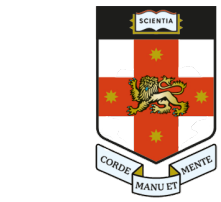
University of New South Wales

Machine Learning and Data Mining
(COMP9417)

In-hospital Mortality Prediction for Intensive Care Unit Patients with Heart Failure



UNSW
SYDNEY

Major Project by NAFT

Novin Wijesundara (z5348315)
Tanya Hollis (z5323560)
Lee Chung Lai (z5210146)
Yun Li (z5231701)
Anish Saraogi (z5428544)

In-hospital Mortality Prediction for Intensive Care Unit Patients with Heart Failure

October 7, 2023

Abstract

In this study, a machine learning-based prediction model was developed to predict in-hospital mortality among ICU-admitted heart failure patients. The team used the MIMIC-III dataset and reduced the feature count to 25, configuring and training eight machine learning classifiers. The Gradient Boosting classifier (GBC) exhibited an accuracy of 95% and was selected as the final model. The models were evaluated using f1-score and precision-recall curves and our final GBC model outperformed the previous model built by other researchers (Li et al. 2021). Our model will offer a practical tool for clinicians to evaluate patient outcomes and make data-driven ICU admission and medical attention decisions. This AI development has the potential to improve clinical decision-making and patient outcomes for at-risk heart failure patients.

1 Introduction

This project is an extension of the results achieved by researchers from Qingdao University (Qingdao, China) and Fudan University (Shanghai, China). The aim of this study is to implement a machine learning model for classification in Python with the use of the Medical Information Mart for Intensive Care (MIMIC-III) dataset, which included data on 1177 heart failure patients with 48 features (Li et al. 2021). The objective is to improve the performance of the previous model by performing experiments with a number of techniques such as feature engineering and selection, sampling, and the use of other ML models.

In Australia, heart failure (HF) affects approximately 480,000 Australians, with over 60,000 new diagnoses made every year. This is expected to further increase and put pressure on the healthcare system. Further data shows if proper care is not provided, around 50% to 75% of people with HF die within 5 years of diagnosis (Richard Whaddon Parsons 2020). Therefore, identifying patients who have a higher probability of dying is a crucial task to ensure appropriate treatment is provided immediately.

Previous researchers developed a final model using XGBoost with an accuracy of 84% to aid medical practitioners with clinical decision-making. However, in light of the high criticality of the use case, there is a pressing need to enhance both the accuracy and the rate of false positives in the current solution.

In this project, an extensive feature analysis was conducted, and additional predictive features were generated. Due to the highly imbalanced nature of the dataset, the SMOTE sampling technique was employed. Moreover, eight different Scikit-Learn models, including decision tree (DT), random forest ensemble (RF), gradient boosting (GBC), neural network (NN), k-nearest neighbors (KNN), support vector machine (SVM), extreme gradient boosting (XGBoost), and voting classifier ensemble (VC) were developed. The performance of these models was evaluated using state-of-the-art validation methods such as ROC, F1 score, precision-recall curve, GWTG-HF score, and accuracy, to ensure robust results.

This proof of concept can be further extended to provide models for further decision-making within hospitals which will allow medical practitioners to focus purely on the patients and provide the best care possible. Furthermore, this will allocate the resources appropriately based on a prioritised manner for critical patients.

2 Exploratory Data Analysis

This project uses the Medical Information Mart for Intensive Care (MIMIC-III) dataset, which included data on 1177 heart failure patients with 48 features. The original dataset contains 1017 rows (86.4%) for the alive class, 159 rows (13.5%) for the death class, and 1 null row. Initial analysis showed that there were 19 feature columns with missing data where eight of which had missing data ranging from 20-25%, while the remaining columns had missing values of less than 2%.

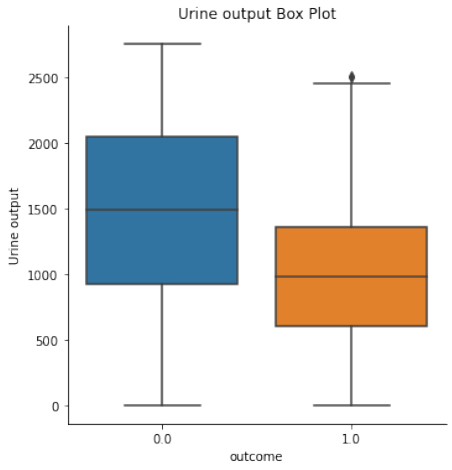


Figure 1: Box plot indicating the urine output of all patients.

The features were graphically represented using four main methods, box plots, histogram, Q-Q plots, and feature values were binned into ten or more histogram bins and plotted with the average death rate to observe the general trends.

All original features were evaluated using the box plots. In Figure 1, the box plot is used to display the distribution of urine output data for patients who survived (first box) and patients who died (second box). The boxes represent the inter-quartile range (IQR), and the mean value marked in the horizontal line. Individual points outside of the whiskers represent the possible outliers. In this particular case, more deaths can be observed in patients who have lower urine output, therefore, this can be expected to be a prediction feature for this classification problem.

