Heart Failure – Predict readmission at 6 months

Ali Shadman Yazdi, Yin Lin, Gaia Vettori, Siavash Taleghani

1. Introduction

Heart failure is a prevalent medical condition with significant implications for patients and healthcare systems worldwide. It is estimated that over 26 million people are affected by heart failure globally, and this number continues to rise due to aging populations and the increasing prevalence of risk factors such as obesity, diabetes, and hypertension [1]. Statistics reveal that nearly one in four heart failure patients experience readmission within a mere 30 days after discharge, while approximately half face readmission within a span of 6 months [2]. Recognizing the negative impact of these readmissions, our project aims to address this issue by developing a predictive model leveraging electronic healthcare records of patients diagnosed with heart failure. By understanding the characteristics associated with readmission, our ultimate objective is to optimize treatment plans and enhance patient care, thus reducing the occurrence of readmissions effectively.

Over the past decade, numerous studies have focused on developing predictive models to forecast the likelihood of hospital readmission in heart failure patients. These models utilize a variety of clinical, demographic, and socioeconomic variables to estimate the risk of readmission. For instance, factors such as age, gender, comorbidities, previous hospitalizations, laboratory values, and medication usage have been identified as important predictors in many studies [3]. Furthermore, researchers have explored the use of advanced machine learning techniques to enhance the accuracy of predictive models for heart failure readmission. These techniques include logistic regression, decision trees, random forests, support vector machines, and artificial neural networks. Machine learning algorithms can effectively analyze complex interactions among numerous variables and provide individualized risk assessments for heart failure patients, thereby assisting healthcare professionals in making informed decisions regarding patient care and resource allocation. The publication, titled "Trends in 30- and 90-Day Readmission Rates for Heart Failure" [4], conducted an analysis on patient data derived from multiple hospitals across the United States that had participated in a Hospital Readmission Reduction Program (HRRP) spanning from 2010 to 2017. Nevertheless, the outcomes revealed that the rates of readmission were higher in the post-HRRP period as compared to the pre-HRRP phase, apart from low-volume hospitals.

The publication, titled "Statistical Models and Patient Predictors of Readmission for Heart Failure" [5], represents a systematic review that explores various statistical models used to assess patient risk and associated characteristics related to hospital readmission among individuals admitted for heart failure. The study focused on English publications spanning the period from 1950 to 2007. However, no specific model was identified that was exclusively designed to compare readmission rates. Instead, the emphasis was primarily on assessing readmission risk. The analysis revealed the presence of diverse approaches within these models, leading to inconsistencies among them. Consequently, the study concludes that, from a policy perspective, there is currently no statistical model available to accurately predict readmission rates.

In conclusion, despite the progress made in developing predictive models for heart failure readmission, there is still a need for further research to improve their accuracy and clinical applicability. Many existing models have limitations such as inadequate accuracy, lack of external validation in different patient populations, and limited integration into real-time clinical workflows. Additionally, the incorporation of novel data sources such as wearable devices, electronic health records, and social determinants of health can potentially enhance the predictive capabilities of these models [6].

In this project, our objective is to address these challenges and develop an accurate predictive model to forecast readmission at 6 months for patients with heart failure. We will leverage a comprehensive dataset of clinical and demographic variables and employ machine learning techniques to train and validate the model. The final objective for this project will be to provide clinicians with a valuable tool that can assist in identifying high-risk patients and tailoring interventions to reduce readmission rates, improve patient outcomes, and optimize healthcare resource allocation.

2. Material and Methods

Dataset cleaning:
The dataset utilized for this project comprises electronic health data pertaining to patients diagnosed with heart failure, who were admitted to a hospital in Sichuan, China, spanning the years 2016 to 2019. This comprehensive dataset encompasses 166 variables, capturing vital information for a total of 2,008 patients diagnosed with heart failure. With a total of 333.328 values of which 61.787 (20%) are missing. The features include 151 numerical ones and 15 categorical ones. The project objective involves setting up a system to predict the likelihood of patient readmission to the medical clinic within six months following their initial admission. Additionally, the aim is to explore and elucidate the potential impact of drug usage on readmission rates.

After identifying patients who had passed away between their admission and the recording of data, we proceeded by removing them from the dataset. This involved excluding patients with positive values in the "death.within.28.days," "death.within.3.months," and "death.within.6.months" features, as well as those categorized as "Died" in the "DestinationDischarge" feature. As a result, the revised dataset now consists of 1946 patients. The target variable in our dataset exhibited a distribution of 60:40. This indicates that out of the 1946 patients in our dataset, approximately 40% of them were readmitted back to the hospital within a 6-month period. This information will be taken into consideration during the modeling and evaluation phases.

