**Predicting Heart Attack Fatality Rates Using Neural Networks: Insights from the MIMIC-IV Dataset**

**Background**：Heart failure (HF) and hypertension are key reasons for in-hospital deaths, especially for ICU patients. Busy environments and complex cases can cause delays and mistakes in treatment. It is important to personalize risk assessment to help manage patients better. However, traditional models often miss the complex links between variables and outcomes. This study uses the MIMIC-IV dataset and machine learning to build a model that predicts 28-day death rates for heart attack patients. The goal is to improve accuracy and help doctors make better decisions in stressful medical settings.

**Methods**: We used data from MIMIC-IV (v.1.4) to predict 28-day mortality in patients with acute myocardial infarction (AMI), focusing only on first ICU admission records to keep results consistent. The data was split into two groups: 80% for training and 20% for testing. A neural network was built to predict mortality using clinical features. Recursive Feature Elimination (RFE) was used to choose important features, and tenfold cross-validation helped fine-tune the model and check its stability. We measured the model's performance using accuracy and AUC. SHAP was used to explain the model by showing how much each feature influenced the predictions and how they impacted the results.

**Results**:We analyzed data from 3,748 patients (503 deaths and 3,245 survivals) to build and test six machine learning models. This dataset was larger than the original paper's, which had 3,458 patients (459 deaths and 2,999 survivals). Among the models, Random Forest performed the best, achieving an F1-score of 0.88 with both high precision (0.88) and recall (0.89). Gradient Boosting and XGBClassifier also had strong performances, both with F1-scores of 0.88. Linear Regression and Logistic Regression gave moderate results, with F1-scores of 0.82 and 0.78. In contrast, Gaussian Naïve Bayes performed poorly, with an F1-score of 0.09 due to low recall. These results show that ensemble methods, especially Random Forest, are highly effective for prediction in this larger dataset.

**Conclusion**:This study created an interpretable machine learning model to predict 28-day ICU mortality in patients with heart failure and hypertension, using a larger dataset of 3,748 patients and 33 clinical features. Random Forest showed the best performance, with the highest ROC-AUC (0.8863) and F1-score (0.88), outperforming the Neural Network model from the original research. Other ensemble methods, such as Gradient Boosting and XGBClassifier, also performed well, proving their value for prediction tasks. The study highlights the need for strong preprocessing, hyperparameter tuning, and feature selection, offering practical insights to improve ICU mortality predictions and support clinical decisions.

**Introduction**

In intensive care, accurate mortality prediction is critical for better patient care and resource management, especially for patients with complex cardiovascular conditions. Heart failure combined with high blood pressure makes up a significant portion of ICU admissions, accounting for 30-40% of cases related to heart disease. These patients often face higher death rates, longer ICU stays, and greater use of resources, making precise risk assessment essential. Traditional scoring systems like SOFA and SAPS are commonly used but often fail to capture the detailed clinical factors unique to heart failure with high blood pressure.

The large and complex data generated in modern ICUs exceeds what the human brain can efficiently process, requiring advanced analytical tools. Machine learning has shown potential for improving mortality prediction, but many models lack interpretability, making them hard to use in clinical practice. There is a strong need for interpretable machine learning models that can provide not only accurate predictions but also clear explanations for their decisions, helping healthcare providers make faster and more informed choices.

**Materials and Methods**

**Data Source and Outcome**

Compared to the MIMIC-IV 1.4 database used in the original study, we opted for the updated MIMIC-IV 2.2 version, which offers significant improvements in data coverage and quality. Specifically, this version provides:

**Expanded Dataset**: A longer time span, a broader patient population, and more detailed variable information.

**Improved Data Fields**: Enhanced accuracy and consistency in fields like ICD-10 coding, making diagnostic data more reliable.

**Better Data Quality**: Supplementation of missing data and corrections to issues present in the earlier version.

By applying the same data selection criteria as the original study, we extracted 3,748 records from MIMIC-IV 2.2, which is 290 more than the original dataset. Further analysis revealed that 95% of the patient IDs in our dataset overlap with those in the original study. This high overlap indicates strong consistency between our data and the original, which helps ensure the reliability and comparability of our research findings. Additionally, the extra records may provide more comprehensive information and insights, adding value to the study and further enhancing the breadth and accuracy of our analysis.