In addition, to observe the trends of features to predict the desired outcome were graphically represented using binned histograms. Figure 2, represent one of the original features created and a clear trend can be observed as patients who have a higher z-score sum is likely to die.

Based on the analysis number of features including diabetes, deficiency anemias, depression, hyperlipemia, renal failure, and COPD were removed from the dataset. These features indicated a

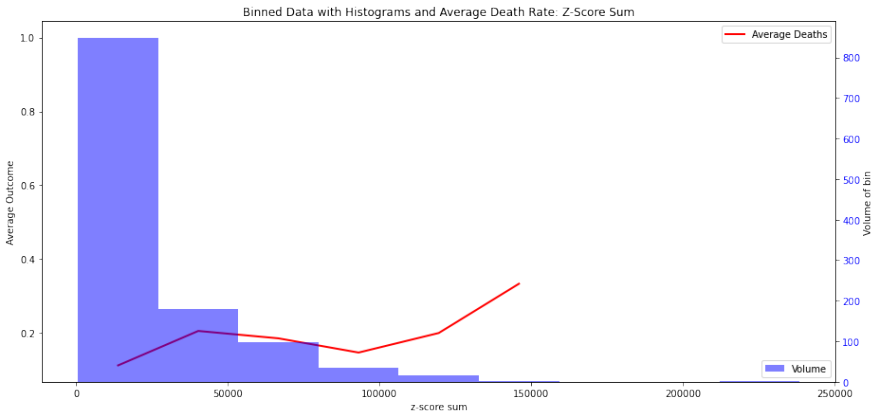


Figure 2: Binned histogram plot indicating the z-score sum and the average death rate of each bin for all patients.

trend in the opposite direction, suggesting that they may not be significant predictors and conflict with the outcome being predicted.

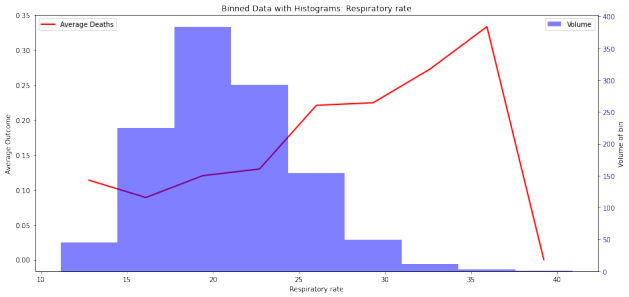


Figure 3: Binned histogram plot indicating the respiratory rate and the average death rate of each bin for all patients.

The binned histograms were also used to evaluate features where a clear trend cannot be observed and based on two factors, observations from the graph and healthy ranges based on research, these features were converted into categorical features. For example, Figure 3, shows the respiratory rate feature of the dataset. The trend line shows a low volume of deaths when the rate is extremely high due to insufficient data. Therefore, these features were converted into a categorical feature where the healthy range between 12-22 (low death rates can be observed from Figure 3) is converted to 0 and the rest is converted to a value of 1.

Moreover, new features created include processed age, correlation and R2 Score, Z-Score calculation, unserious diseases count, and health indicator calculation. These features were combined with the original features and the most predictive features were obtained for the final model. Further details on the original features and the methods used to reduce the number of features used in the final model can be found in the sections below.

To assess the distribution of our features, we used histogram and Q-Q plots. An example shown is feature NT-proBNP. From the histogram in Figure 4, it shows the feature is right skewed. The skewness is also shown in Q-Q plot in Figure 5.

In addition to the above-mentioned methods, further analyses were done to find the most predictive features. Namely, we performed Recursive Feature Elimination (RFE), VIF filtering, and filtering features based on the p-value and Information Value (IV). Most important features were derived using these methods and then a correlation matrix was used to determine the most correlated

value. Highly correlated value with the lower IV was dropped to reduce the number of features in the final model.

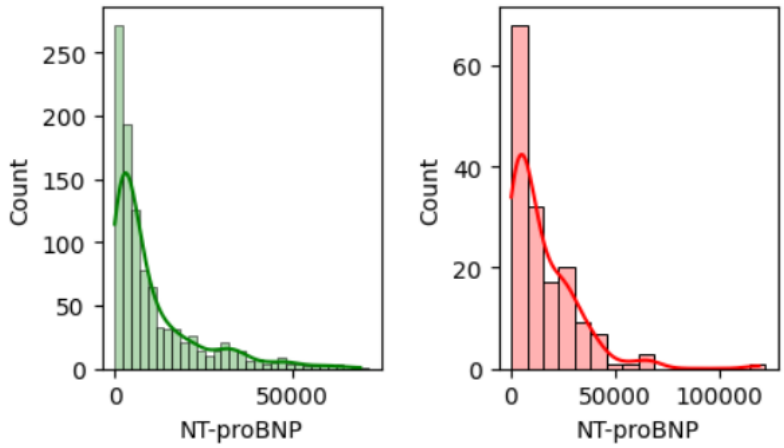


Figure 4: NT-proBNP histogram for negative (left) and positive (right) labels, binned using Freedman-Diaconis's Rule