The next step was to remove the features that wouldn't have been available at admission as they were recorded in admissions or events after the first admission, such as the deaths mentioned before and the readmissions at 28 days and 3 months, "outcome.during.hospitalization", "time.to.emergency.department.within.6.months", "time.of.death..days.from.admission.", "re.admission.time..days.from.admission.", "DestinationDischarge", and "return.to.emergency.department.within.6.months".

To mitigate biases arising from missing data, we implemented a strategy to remove any features that exhibited a missing value rate exceeding 50 percent. This decision was made to ensure that variables with such high levels of missingness, which would possess limited statistical power, were excluded. By taking this approach, we aimed to prevent potential biases that could arise during the process of imputing missing values.

In addition to the electronic healthcare data of the patients, the provided dataset also included a separate dataset containing information about medications administered during their hospitalization. Upon analysis, we identified a total of 25 unique drugs in the dataset. Out of the 2008 patients, drugs were administered to 2004 individuals. The primary drug categories featured in the dataset were diuretics, inotropes, and vasodilators. However, it is worth noting that there were drugs that did not fall into these three categories, such as Shenfu injection and sulfotanshinone sodium injection, which are drugs generally used in Traditional Chinese Medicine.

To incorporate this medication information into our analysis, we employed dummy variables. A value of 1 was assigned to indicate that a patient had been administered a particular drug, while a value of 0 indicated that they had not received it. Subsequently, we merged the drug dataset with the original healthcare data, thereby combining the information from both datasets into a unified dataset for comprehensive analysis.

Outlier removal:
After conducting a visual inspection of the dataset, we detected the presence of outliers in certain numerical features. To address this issue, we employed various techniques for outlier detection and removal. These techniques included the utilization of the z-score method, interquartile range (IQR), and selection based on feature variance.

Initially, we applied the z-score method by setting a z threshold of 2. However, due to the non-Gaussian distribution of our data, we decided not to incorporate this method into the preprocessing pipeline. Next, we employed the IQR method to identify and remove values lying outside the interquartile range. Specifically, we eliminated values falling below the 0.5th percentile and above the 95th percentile. By doing so, we aimed to eliminate extreme values from the dataset, particularly the outliers observed in the tails of the box plot. It is important to note that while removing the outliers did lead to a slight improvement in accuracy, however this was carried out to enhance the overall robustness of the model.

Feature selection:
After dividing categorical and numerical variables, we obtain two datasets with 14 categorical variables (with 26 missing values in occupation) and 126 numerical ones (with 32907 missing values). Before addressing the missing values, we initially applied the "remove_collinear_features" function to the numerical dataset, employing a collinearity threshold of 0.80. This function calculated the linear correlation between features, and if the collinearity exceeded the threshold, it dropped the feature with the higher percentage of missing values. As a result, we obtained a dataset containing 1946 patients, comprising 102 numerical features with 5856 missing values and 13 categorical features with 26 missing values.

Additionally, feature selection was employed utilizing a model to effectively decrease the dimensionality of our database. Adaboost, chosen for its commendable performance, was the model of choice. Encouragingly, after implementing this processing step, the model exhibited enhanced performance.

Training and test sets and noise addition:
As part of the pre-processing phase, prior to proceeding with the remaining steps, we focused on the stratified splitting of the dataset into training and test sets. Our aim was to enhance the dataset, specifically the training set, to improve the performance of the model. To achieve this, we employed oversampling techniques while being cautious about overfitting.

The dataset was divided into a 75% training set and a 25% test set. Subsequently, we quadrupled the size of the training set by generating additional samples. Each of these four subsets underwent a different transformation:

1. The first subset had normal zero-mean noise added to the features with a 10% variance.
2. The second subset had normal zero-mean noise added to the features with a 20% variance.
3. The third subset had Poisson zero-mean noise added to the features with a 10% variance.
4. The fourth subset retained the original dataset without any noise or oversampling.

This approach was applied at this stage, prior to the remaining pre-processing steps, to ensure that the generated samples maintained realistic feature values. It should be noted that no noise or oversampling was introduced to the test set. This decision was made to evaluate the model's performance without any biases and assess its robustness under unbiased conditions.

Missing values:
Given that all the missing categorical values are confined to the "occupation" feature, our approach involved replacing these missing values with the label "Other", consequently, the categorical dataset was transformed into 48 dummy features. On the other hand, for the numerical missing values, we

employed K-nearest neighbors (KNN) imputation. After experimentation, we determined that setting the neighbor parameter to 5 yielded the most favorable outcomes in terms of imputation accuracy.

Standardization:
The final step of the pre-processing phase involves standardization, which is performed by applying the standard scaler on the numerical features. We also experimented with the MinMax scaler, but found that the standard scaler yielded better results for our specific case.

Randomization:
To mitigate overfitting and enhance the overall performance of our model, we employed the technique of randomizing the training set prior to model training and evaluation. This process involved shuffling the dataset to ensure that the data points were not ordered in any sequence.