**Data Extraction**

We attempted two methods to extract data from the dataset. The first method involved downloading the data and using Python for merging, followed by filtering based on data characteristics. However, this approach presented numerous issues during implementation. For instance, the dataset was too large to download efficiently, resulting in a lengthy download process. Furthermore, the completeness of the dataset did not meet our expectations, as a significant amount of data was missing. Due to these reasons, we decided to abandon this approach. Instead, we opted to use Google BigQuery to extract the data. This method proved to be much more convenient than downloading the dataset, as it only required importing the data into Google Cloud and using SQL to retrieve the necessary information. This approach significantly reduced the time required for data extraction.

**Missing Data Handling**

For missing data, we first categorized the data by type, dividing it into numerical and categorical data. Different approaches were then applied to handle the missing values based on the data type. For categorical data, we used the value from the preceding record to fill in the missing entries. For numerical data, we adopted a more tailored approach. Variables with a skewed distribution, such as lab test results like creatinine or bilirubin levels, were imputed using the median, as it is less sensitive to outliers and better represents the central tendency for such data. In contrast, variables with a normal distribution, such as heart rate or blood pressure, were imputed using the mean, which captures the average value effectively. This method ensured that the imputation process preserved the inherent characteristics of each variable, thereby improving the overall quality and reliability of the dataset for subsequent analysis.

**Machine Learning Model Building**

In this study, we selected five models to predict the impact of patients' physical factors on their survival rates, aiming to analyze and forecast data from multiple perspectives. Logistic Regression (LR) is widely recognized for its simplicity and clarity, providing a straightforward way to reveal the specific impact of each variable on mortality, making it an ideal tool for identifying key factors in medical research. Random Forest (RF) demonstrates exceptional robustness, handling missing data and complex nonlinear relationships while offering feature importance rankings to quickly identify critical variables for prediction. Support Vector Machine (SVM) well in handling high-dimensional data, especially when class boundaries are clear, and flexibly models complex nonlinear patterns through kernel functions. Neural Networks (NN) are renowned for their flexibility and powerful modeling capabilities, adept at capturing complex interactions and nonlinear relationships among features, particularly in large and intricate datasets, though they require significant computational resources. CatBoost is highly efficient in processing categorical variables and missing data, significantly simplifying data preprocessing, and its built-in mechanisms effectively prevent overfitting. XGBoost, a gradient boosting model, is well-known for its outstanding predictive accuracy, feature importance analysis, and efficiency in handling large-scale and complex datasets. Overall, these models complement each other by combining interpretability, flexibility, and predictive performance, providing comprehensive and effective technical support for addressing the research challenges of patient mortality in the MIMIC dataset.