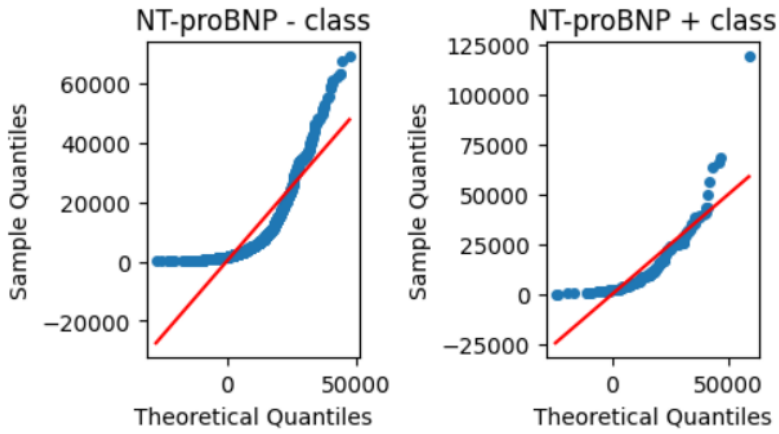


Figure 5: NP-proBNP Q-Q plot

3 Methodology

In this section, a detailed explanation will be provided regarding data preprocessing, feature engineering, feature selection and model development and validation.

3.1 Data Preprocessing

3.1.1 Handling missing data

We filled missing values with either mean or median, based on the skewness of the column. Skewness (γ) measures data asymmetry. For ≤ 0.5, the mean is appropriate, maintaining central tendency. For > 0.5, the median is more robust, being less sensitive to extreme values. By doing

this, we tailor the imputation approach to each variable's distribution, minimizing the impact on subsequent data processing. This enhances the dataset's quality and reliability, enabling our model to make more accurate inferences.

3.1.2 Sampling using SMOTE

The dataset included a significantly underrepresented class where only 13.5% of the dataset included death classes, therefore, illustrating that this is an imbalanced dataset. This can lead to biased model predictions and poor performance. One common technique used to address this issue is Synthetic Minority Over-sampling Technique (SMOTE). SMOTE generates synthetic samples for the minority class by creating synthetic instances along the line segments that connect the minority class samples. The most optimal percentage was 18% sampling rate. This helped to balance the class distribution by increasing the number of rows where the outcome was a death which then improved the model's ability to learn.

3.2 Feature Engineering

3.2.1 Building up a healthy range for each indicator

The `health_dictionary()` function initializes the "Good_range" dictionary with healthy values for both male and female patients for numerous features such as age, BMI (Organisation 2010), blood pressure (Cardiology 2018), respiratory rate (Publishing 2020), temperature (Clinic 2022), and various blood parameters. These healthy ranges are based on medical guidelines and serve as a reference to assess the health status of the ICU-admitted HF patients in our dataset.

3.2.2 Age Processing

We transformed the age feature using a custom 'more_extrem' function, which takes into account the gender and good age range for both male and female patients. This transformation helped to emphasize extreme age values outside the good age range, thereby capturing the potential impact of age on in-hospital mortality. The mathematical principle behind this transformation is the normalization of age values using a sigmoid function: Normalized Age = $\frac{1}{1+e^{(-k\times(age-x0))}}$ Where 'k' represents the steepness of the curve, 'x0' is the midpoint of the curve, and 'age' is the original age value.

3.2.3 Correlation and R2 Score

We calculated the correlation and R2 score between each feature and the target variable (in-hospital mortality). These values help to understand the importance of each feature in predicting the target variable. The correlation is a measure of the strength and direction of a linear relationship between two variables, computed using Pearson's correlation coefficient formula: $r = \frac{\sum((x-\bar{x})(y-\bar{y}))}{\sqrt{(\sum(x-\bar{x})^2 * \sum(y-\bar{y})^2)}}$ The R2 score is the proportion of the variance in the target variable that can be explained by the features, calculated as the square of the correlation coefficient: $R^2 = r^2$.

3.2.4 Z-Score Calculation

We calculated the z-score for each feature by standardizing the data, subtracting the mean, and dividing by the standard deviation. This normalization ensures that features are on the same scale, allowing the model to better capture the relationships between features and the target variable. The mathematical principle behind z-score calculation is as follows:
Z-score = $\frac{x-\mu}{\sigma}$ Where 'x' is the original value , 'μ' is the mean of the dataset, and 'σ' is the standard deviation of the dataset.

3.2.5 Dimensionality Reduction

We applied Principal Component Analysis (PCA) to reduce the dimensionality of the dataset while retaining the most important information. PCA transforms the original features into a new set of linearly uncorrelated features, called principal components. The mathematical principle behind PCA involves computing the eigenvectors and eigenvalues of the data's covariance matrix: Covariance Matrix = $\frac{1}{(n-1)} * \sum((X-\mu)(X-\mu)^T)$ Eigenvectors and eigenvalues are then computed, and the eigenvectors corresponding to the largest eigenvalues are selected as the principal components.

3.2.6 Binned Feature Generation

A number of categorical features were created using continuous columns to allow the model to interpret the data more effectively. As shown in Figure 3, for certain columns the trends of the death rates were not as expected. In unhealthy ranges, a low number of deaths were observed. Therefore, by creating a categorical column where the values within the unhealthy ranges are converted to 1 and healthy ranges are converted to 0, the model was able to perform better. These features include respiratory rate, temperature, SP O2, hematocrit, RBC, MCH, MCHC, RDW, lymphocyte, blood sodium, blood calcium, chloride, and magnesium ion. Once these features were concerted to categorical, the original features were dropped from the dataset to avoid any duplicate information.