3. Results

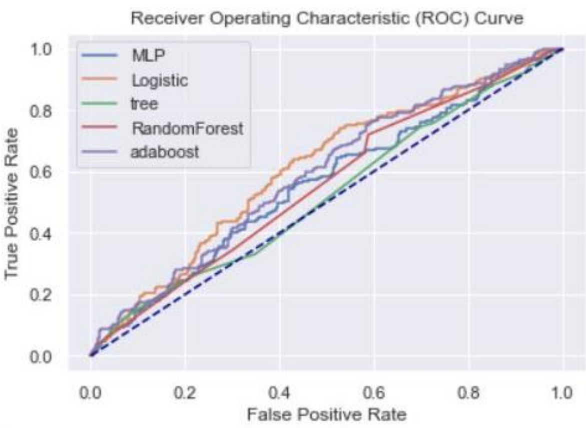

Figure 1

After the pre-processing phase, we pursued two strategies for model selection. Initially, we employed grid search to identify the optimal parameters among various machine learning models. Subsequently, we explored different combinations of hyperparameters to evaluate their impact on the model's performance, with the objective of maximizing accuracy and minimizing the loss function. As depicted in the Figure 1 on the left, the Logistic Regression Classifier demonstrated superior performance, achieving an accuracy of 0.64 and an AUC of 0.59.

Secondly, we also found a highly effective model in the context of neural networks, and our approach was more iterative. We began with a general architecture, such as an encoder or decoder structure commonly employed in tasks like sequence-to-sequence modeling. From there, we started a process of fine-tuning, where we modified various hyperparameters to improve performance. During this fine-tuning process, we experimented with factors such as the number of layers, seeking to strike the right balance between model complexity and capacity to capture patterns in the data. We also adjusted the number of neurons per layer, carefully considering the impact on model capacity and avoiding overfitting or underfitting. Furthermore, we explored different batch sizes, which determine the number of training examples processed in each iteration, and the number of epochs, representing the number of times the entire dataset is passed through the network during training. By iteratively adjusting these hyperparameters and evaluating model performance, we were able to select the combination that yielded the highest accuracy at 0.65-0.70 and AUC at 0.60. This iterative approach allowed us to fine-tune our neural network models, harnessing their full potential and improving the overall performance. The model we obtained is a 1-dimensional convolutional neural network with two convolutional layers.

The pre-processing steps were meant to prepare the data so that the model could achieve the best performance and accuracy. The removal of the collinear feature allowed us to work with a smaller dataset and increase the Neural Network model accuracy from 0.61 to 0.65. The oversampling method employed consisted in quadruplicating the training set by adding noise with different distributions to the duplicates of the original set. This step drastically improved both the machine learning models and the neural network and especially reduced the overfitting. The outlier removal through IQR and adding the new feature of hypertension increased both accuracy and AUC. Finally, the KNN imputation for missing values and the application of the standard scaler were crucial for achieving the AUC over 50%.

4. Discussion and conclusion

The papers cited in the introduction highlight the limited success of prediction models for hospital readmission in attaining high performance. Although our solutions achieve an accuracy of 0.65 and an AUC of 0.60, the sensitivity falls short, which is undesirable for classifiers in medical applications. Consequently, we opted for the logistic regression classifier over the neural network due to its superior sensitivity (0.49 vs 0.35) and weighted F1 score (0.65 vs 0.64).

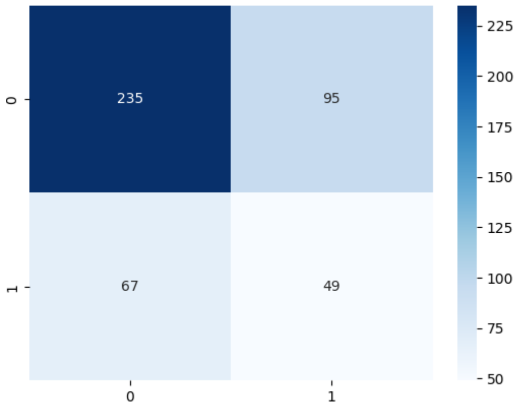

Figure 2

Despite both the logistic regression and neural network models yielding similar results, the logistic regression model is better suited to our project requirements. This choice is attributed to the model's capacity for implementing different interpretability techniques, enabling a comprehensive understanding of its functionality and the impact of individual features on the outcomes. Such interpretability was a primary goal of our project. Additionally, the logistic regression model outperforms the neural network in terms of sensitivity, resulting in a lower false negative rate. Thus, considering its interpretability advantages and superior sensitivity, the logistic regression model is the preferred choice for our project.