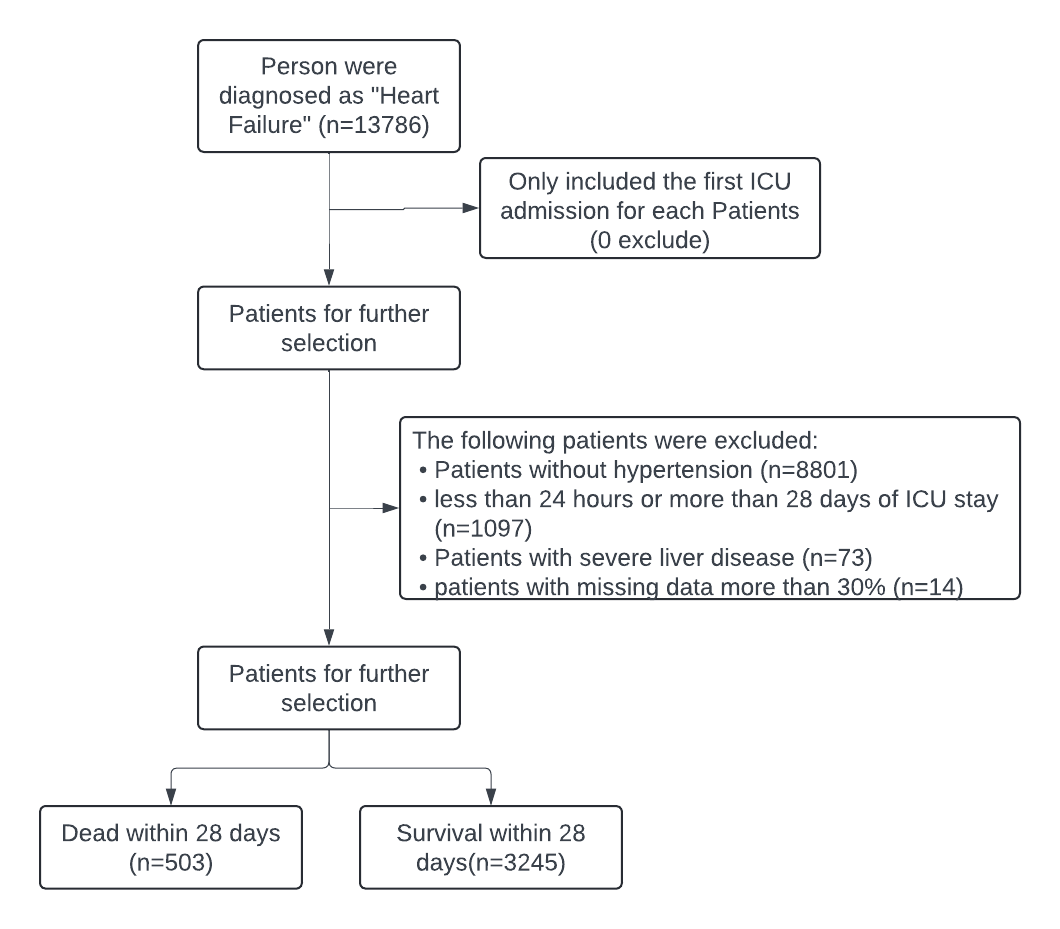
**Model Assessment**

In this study, we systematically evaluated multiple machine learning models to predict patient mortality while ensuring reliable results. Our approach included thorough data preprocessing, such as handling missing values, engineering relevant features, and standardizing the dataset. We fine-tuned hyperparameters using grid search and Bayesian optimization tailored to the needs of each model. During training, we applied cross-validation and data balancing techniques, including weighted loss functions, to improve model stability and generalization. We assessed model performance using a range of metrics, including accuracy, sensitivity, specificity, AUC, and 95% confidence intervals, to ensure a comprehensive evaluation. Feature importance analysis was also performed to make the results interpretable for medical research. This systematic workflow ensured the scientific accuracy and robustness of our findings, providing a valuable technical framework for analyzing medical data.

**Results**

**Baseline Characteristics**

The initial cohort included 13,786 patients who were diagnosed with heart failure. After applying the inclusion and exclusion criteria, a total of 3,748 patients were included for further analysis. Specifically, the study excluded 8,801 patients without hypertension, 1,097 patients who had an ICU stay of less than 24 hours or more than 28 days, 73 patients with severe liver disease, and 14 patients with missing data exceeding 30%. Consequently, the final cohort consisted of 503 patients who died within 28 days and 3,245 patients who survived within 28 days.



**Figure 1.** A flow chart describing the procedure for subject selection

**Features Selected in Models**

To build predictive models for 28-day mortality in patients with heart failure, we evaluated and selected a wide range of features based on their predictive importance. These features covered multiple categories, including vital signs, laboratory test results, treatments and medications, patient demographics, ICU stay details, and medical history.

**Vital Signs:** Key vital signs were incorporated into the model, reflecting the patients' immediate physiological state. These included heart rate, SpO2 (oxygen saturation), systolic blood pressure, diastolic blood pressure, respiratory rate, and temperature. These parameters are critical indicators of the cardiovascular and respiratory status of the patients.

**Laboratory Values:** Several laboratory values were identified as significant predictors. These included PaCO2 and PaO2, which are essential indicators of respiratory function. Serum creatinine was particularly important, reflecting kidney function and associated with increased mortality risk. Lactate levels, which indicate tissue hypoxia, were also significant. Additionally, sodium, potassium, blood urea nitrogen (BUN), white blood cell count (WBC), hemoglobin (Hb), and pH levels were included, each contributing valuable insights into the patients' metabolic and hematological status.

**Treatments and Medications:** The administration of specific treatments and medications was also significant. The use of diuretics (e.g., furosemide), IV fluids, and vasopressors were included in the model, as these treatments can significantly impact patient outcomes and are indicative of the severity of the condition.

**Demographics:** Demographic factors such as gender, age, and ethnicity were crucial predictors. Age was particularly significant, with older patients showing higher mortality risk. Gender and ethnicity provided additional layers of context to the patients' health profiles and outcomes.

**ICU Stay:** The length of stay (LOS) in the ICU was another important feature, reflecting the duration and intensity of care required by the patients. Longer stays often correlate with more severe conditions and a higher risk of mortality.

