final training and validation sets. The conclusion of all the splitting of the database resulted a distribution 53.6%, 26.4%, and 20% for the training, validation, and testing sets, respectively. The primary goal of using a validation set is to determine the performance of the model over different epochs. Epochs are the cycles of training that a model goes through. In the case of the proposed model, the total number of epochs is set to 200. The model is trained using backpropagation. Backpropagation works by calculating the loss of the model after making a prediction and comparing it to the correct output. The weights and parameters of the model are then adjusted to decrease the loss of the following predictions. In the case of the proposed model, the mean square error (MSE) function is used to calculate the loss. Adam optimization was used to perform the backpropagation on the model.

7 Experimental Results and Analysis

To measure the performance of the model, various metrics are observed. The metrics consist of the mean square error, accuracy and precision. The performance of the model is measured once after the model is trained as well as during different epochs during the training phase. The measurement of the performance after training is conducted with the help of the testing set, while measurement of the performance during training utilizes the validation set. The validation set helps to measure the progress of the model throughout the training.

A screenshot of a cell phone

Description automatically generated

**Figure 4**: Model accuracy.

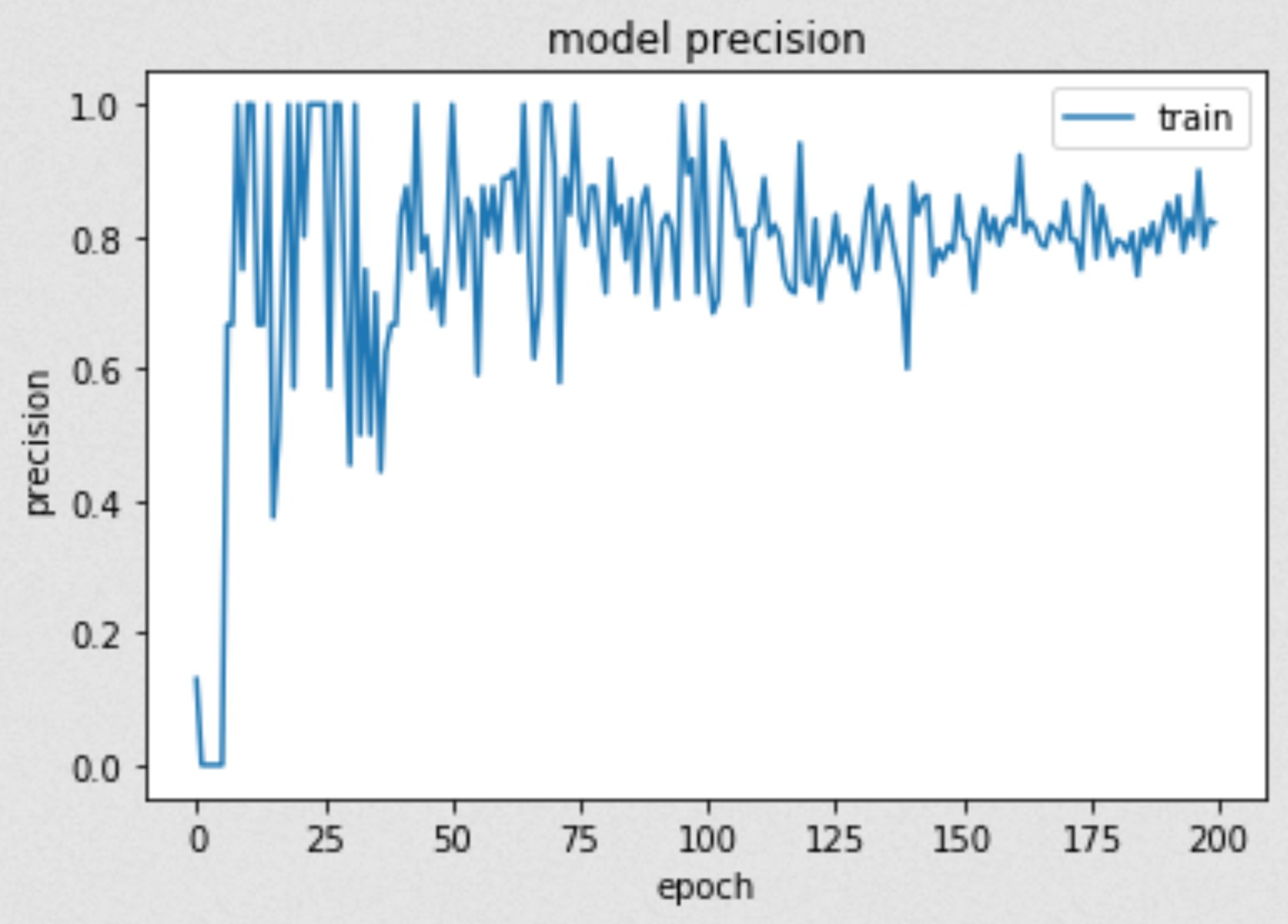
The accuracy of the model is the most significant indicator of performance. By comparing the predictions of the model to the correct outputs in the dataset for a given series of inputs, a reliable estimation of the correctness of the predictions can be obtained. The model that this study proposed yields an overall accuracy of 88.29%. Figure 4 showcases the performance of the model during the two hundred (200) epochs. By looking at Figure 4, it can be understood that the model optimizes its parameters to reach the overall accuracy during the testing phase in the first few epochs. Over the next iterations, the model increases its accuracy slowly when measured with the training set, but it remains stable while using the validation set. It is possible that this phenomenon occurred because the model adjusted its weights to learn the patterns in the training set, not the validation set. Additionally, the validation set may have contained edge case entries that the training set did not contain. A larger database would support the mitigation of this issue.