3.2.7 Other features generated

- **Unserious Diseases Count:** We calculated the number of unserious diseases each patient has by counting the diseases present in the 'not_serious_list'. This feature may help to understand the overall health condition of the patients and its potential impact on their mortality. The mathematical principle behind this feature is the summation of binary values: Unserious Diseases Count = $\sum(disease_i)$ Where '$disease_i$' is a binary value indicating the presence (1) or absence (0) of an unserious disease.

- **Health Indicator Calculation:**
  We computed the number of health indicators outside the good range for each patient using the 'count_unhealthy_indicator' function. This feature may highlight patients with a higher number of abnormal health indicators, potentially increasing their risk of mortality. The mathematical principle supporting this feature is a summation of binary values, similar

to the unserious diseases count:
Health Indicator Count = $\sum(indicator_i)$
Where '$indicator_i$' is a binary value indicating whether a health indicator is outside the good range (1) or within the good range (0).

- **Z-Score Weighted by Correlation and R2:**
  To further emphasize the importance of features with a high correlation and R2 score, we multiplied the z-score by the correlation and R2 score for each feature. This adjusted z-score emphasizes the influence of highly correlated features in the model.

3.3 Feature Selection

3.3.1 Recursive Feature Elimination (RFE)

We used Recursive Feature Elimination (RFE) to select the most important features contributing to the prediction of in-hospital mortality. RFE is a backward feature elimination technique that iteratively removes the least important features based on the model performance. The mathematical principle behind RFE is the evaluation of feature importance using the model's coefficients or feature importance.

3.3.2 VIF filtering

Due to the large amount of features, to further improve interpretability and remove multicollinearity, variance inflation factor (VIF) filtering is applied to numerical features as our last step in feature selection. The filtering is an iterative method. In each iteration, VIF for each numerical feature is computed, and the feature with the highest VIF is removed from our numerical feature set. The filtering stops when VIF for all feature is below a specified threshold. The remaining numerical features is than combined with the categorical features to form our final feature set. In our model, we have chosen a VIF threshold of 7. Since there isn't universal agreement on VIF threshold, we chose a threshold between 10 and 5, two popular VIF thresholds. Using VIF filtering, Neutrophils, PH, Anion gap, Bicarbonate, MCV, howmany, Blood potassium, age, heart rate and Systolic blood pressure are removed.

3.3.3 Correlation Analysis

Correlation analysis was done to obtain the most correlated features. Then out of these, the least predictive features were dropped to allow the model to reduce overfitting and to build a general model for better performance.

3.4 Model Training and Validation

Prior to the finalisation of the feature selection, estimator-model evaluation and selection began on numerous classification estimators with the default hyper-parameters. The data for the development

environment included 90% of the total data set, with 3-Fold cross validation. This data was analysed for linear relationships with the outcome (target)1, however data did not seem to be linearly separable. We identified several models over-fitting, however, with only just over 1100 instances, this was not easy to remedy beyond the extensive work already carried out from Feature Selection and Engineering. During this early model analysis, the nine best performing models (based on Accuracy) were chosen for further analysis and hyper-parameter tuning. The nine best included five classifier models: Multi-Layer Perception (NN), Extreme Gradient Boosting / XG-Boost (XGB), Decision Tree (DT), K-Nearest Neighbors (KNN) and Support Vector Machine - Kernel (SVC) and four ensemble models: Random Forrest (RF), AdaBoost (ADB), Gradient Boosting Machine (GBC) and Voting Classifier (VC). AdaBoost was dropped from this list due to extreme over-fitting identified early on compared to the other three ensemble classifiers. So, the final list included the remaining eight.

After applying the aforementioned feature engineering techniques and selecting the most relevant features, the eight different machine learning models were trained. We used cross-validation to assess the model's performance on unseen training validation data and fine-tuned its hyper-parameters to optimise the prediction accuracy. Most hyper-parameter values were found using Sci-kit Learns' "Halving Grid Search CV" library (Scikit-Learn 2023). See the results section for a snap-shot of the cross validation results 1.

In our model evaluation process, we utilised the GWTG-HF score as a benchmark for comparison. However, we encountered the challenge of imbalanced class distribution. To address this issue, we primarily relied on metrics such as f1-score and precision-recall curve to assess the performance of our models. The results are discussed further in details in the following sections.