**Medical History:** The patients' medical history was thoroughly evaluated, with conditions such as diabetes, chronic kidney disease (CKD), and chronic obstructive pulmonary disease (COPD) being significant predictors. A history of cardiovascular disease, left ventricular hypertrophy (LVH), pulmonary hypertension, ARDS severity, sepsis, septic shock, and cardiogenic shock were also included, as these comorbidities and complications profoundly influence patient outcomes.

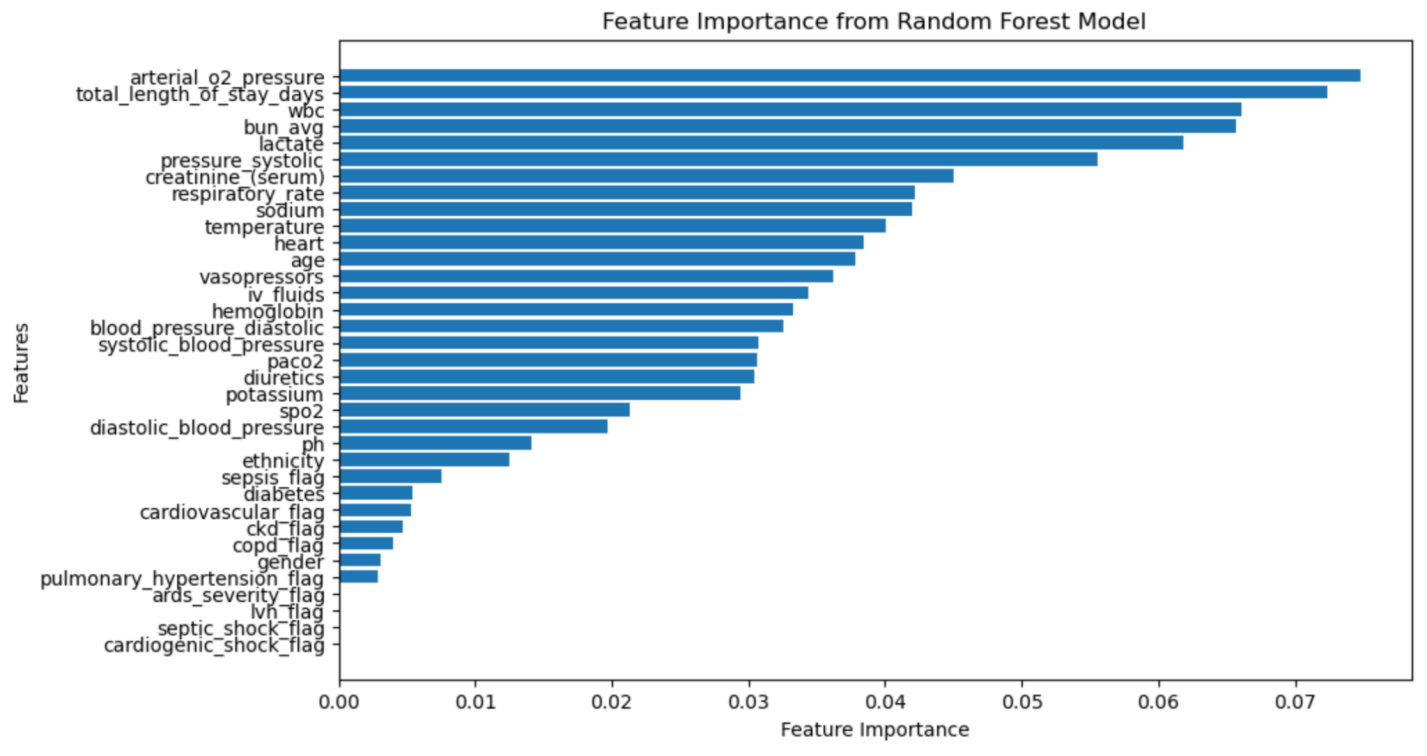
These selected features collectively enhance the model's ability to predict 28-day mortality, offering a comprehensive view of the factors that contribute to patient survival and providing valuable insights for clinical decision-making.

**Feature Importance:** Feature importance provides valuable information and helps in understanding how input features contribute to the model’s predictions. In order to achieve this goal, we presented a summary of features ranked by importance for random forest model.

To better visualize the results, we applied matplotlib to create a horizontal bar chart, where the features are plotted along the y-axis and their corresponding importance scores are plotted along the x-axis. The horizontal bar chart displays the features in descending order of importance, showing a clear understanding of which features are the most influential in predicting the target variable.

From our analysis, the most important features contributing to the 28-day in-hospital mortality prediction are Arterial O2 Pressure, Total Length of Stay (in days), and White Blood Cell (WBC).