To address the similarity in target distribution caused by equal data distribution within the same range, we need to incorporate more diverse features. By increasing the sample size and introducing factors like family history, ECG data, or other relevant variables, we can enhance the model's performance. This will improve the dataset's representativeness and variability, allowing for a more comprehensive analysis. It's important to note that drugs alone are insufficient for achieving optimal performance; additional features are necessary for better results.

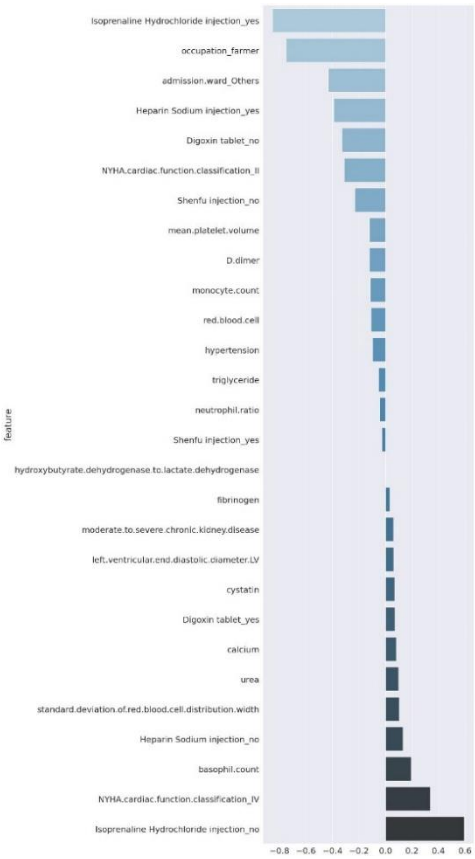

Figure 3

Moreover, due to the relatively small size of the dataset, we had to perform oversampling on the training set to ensure an adequate number of samples for effective model training. However, it is worth noting that a larger dataset would likely enhance the performance of the model.

In Figure 3, the feature importance analysis of our model is displayed, showcasing the significance of each feature. The positive segment of the figure represents the instrumental and crucial features that contribute to the accurate classification of readmission rates. Conversely, the negative segment highlights the features that lead to higher loss functions within the model.

Upon careful examination, it becomes evident that the drug feature played a crucial role in our predictive model. However, it is worth noting that the overall performance of the model was not exceptionally strong. From this observation, we can draw the conclusion that while the inclusion of drug features and data from electronic health records aids in predicting the model's outcomes, relying solely on these factors would not suffice to establish a robust and reliable model. To achieve a more dependable model, additional factors and features should be taken into consideration.

References:

[1] Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JGF, Coats AJS, Falk V, González-Juanatey JR, Harjola VP, Jankowska EA, Jessup M, Linde C, Nihoyannopoulos P, Parissis JT, Pieske B, Riley JP, Rosano GMC, Ruilope LM, Ruschitzka F, Rutten FH, van der Meer P; ESC Scientific Document Group. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC)Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. Eur Heart J. 2016 Jul 14;37(27):2129-2200. doi: 10.1093/eurheartj/ehw128. Epub 2016 May 20. Erratum in: Eur Heart J. 2016 Dec 30;: PMID: 27206819.

[2] Virani SS, Alonso A, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, Chamberlain AM, Chang AR, Cheng S, Delling FN, et al. American heart association council on epidemiology and prevention statistics committee and stroke statistics subcommittee. heart disease and stroke statistics-2020 update: a report from the American Heart Association. Circulation. 2020;141:e139–e596. doi: 10.1161/CIR.0000000000000757. Epub 2020 Jan

[3] Kansagara D, Englander H, Salanitro A, Kagen D, Theobald C, Freeman M, Kripalani S. Risk prediction models for hospital readmission: a systematic review. JAMA. 2011 Oct 19;306(15):1688-98. doi: 10.1001/jama.2011.1515. PMID: 22009101; PMCID: PMC3603349.

[4] Trends in 30- and 90-Day Readmission Rates for Heart Failure
Muhammad Shahzeb Khan, MD, MSc, Jayakumar Sreenivasan, MD, MSc, Noman Lateef, MD, Marwan S. Abougergi, MD, Stephen J. Greene, MD, Tariq Ahmad, MD, MPH, Stefan D. Anker, MD, PhD, Gregg C. Fonarow, MD, Javed Butler, MD, MPH, MBA

[5] Ross JS, Mulvey GK, Stauffer B, et al. Statistical Models and Patient Predictors of Readmission for Heart Failure: A Systematic Review. Arch Intern Med. 2008;168(13):1371–1386. doi:10.1001/archinte.168.13.1371

[6] Jason P Burnham, Chenyang, Lu, Lauren H Yaeger, Thomas C Bailey, Marin H Kollef; Using wearable technology to predict health outcomes: a literature review. J Am Med Inform Assoc. 2018 Sep; 25(9): 1221–1227. doi: 10.1093/jamia/ocy082