4 Results

Once the complete feature set was finalised, the cross validation development environment could finally configure accurate hyper-parameters and produce encouraging results which were competitive with our team's comparison model of 85% (Li et al. 2021). Here are the cross validation accuracy results of the estimators, subsequent to finding optimum parameters:

Model	Fold 1	Fold 2	Fold 3	CV Mean Accuracy
GBC	0.903	0.892	0.875	0.89
XGB	0.898	0.884	0.878	0.887
ADB	0.89	0.876	0.884	0.883
VC	0.878	0.859	0.875	0.871
NN	0.867	0.854	0.859	0.86
DT	0.845	0.854	0.864	0.854
RF	0.848	0.848	0.848	0.848
SVC	0.848	0.848	0.848	0.848
KNN	0.845	0.812	0.853	0.837

Table 1: Classifier Models with their training Cross Validation 3-Fold Validation results

For final model performances on unseen test data, we will use PR curve, F1-score and AP due to class imbalance as accuracy and AUC-ROC will be inflated due to high amount of negative class. Similar to previous research Li et al. 2021, we are using Get With the Guidelines-Heart Failure (GWTG-HF) risk score as benchmark for comparison. The risk score is based on age, systolic blood pressure, blood urea nitrogen, heart rate, serum sodium, COPD and ethnicity. In our dataset, ethnicity feature is missing. However, it only contributes to a score of 3 out of 100, so it is a good enough estimate. The risk score is calculated according to Figure 6 (Peterson et al. 2010).

Model	Accuracy	F-score	AP
MLP	0.839	0.30	0.522
XGB	0.907	0.72	0.821
GBC	0.949	0.86	0.922
DT	0.856	0.45	0.523
KNN	0.814	0.21	0.280
SVC	0.822	0.0	0.535
RF	0.822	0.0	0.672
VC	0.881	0.5	0.885

Table 2: Models and their corresponding F-score and average precision on test set

The score for each prediction is then used to generate AP and PR curves. The metrics of test set for different models are shown in Table 2, and PR curves of our models and GWTG-HF risk score are shown in Figure 7. The confusion matrix for the top 2 models (GBC and XGB) and the worst model (SVC) is shown in Figure 8 and 9 respectively.

Systolic BP	Points	BUN	Points	Sodium	Points	Age	Points
50-59	28	≤9	0	≤130	4	≤19	0
60-69	26	10-19	2	131	3	20-29	3
70-79	24	20-29	4	132	3	30-39	6
80-89	23	30-39	6	133	3	40-49	8
90-99	21	40-49	8	134	2	50-59	11
100-109	19	50-59	9	135	2	60-69	14
110-119	17	60-69	11	136	2	70-79	17
120-129	15	70-79	13	137	1	80-89	19
130-139	13	80-89	15	138	1	90-99	22
140-149	11	90-99	17	≥139	0	100-109	25
150-159	9	100-109	19			≥110	28
160-169	8	110-119	21				
170-179	6	120-129	23				
180-189	4	130-139	25				
190-199	2	140-149	27				
≥200	0	≥150	28				

Heart Rate	Points	Black Race	Points	COPD	Points
≤79	0	Yes	0	Yes	2
80-84	1	No	3	No	0
85-89	3				
90-94	4				
95-99	5				
100-104	6				
≥105	8				

Total Score	Probability of Death
0-33	**<1%**
34-50	**1-5%**
51-57	**>5-10%**
58-61	**>10-15%**
62-65	**>15-20%**
66-70	**>20-30%**
71-74	**>30-40%**
75-78	**>40-50%**
≥79	**>50%**

Figure 6: GWTG-HF risk score calculations

With the confusion matrices below, we have in-hospital ICU mortality = 1 (+) and survival = 0 (-). As such, with the GBC confusion matrix, 94 ICU patients were correctly classified (survived), 18 patients were correctly classified as in-hospital mortality (deaths), 3 patients were predicted to have died, however, they survived, and finally, the false negative case, 3 patients were predicted to

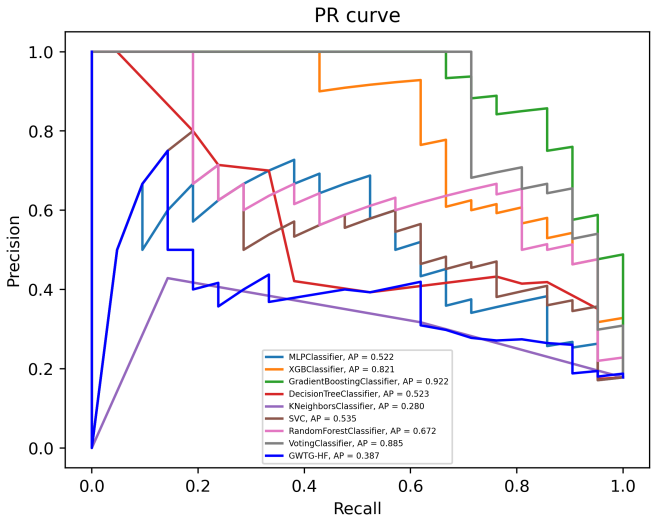


Figure 7: Precision-Recall Curve for our models and GWTG-HF risk score

live (-), however, patients actually died (+). This value, the False negative, is the value our team wishes to minimise for ethical and medical professionalism reasons.