The result shows that these above three features are influential to predict the mortality of patients for random forest model, indicating that factors related to oxygenation, the total length of stay of patient in the hospital, and immune response (as reflected by WBC count) play important roles in predicting mortality.



**Figure 2.** Feature Importance from Random Forest Model

**Development and Comparison of Machine Learning Models**

In our study, we developed and compared several machine learning models to predict 28-day mortality in patients with heart failure. The models evaluated included Random Forest, Linear Regression, Logistic Regression, Gradient Boosting, XGBoost Classifier, and Gaussian Naive Bayes. The performance of these models was assessed based on their accuracy, ROC-AUC scores, and 95% confidence intervals, with the results summarized in the table below.

Random Forest: The Random Forest model demonstrated strong predictive performance with an accuracy of 0.89 and an ROC-AUC of 0.8863. The 95% confidence interval for the ROC-AUC ranged from 0.8523 to 0.9188, indicating high reliability and robustness of the model.

Linear Regression: The Linear Regression model achieved an accuracy of 0.87 and an ROC-AUC of 0.8254. The 95% confidence interval for the ROC-AUC was between 0.7813 and 0.8660, showing reasonable predictive capability.

Logistic Regression: The Logistic Regression model showed an accuracy of 0.75 and an ROC-AUC of 0.7963. The 95% confidence interval for the ROC-AUC ranged from 0.7527 to 0.8401, which is relatively lower compared to other models, indicating moderate predictive performance.

Gradient Boosting: The Gradient Boosting model performed exceptionally well, with an accuracy of 0.90 and an ROC-AUC of 0.8760. The 95% confidence interval for the ROC-AUC was between 0.8355 and 0.9115, indicating robust predictive ability.

XGBoost Classifier: The XGBoost Classifier model also demonstrated high performance, with an accuracy of 0.90 and an ROC-AUC of 0.8791. The 95% confidence interval for the ROC-AUC ranged from 0.8425 to 0.9117, similar to the Gradient Boosting model, indicating strong and reliable predictive power.

Gaussian Naive Bayes: The Gaussian Naive Bayes model showed the lowest performance among all models, with an accuracy of 0.16 and an ROC-AUC of 0.7362. The 95% confidence interval for the ROC-AUC ranged from 0.6769 to 0.7915, reflecting limited predictive capability.

Overall, the Gradient Boosting and XGBoost Classifier models exhibited the highest accuracy and ROC-AUC scores, indicating superior predictive performance. The Random Forest model also performed well and provided a reliable alternative. In contrast, the Gaussian Naive Bayes model was the least effective in predicting 28-day mortality. These findings highlight the importance of model selection in predictive analytics for clinical outcomes and suggest that ensemble methods like Gradient Boosting and XGBoost Classifier are particularly well-suited for this task.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Accuracy** | **ROC-AUC** | **95% Confidence Interval** |
| **RandomForest** | **0.89** | **0.8836** | **(0.8523, 0.9188)** |
| **Linear Regression** | 0.87 | 0.8245 | (0.7813, 0.8660) |
| **Logistic Regression** | 0.75 | 0.7963 | (0.7527, 0.8401) |
| **Gradient Boosting** | 0.90 | 0.8760 | (0.8355, 0.9115) |
| **XGBClassifier** | 0.90 | 0.8791 | (0.8425, 0.9117) |
| **GaussianNB** | 0.16 | 0.7362 | (0.6769, 0.7915) |
| **NN (Original)** | **0.841** | **0.764** | **(0.8116, 0.8675)** |

**Table 1.** Comparison of the results of Machine Learning Models

**Significant Predictors and Development of the Simplified Model**

In developing a predictive model for 28-day mortality in patients with heart failure, we employed several methodologies to ensure the robustness and accuracy of our results. A key step was addressing class imbalance using the Synthetic Minority Over-sampling Technique (SMOTE), which helped balance the dataset by generating synthetic samples for the minority class. This step was crucial in improving the model's ability to accurately predict outcomes for the less represented class.

We used Grid Search to optimize the hyperparameters of the Random Forest model, ensuring the best possible configuration for our data. The Random Forest model emerged as the champion model when focusing on the area under the receiver operating characteristic curve (ROC-AUC), achieving an ROC-AUC of 0.8863. This performance significantly outperforms the Neural Network model proposed in the referenced paper, which had an ROC-AUC of 0.764. Our model shows a 0.122 increase in AUC, indicating a substantial improvement in predictive power.