A screenshot of a cell phone

Description automatically generated

**Figure 5**: Model loss.

The loss of the model is a measurement of the error of the predictions. In this study, the error is measured as the mean square error. This metric is used to tune the weights of the model with the help of backpropagation. The goal of training the model is to decrease the loss as much as possible. Decrease in loss is corelated with increase in performance overall. The lower the loss is, the better the performance will be. Figure 5 showcases the loss over the training period. The loss decreases over time when tested with the training set. However, it seems to be slightly be going upward when tested with the validation set. This slight increase may be due to the model still lightly overfitting on the training data although it has a dropout layer. A larger database with a more diverse distribution may resolve this issue.



**Figure 6:** Model precision

The last metric that is observed to measure performance is precision. Precision is the ability for the model to provide consistent output for a given input. Analysis of the precision can be interpreted as a measurement of the model’s change in prediction difference over the training period. The closer the value is to one hundred percent (100%), the more similar the output is for a given series of inputs and the more stable the model is becoming. Stability of the model can suggest that the model is not learning much more. In some cases, this indicates that the model is reaching a minimum in the loss and thus, an optimal set of weights. Figure 6 describes the precision of the model proposed over the lifecycle of the training. It can be observed that as the number of epochs increase, the more stable the precision becomes and the closer to one (1) it is.

8 Future Work and Conclusion

This study has reached a conclusion that non-medical factors associated with patients do in fact play a role in predicting patients’ survival after an ICU stay with diagnoses of CHF and/or AFIB. With these attributes (GEN, AGE, ADMIT, INS, MAR, DIAG, PREV, and CURR), this study is able to reach approximately 88% accuracy in predicting whether a patient with cardiac diagnoses (CHF and AFIB in this case) will survive their stay in the ICU wing. Aside from patients’ medical background, this study views demographic socioeconomic factors as viable predictors of their survival chance using data mining tools. Further development or research into these factors is crucial to create a better predictive model, as this would result in more accurate tool that could ultimately assist doctors in their medical decision-making that could lead to an increase likelihood of patient survival.

This study would propose several avenues that future research could pursue. As mentioned previously, this study has encountered significant overfitting issues due to the higher ratio of patient surviving compared to those that expired during their stay in the hospital. Acquiring more datasets or the usage of other sampling methods to counteract this issue would be ideal.

Another suggestion that this study would offer is the testing of dataset retrieved from hospitals located in less prosperous neighborhood. In the interest of socioeconomic factors affecting patient’s survival, this potential future study may further corroborate our findings, or perhaps it could lead to a different conclusion altogether. A study done on hospitals that only have general ICU wings compared with hospitals that are equipped with cardiac-specific ICU wings can also be considered a viable/noteworthy topic to pursue.

Finally, this study would like to suggest future research to consider other attributes that this study did not examine. Researching for more potential correlations between other socioeconomic factors and patients’ survival are always welcomed and would further strengthen/improve overall predictive accuracy.

9 References

Donal, E., Flecher, E., & Tattevin, P. (2017). Machine learning to support decision-making for cardiac surgery during the acute phase of infective endocarditis. *Heart, 103*(18), 1396–1397. doi: 10.1136/heartjnl-2017-311512

Farmer, S. A., Grady, K. L., Wang, E., Mcgee, E. C., Cotts, W. G., & Mccarthy, P. M. (2013). Demographic, Psychosocial, and Behavioral Factors Associated with Survival After Heart Transplantation*. The Annals of Thoracic Surgery, 95*(3), 876–883. doi: 10.1016/j.athoracsur.2012.11.041

Gerward, S., Tyden, P., Hansen, O., Engstrom, G., Janzon, L., & Hedblad, B. (2006). Survival rate 28 days after hospital admission with first myocardial infarction. Inverse relationship with socio-economic circumstances. *Journal of Internal Medicine, 259*(2), 164–172. doi: 10.1111/j.1365-2796.2005.01594.x

Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., … Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet. *Circulation, 101*(23). doi: 10.1161/01.cir.101.23.e215

Johnson, A. E., Pollard, T. J., Shen, L., Lehman, L.-W. H., Feng, M., Ghassemi, M., … Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. *Scientific Data, 3*(1). doi: 10.1038/sdata.2016.35

Lecun, Y. A., Bottou, L., Orr, G. B., & Müller, K.-R. (2012). Efficient BackProp. *Lecture Notes in Computer Science Neural Networks: Tricks of the Trade*, 9–48. doi: 10.1007/978-3-642-35289-8\_3

Nanayakkara, S., Fogarty, S., Tremeer, M., Ross, K., Richards, B., Bergmeir, C., … Kaye, D. M. (2018). Characterising risk of in-hospital mortality following cardiac arrest using machine learning: A retrospective international registry study. *PLOS Medicine, 15*(11). doi: 10.1371/journal.pmed.1002709