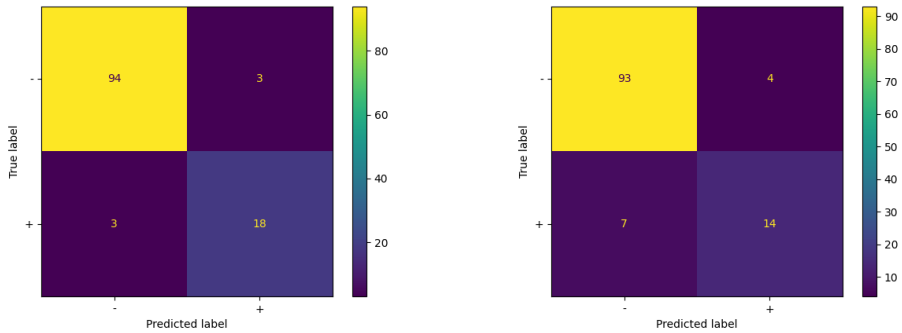


Figure 8: GBC confusion matrix (left) and XGBoost confusion matrix (right)

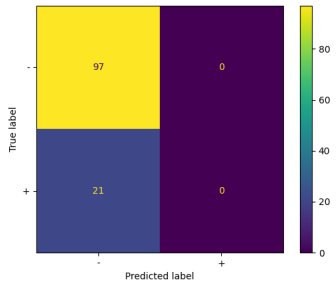


Figure 9: SVC confusion matrix

The confusion matrix of the SVM shows 0 predictions for the death class. This illustrates that the problem is too complex to be solved using an SVM or further testing is required to find an appropriate kernel etc.

Our final GBC model was built on 25 features, our own engineered and original features, and was able to outperform the final model developed by Li et al. 2021 by approximately 10% higher accuracy. More importantly, the model was able to reduce the number of false positives which will allow the medical staff to provide the necessary care for the most urgent patients and that critical patients will not be neglected.

Finally, further improvements to be made include, it can be trained using much more complex models to increase the F1 score even further, and a probability can be provided using a model such

as a Bayesian model to provide confidence in the predictions made by the model. This will allow the ability to reduce the number of false positive values even further. This proof of concept addressed patients with HF, however, if more features are used a similar model can be used to make other areas of hospitals more efficient as well.

5 Discussion

The project involved building a machine learning-based prediction model for predicting in-hospital mortality among ICU-admitted heart failure patients. The MIMIC-III dataset was used and features were reduced to 25 most predictive features. Finally eight machine learning classifiers were configured, trained, and evaluated.

Multiple challenges were presented to our team, these challenges included; a small number of labelled instances, skewed dataset (the dataset consisted of mostly patients who had survived the ICU), missing values and non-linear features just to name a few which potentially compromised delivering an accurate classification model. Our team met these challenges via; AI model selection, feature engineering techniques and applying medical heuristic research knowledge to the data. All of these efforts resulted in a more effective and accurate model. Decisions resulting from medical research included removal of certain features, normalising continuous columns, generating features, and statistical methods such as Variance Inflation Factor (VIF) to eliminate multicollinear features.

It is essential to address missing values to maintain data integrity and ensure the model can accurately learn from the dataset. By properly handling missing data, biases, and inaccuracies in predictions stemming from unaddressed gaps in the data can be avoided. We used SMOTE sampling to tackle class imbalance in the dataset. This technique creates synthetic data points for the minority class, enabling the model to better distinguish between classes. As a result, prediction accuracy improved, and the impact of class imbalance on the model's performance was minimized.

Following the feature engineering process, the dataset was streamlined to 25 features. This not only simplified the model but also enhanced its generalisation capabilities. By carefully selecting and engineering features, the developed models produced more efficient and precise predictions.

The AI Model selection involved developing, optimising and training eight pre-selected models and finally evaluated their performance on the preserved unseen test set. The result of these can be found in Table 2. Referring to Accuracy as the measure, four of the eight models outperformed our comparison measures presented in the article by Li et. all, they were XG Boost with 90.7%, Gradient Boost with 94.9%, Decision Tree with 85.6% and Ensemble Voting Classifier (which combined all of the other estimators) with 88.1%. It can be observed clearly that XGBoost, Voting Classifier and GBC can capture complex non-linear relationships in the data due to their ability to create ensembles of weak learners and combine their predictions. This allows them to effectively model complex patterns in the data, making them suitable for such a task compared to much simpler approaches such as SVMs and DTs.

In comparison to a neural network, XGBoost and GBC have built-in regularisation techniques, such as pruning of decision trees and early stopping, which help prevent over-fitting. Although we

performed hyper parameter tuning for the neural network to improve the performance, it was not able to outperform the tree-based ensemble models.

On the other hand, SVM, decision trees, and neural networks (such as MLP, Multi-Layer Perceptron) may have limitations in dealing with complex datasets with many features. Handling high-dimensional data, or preventing overfitting, reduced the final performance of the other models in the experiment. However, the goal was to produce a model that had the ability to outperform the model developed by Li et al. 2021. We belive the over-fitting problem could be reduced in the future with a larger dataset to work with.

Without question, the Gradient Boost Machine Classifier estimator, was the most accurate predictor for in-hospital mortality, boasting a test accuracy of 94.5%, and was able to improve on our teams' benchmark accuracy score (of 85%) by approximately 10%. Furthermore, a key metric our team aimed for - the model reduced the number of false positives, therefore, allowing the medical staff to provide the required care for critical patients promptly.