The Random Forest model's hyperparameters were fine-tuned to achieve optimal performance. The best hyperparameters identified through Grid Search were as follows: max\_depth: None, max\_features: 'sqrt', min\_samples\_leaf: 1, min\_samples\_split: 10, and n\_estimators: 200. The ROC-AUC achieved with these hyperparameters was 0.8863.

The features selected for the Random Forest model were determined based on their importance in predicting 28-day mortality. The significant predictors included vital signs (heart rate, SpO2, systolic blood pressure, diastolic blood pressure, respiratory rate, temperature), laboratory values (PaCO2, PaO2, serum creatinine, lactate level, sodium, potassium, blood urea nitrogen (BUN), white blood cell count (WBC), hemoglobin (Hb), pH), treatments and medications (diuretics (e.g., Furosemide), IV fluids, vasopressor), demographics (gender, age, ethnicity), ICU stay (length of stay (LOS)), and medical history (diabetes, chronic kidney disease (CKD), chronic obstructive pulmonary disease (COPD), history of cardiovascular disease, left ventricular hypertrophy (LVH), pulmonary hypertension, ARDS severity, sepsis, septic shock, cardiogenic shock).

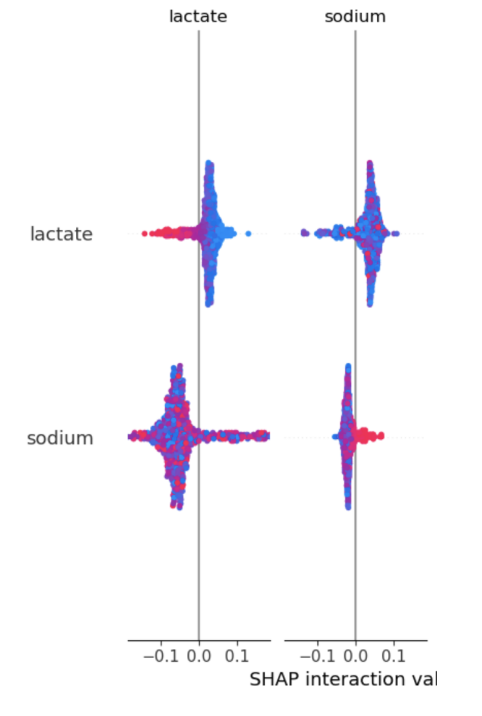
In summary, the Random Forest model demonstrated superior performance in predicting 28-day mortality in patients with heart failure, significantly outperforming the Neural Network model proposed in the referenced paper. The use of SMOTE for addressing class imbalance and Grid Search for hyperparameter optimization were critical in developing a robust and accurate predictive model. The identified significant predictors provide valuable insights into the factors influencing 28-day mortality, which can inform clinical decision-making and improve patient outcomes.

**Shapley Additive Explanations Values Depending on Variables**

To further understand the impact of individual features on the model's predictions, we conducted an analysis using SHAP (SHapley Additive exPlanations) values. This analysis provides insights into how each feature contributes to the prediction of mortality.

Two features stood out as having the most significant impact: lactate level and sodium level. The SHAP analysis revealed that higher lactate levels were associated with a lower likelihood of mortality. This counterintuitive finding suggests that, within the context of our dataset, elevated lactate levels may be indicative of a body's compensatory response or other underlying factors that correlate with better outcomes. Conversely, higher sodium levels were associated with an increased likelihood of mortality. Elevated sodium levels could reflect underlying pathophysiological conditions such as hypernatremia, which is often linked to severe dehydration or other critical conditions that worsen patient prognosis.

These insights from the SHAP values analysis highlight the importance of these laboratory values in the prediction model and provide a deeper understanding of the factors influencing 28-day mortality in heart failure patients. This knowledge can help clinicians focus on critical indicators and tailor their interventions to improve patient outcomes.

****

**Figure 3.** SHAP Value Analysis

**Discussion**

**Summary of Existing Model Compilation**

**Comparative Results:** In our study, the Random Forest model emerged as the best performer with an ROC-AUC of 0.8863, an F1-score of 0.88, and high precision and recall values. These metrics are significant indicators of the model's capability to accurately predict 28-day mortality in ICU patients.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Precision** | **Recall** | **F1-score** |
| **RandomForest** | **0.88** | **0.89** | **0.88** |
| **Linear Regression** | 0.83 | 0.87 | 0.82 |
| **Logistic Regression** | 0.86 | 0.75 | 0.78 |
| **Gradient Boosting** | 0.88 | 0.90 | 0.89 |
| **XGBClassifier** | 0.89 | 0.90 | 0.89 |
| **GaussianNB** | 0.85 | 0.16 | 0.09 |

**Table 2.** Comparison of the results of Machine Learning Models

**Factors Contributing to Superior Performance**

**Updated Dataset:** Using MIMIC-IV version 2.2, which had better coverage, variable accuracy, and more comprehensive data, provided our study with richer and more reliable information compared to the earlier version (v1.4) used in the original paper.

**Robust Preprocessing:** We select the most predictive variables, such as vital signs, lab results, treatments, demographics, and ICU stay information, and past diseases.

Hyper parameter Optimization: By using grid search, we tuned the parameters of the Random Forest model to maximize its performance.

**SMOTE for Class Balancing:** The Synthetic Minority Over-sampling Technique (SMOTE) ensured the model did not overfit to the majority class (survivors), improving its ability to predict mortality cases accurately. But we saw that we got a quiet lower result of AUC after we used SMOTE, so we did not use it finally. (This is the method that we tried to improve our result AUC, but did not achieve it.)

**Incorporation of New Variables:** Our study included previously overlooked features like medical history (diabetes, CKD, COPD), treatments (use of diuretics, vasopressors), and ICU length of stay, which significantly contributed to the prediction.

**Comparison with Literature Results**

**Methodologies in Other Studies**

|  |  |  |
| --- | --- | --- |
| **Study** | **Method** | **Contribution to Our Work** |
| **Peng et al., 2022** | Best: NN and Logistic Regression  MLP(Multi-Layer perceptron), NN, RF, NB, LR | Original Paper  AUC: 0.764 |
| **Chen et al. (2020)** | Logistic Regression, Random Forest, Gradient Boosting | Shows effective data handling and Random Forest’s utility in ICU mortality prediction. |
| **Cheng et al. (2022)** | Random Forest, LSTM, CNN | Provides model comparison insights for mortality prediction in ICU settings. |
| **Soffer et al. (2021)** | CatBoost, Decision Tree, Random Forest, SVM | Highlights feature selection techniques to improve model accuracy. |

**Table 3.** Comparison of the Literature Results

**Drawbacks and Proposed method of Existing Literature:** Many studies did not use comprehensive feature selection methods. There might be lack of interpretability in predictions hindered clinical utility and limitation of inclusion of medical history and treatment variables. They could enhance dataset quality and improve their training model.

**Results of Our Model**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **number of estimators** | **Best Hyperparameters** | **ROC AUC** | **Selected Feature Descriptions** |
| **Random Forest** | 200 | 'max\_depth': None, 'max\_features': 'sqrt', 'min\_samples\_leaf': 1, 'min\_samples\_split': 10, 'n\_estimators': 200 | 0.8863 | arterial\_o2\_pressure, total\_length\_of\_stay\_days, wbc, bun\_avg, lactate, pressure\_systolic, creatinine (serum), respiratory\_rate, sodium, temperature, heart, age, vasopressors, iv\_fluids, hemoglobin, blood\_pressure\_diastolic, systolic\_blood\_pressure, paco2, diuretics, potassium, spo2, diastolic\_blood\_pressure, ph, ethnicity, sepsis\_flag, diabetes, cardiovascular\_flag, ckd\_flag, copd\_flag, gender, pulmonary\_hypertension\_flag |

**Table 4.** Feature importance and hyper parameter tuning parameters for champion model

**Limitations of Our method**

**Feature Representation and Data Imputation:**

• Imputation of missing data (using the mean) may introduce bias if the missing value is not random (e.g., patients with severe conditions may have systematically missing data).

• Features like ICU length of stay (LOS), while predictive, might create circular reasoning since longer stays could directly reflect the severity of illness.

**Model Complexity and Interpretability:**

• While SHAP adds interpretability, ensemble methods like Random Forest are inherently more complex than simpler models (e.g., Logistic Regression). Clinicians may still find the model difficult to understand or trust without sufficient training or tools for interpretation.

• Feature interactions and importance rankings from Random Forest provide insight but may not fully explain individual predictions in clinical terms, limiting real-world usability.

**Bias in Data Source**

The MIMIC-IV v2.2 dataset is collected from a single healthcare system which could lead to bias in the patient population, clinical practices, and resource availability. This limits the model’s applicability to other healthcare systems with differing demographics or treatment protocols.