6 Conclusion

The project was to develop a prediction model for in-hospital mortality among ICU-admitted heart failure patients. Our team aimed to develop a superior model compared to previous research in this area. Using machine learning techniques, we sought to create a model that could improve clinical decision-making for at-risk heart failure patients and ultimately improve patient outcomes.

We have trained a number of different models for this particular problem and the Gradient Boosting Classifier (GBC) model demonstrated high accuracy of 94.9%, indicating the potential of machine learning in predicting patient outcomes as soon as their results are available. The GBC model was evaluated against the GWTG-HF score benchmark and showed the best f1-score of 0.86, making it the optimal model for low false negatives. However, after extensive analysis and evaluation, the GBC model was selected as the final model for its concise and wider net benefit threshold probability range.

Due to the imbalanced class in our test set, we primarily used the f1-score and precision-recall curve to evaluate our model's performance. The GBC model's wider net benefit threshold probability range makes it a valuable tool for healthcare professionals to optimize clinical decision-making for at-risk heart failure patients.

While the developed machine learning-based prediction model demonstrated high accuracy in predicting in-hospital mortality among ICU-admitted heart failure patients, there is always room for improvement. One potential area for improvement is to expand the feature set to include additional variables that may impact patient outcomes. Additionally, the model could be further validated using data from external sources to assess its generalizability to other patient populations. A key area of improvement could be training the GBM model with the latest data weekly to enhance the model's up to date accuracy.

Overall, our machine learning-based prediction model presents a valuable contribution to the medical field by leveraging technology to improve clinical outcomes for patients with heart failure.

References

Cardiology, American College of (2018). "2017 Guideline for High Blood Pressure in Adults." In: URL: https://www.acc.org/latest-in-cardiology/ten-points-to-remember/2017/11/09/11/41/2017-guideline-for-high-blood-pressure-in-adults.

Clinic, Mayo (2022). "Fever." In: URL: https://www.mayoclinic.org/diseases-conditions/fever/symptoms-causes/syc-20352759.

Li, Fuhai, Hui Xin, Jidong Zhang, Mingqiang Fu, Jingmin Zhou, and Zhexun Lian (July 2021). *Prediction model of in-hospital mortality in intensive care unit patients with heart failure: machine learning-based, retrospective analysis of the MIMIC-III database*. en.

Organisation, World Health (2010). "A healthy lifestyle - WHO recommendations." In: URL: https://www.who.int/europe/news-room/fact-sheets/item/a-healthy-lifestyle---who-recommendations.

Peterson, Pamela N., John S. Rumsfeld, Li Liang, Nancy M. Albert, Adrian F. Hernandez, Eric D. Peterson, Gregg C. Fonarow, and Frederick A. Masoudi (Jan. 2010). "A Validated Risk Score for In-Hospital Mortality in Patients With Heart Failure From the American Heart Association Get With the Guidelines Program." In: *Circulation: Cardiovascular Quality and Outcomes* 3.1, pp. 25–32. DOI: 10.1161/circoutcomes.109.854877. URL: https://doi.org/10.1161/circoutcomes.109.854877.

Publishing, Harvard Health (2020). "The respiratory rate: how fast you breathe." In.

Richard Whaddon Parsons Danny Liew, A. Munro Neville (2020). "The epidemiology of heart failure in the general Australian community - study of heart failure in the Australian primary carE setting (SHAPE): methods." In: *BMC Public Health*. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7216401/.

Scikit-Learn (2023). "Tuning the hyper-parameters of an estimator." In: URL: https://scikit-learn.org/stable/modules/grid_search.html.

Appendices

1. Scatter Plots

   Searching for features which have a hint of linear separation in data w.r.t outcome (dead, alive). The following plots show from early development, if any feature combinations could present some kind of linear separation - visually. From these (the best) we can see there is not really a clear linear separation.