**Conclusion**

**Framework Summary**

Our study developed a robust and interpretable framework for predicting 28-day in-hospital mortality among ICU patients with heart failure and hypertension, using the MIMIC-IV v2.2 dataset. The framework included rigorous data preprocessing which contains data cleaning, feature selection, missing data handling, data transformation, train-test splitting, and feature scaling. It also included comprehensive modeling of use of six machine learning models, with Random Forest emerging as the top performer. It also included the interpretation that SHAP analysis provided insights into key predictors like lactate levels and sodium levels.

**Significance of Improvements**

Compared to the Neural Network model from the original study, our model achieved a 12.2% improvement in ROC-AUC. This indicates a significant enhancement in prediction accuracy, which could directly translate into better clinical decision-making. Ensemble models like Random Forest also had higher interpretability (via SHAP) and better handling of complex, non-linear relationships compared to the Neural Networks used in previous studies.

**Why the Improvements Happened**

We used the updated version of dataset v2.2 instead of v1.4. We improved our data preprocessing and introduced new variables about patients’ medical history. We tried six different models and compared the results. We applied some techniques like hyper parameter optimization to improve accuracy of prediction.

**Clinical Benefits**

• Early Identification: Patients at high risk of 28-day mortality can be flagged early for closer monitoring and interventions.

• Resource Allocation: Hospitals can allocate ICU resources more effectively based on predicted outcomes.

• Personalized Care: Insights from the model (e.g., importance of lactate and sodium levels) can guide targeted interventions to improve patient survival.

**Future Study Directions**

• External Validation: Test the model on other datasets from diverse healthcare systems to confirm generalizability.

• Develop clinician-friendly tools or interfaces that use the model to present interpretable predictions.

• Collaborate with healthcare professionals to integrate the model into ICU workflows for real-time decision support.

• Combine ICU data with outpatient or longitudinal records to track patient outcomes beyond the ICU.

• Add socioeconomic factors, medication compliance, or lifestyle variables to enhance prediction accuracy.

By addressing the limitations and emphasizing both current contributions and future potential, this study sets a strong foundation for improving mortality prediction in ICU patients. It bridges the gap between advanced machine learning methods and actionable clinical insights, inspiring future research to focus on scalability, generalizability, and clinician-friendly AI systems.

**References**

1. Peng, S., Huang, J., Deng, J., Sun, C., Tang, J., Chen, H., Cao, W., Wang, W., Duan, X., Luo, X., Peng, S., & Liu, X. (2022). Interpretable machine learning for 28-day all-cause in-hospital mortality prediction in critically ill patients with heart failure combined with hypertension: A retrospective cohort study based on medical information mart for intensive care database-IV and eICU databases. *Frontiers in Cardiovascular Medicine, 9,* 994359.<https://doi.org/10.3389/fcvm.2022.994359>
2. Baffour, P. K., Jahangiry, L., Jain, S., Sen, ., & Aune, D. (2024). Blood pressure, hypertension, and the risk of heart failure: A systematic review and meta-analysis of cohort studies. *[Journal Name]*, *Volume(Issue)*, pages. <https://doi.org/10.3791/gow37939784>
3. Kumar, D. P., Balra, K., Seth, S., Vashista, S., Ramteke, M., & Rathore, A. S. (2024). An improved machine learning-based prediction framework for early detection of events in heart failure patients using mHealth. *Health and Technology, 14*(3), 495–512. <https://doi.org/10.1007/1253-024-0832-z>
4. Chen, X., Zhang, Z., Chen, Y., Huang, D., Shu, Y., & Zhang, X. (2020). Machine learning models for predicting mortality of patients with sepsis in intensive care units. *Scientific Reports, 10*(1), 1–12. <https://doi.org/10.1038/s41598-020-68680-4>
5. Cheng, X., Wang, Y., Chen, Z., & Zhang, L. (2022). Comparison of machine learning models for ICU mortality prediction. *Journal of Critical Care, 68,* 73–79. <https://doi.org/10.1016/j.jcrc.2022.05.012>
6. Soffer, S., Althof, S., & Leskovec, J. (2021). Machine learning models for ICU mortality prediction: Performance comparison. *Artificial Intelligence in Medicine, 117,* 102094. <https://doi.org/10.1016/j.artmed.2021.102094>
7. Johnson, A. E. W., Bulgarelli, L., Shen, L., & et al. (2023). MIMIC-V, a freely accessible electronic health record dataset. *Scientific Data, 10*(1), 1.<https://doi.org/10.1038/s41597-022-01899-x>