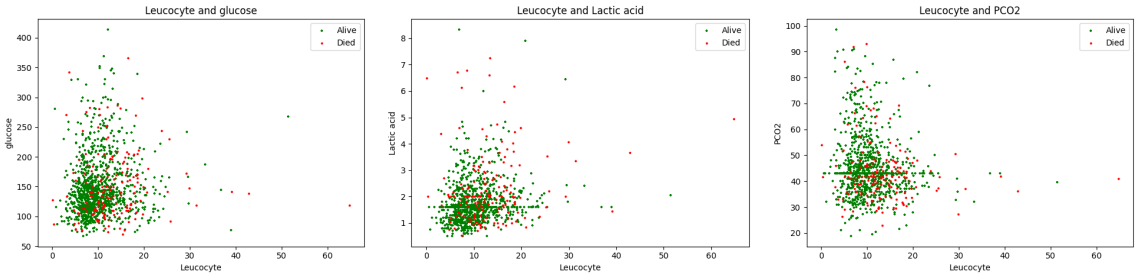


Figure 10: Lecocyte alongside; glucose, lactic acid and PC02

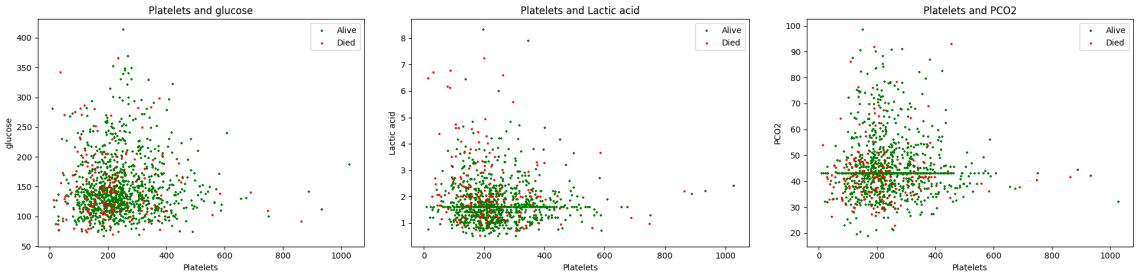


Figure 11: Platelets alongside; glucose, lactic acid and PC02

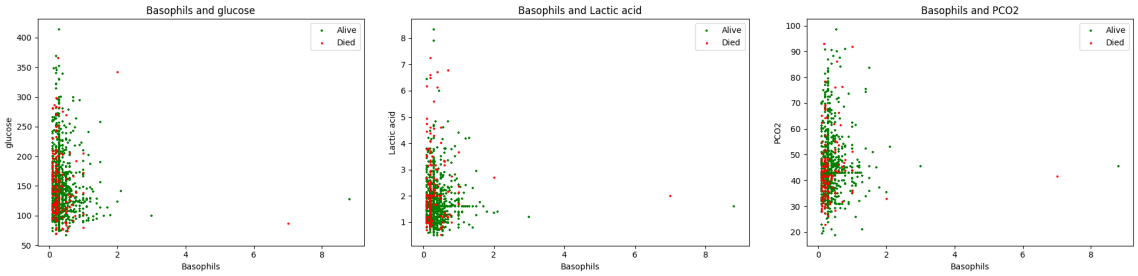


Figure 12: Basophils alongside; glucose, lactic acid and PC02

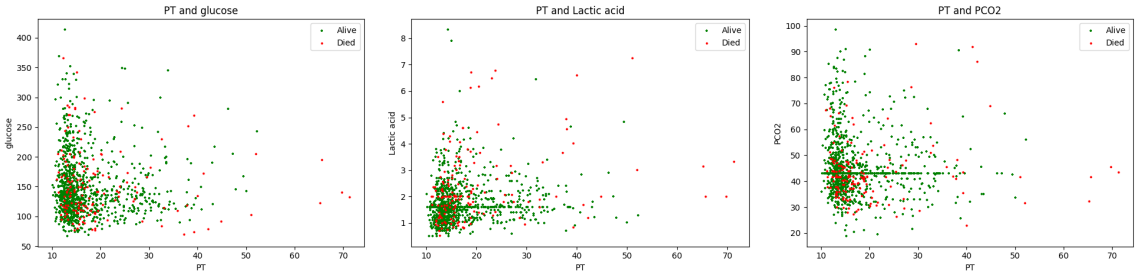


Figure 13: PT alongside; glucose, lactic acid and PC02

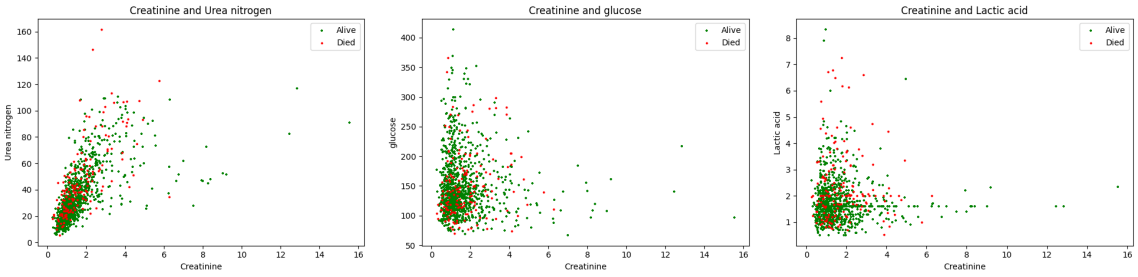


Figure 14: Creatinine alongside; urea nitrogen, glucose and lactic acid

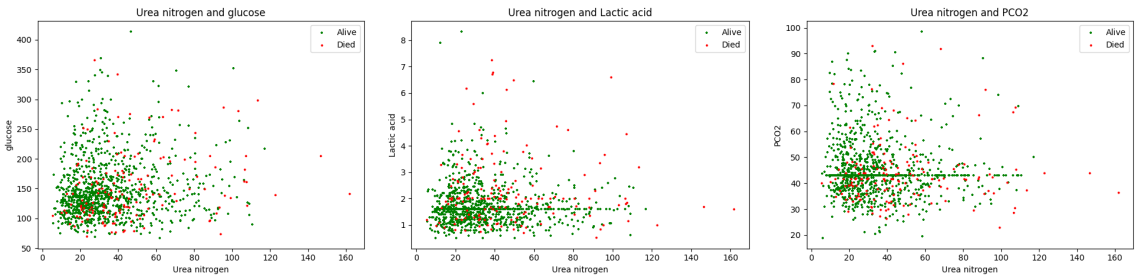


Figure 15: Urea Nitrogen alongside; glucose, lactic acid and PC02