30 Day ICU Readmissions Predictive Model Building and Analysis of Different Model Efficiencies

Quinn Ramsay & Crawford James

Project Code and Implementation:
[ICU Readmissions Model GitHub Repository]

Abstract

ICU readmissions present significant challenges for both patient outcomes and healthcare systems. This study utilized the MIMIC-IV dataset to develop and evaluate machine learning models, including Logistic Regression, Random Forest, and XGBoost, using structured clinical data such as demographics, laboratory results, and comorbidities. Model performance was assessed through standard predictive metrics as well as the potential financial impact of their predictions. The study highlights the utility of machine learning in identifying patients at risk of ICU readmission and underscores the importance of integrating clinical and operational considerations when developing predictive models for healthcare applications.

Overview

ICU readmissions are linked to a six- to sevenfold increase in the odds of death, independent of other patient- or hospital-related factors. This challenge is evident both nationally and locally, particularly in the Piedmont region of North Carolina, where Surgical Trauma ICU readmissions remain a significant concern. In this area, respiratory failure is the most common predictor of Surgical Trauma ICU readmission.

Previous studies report readmission rates as high as 13.4% within just seven days, which occurs well before the full 30-day window examined in our study. Among Medicare beneficiaries, one in five patients is readmitted within 30 days, and 67% are readmitted within 90 days. These readmissions not only pose serious risks to patient health but also place a substantial financial strain on healthcare systems. In 2004 alone, avoidable Medicare readmissions cost more than $17 billion. Nationally, the annual cost of patient readmissions exceeds $52.4 billion, with an average cost of approximately $15,200 per readmission.

Methods

Logistic Regression is a method used to predict the probability of a binary outcome, such as yes/no or 1/0. It achieves this by applying the logistic function to model the relationship between variables in the dataset, transforming the output into a probability ranging from 0 to 1. A decision threshold is then applied to classify outcomes; in this case, the threshold was set at 0.52. Logistic Regression models produce a sigmoid-shaped curve between 0 and 1, representing the likelihood of an event, such as a patient being readmitted to the ICU.

Random Forest is a machine learning method used for classification and regression. It generates multiple decision trees during training and combines their outputs to improve prediction accuracy and reduce overfitting. Each tree is trained on a random subset of the data and considers a random subset of features when making splits. For classification tasks, the final prediction is determined by a majority vote across all trees, while for regression tasks, predictions are averaged. Random Forest models are particularly effective at capturing complex relationships in the data and can handle large datasets with diverse variables. In the context of ICU readmission, the model estimates the likelihood of a patient being readmitted, which can then be compared to a threshold to make a yes/no prediction.

XGBoost (Extreme Gradient Boosting) is an advanced machine learning algorithm that constructs a sequence of decision trees, with each new tree designed to correct the errors of the previous ones. The algorithm optimizes a loss function using gradient descent and incorporates regularization techniques to prevent overfitting. XGBoost is highly efficient and performs well on large datasets with numerous features. For ICU readmission prediction, the algorithm produces a probability score for each patient, representing the risk of readmission. This score can be compared to a selected threshold to classify patients as likely or unlikely to be readmitted. XGBoost often achieves higher predictive accuracy than single-tree models because it iteratively focuses on the cases that are most difficult to predict.

The Dataset

The MIMIC-IV dataset, version 3.1, available through PhysioNet, was selected for this project. Access to the dataset was obtained by completing the "Data or Specimens Only Research" certification, which authorizes researchers to work with medical data. The Medical Information Mart for Intensive Care (MIMIC-IV) is a large, de-identified dataset containing information on patients admitted to the emergency department (ED) or an intensive care unit (ICU) at the Beth Israel Deaconess Medical Center in Boston, Massachusetts. MIMIC-IV includes hospital and critical care records for patients admitted between 2008 and 2022, encompassing more than 65,000 ICU admissions and over 200,000 ED visits.

MIMIC-IV is organized into two main components: ICU and HOSP. The HOSP module contains hospital-wide data extracted from electronic health records, including demographics, admissions, and laboratory results. The ICU module provides detailed clinical information such as vital signs, medications, and fluid inputs. This comprehensive dataset supports the development of predictive models that incorporate multiple types of algorithms tailored to different data structures.

Results

When evaluating the performance of predictive models, there are several approaches commonly used in other studies. One approach focuses on patient outcomes, prioritizing the highest recall rate to minimize the number of patients readmitted by ensuring that at-risk patients are correctly identified. Another approach emphasizes the statistical performance of the model, optimizing metrics such as precision, accuracy, and the area under the receiver operating characteristic curve (AUC-ROC) to assess the overall quality and reliability of predictions. While both approaches can be valid, this project adopted a different methodology due to the inherent limitations of relying solely on these metrics.

In this project, model quality was assessed not only based on AUC-ROC, accuracy, and precision but also on the financial impact of the predictions. To calculate this, the confusion matrix was used, detailing the model's predictions as true positives, true negatives, false positives, and false negatives. Financial data obtained from previous research and studies were applied to determine the costs associated with patients not readmitted, the costs of false readmissions, and the savings achieved by preventing unnecessary readmissions. This information was then used to calculate the net financial savings relative to a scenario in which the model had not been implemented.

The following table summarizes the performance metrics and financial impact for each predictive model:

Logistic Regression

- AUC-ROC Score: 0.74
- Accuracy: 76%
- Precision: 59.5%
- Recall: 57%

- Original Cost: $28,196,000
- Comparative Savings: $16,315,200

Random Forest

- AUC-ROC Score: 0.744
- Accuracy: 76%
- Precision: 59%
- Recall: 56%
- Original Cost: $28,196,000
- Comparative Savings: $16,162,200

XGBoost

- AUC-ROC Score: 0.758
- Accuracy: 83%
- Precision: 62.5%
- Recall: 45%
- Original Cost: $28,196,000
- Comparative Savings: $16,811,000

Conclusion

This project assessed the performance of three predictive models, Logistic Regression, Random Forest, and XGBoost, using both statistical metrics and financial impact to evaluate model quality. XGBoost achieved the highest AUC-ROC score of 0.758 and accuracy of 83 percent, though its recall rate was lower at 45 percent compared to Logistic Regression at 57 percent and Random Forest at 56 percent. Logistic Regression demonstrated a balanced performance with strong recall and substantial financial savings of $16,315,200. Random Forest showed similar predictive performance but slightly lower savings of $16,162,200. XGBoost, while lower in recall, generated the highest comparative financial savings at $16,811,000. These results highlight the trade-offs between accuracy, recall, and financial impact in predictive modeling for ICU readmissions.

Overall, all three models demonstrated significant potential to reduce the costs associated with unnecessary readmissions, with comparative savings exceeding $16 million in each case. The findings emphasize that evaluating predictive models in healthcare requires considering both clinical outcomes and economic impact. Logistic Regression provided the most balanced approach between identifying at-risk patients and maximizing recall, while XGBoost excelled in overall predictive accuracy and achieved the highest financial savings. These insights underscore the importance of selecting models that align with both patient care objectives and operational priorities in healthcare settings.

References

[1] Alsentzer, E. (2020). *Bio_ClinicalBERT* [Pretrained language model]. HuggingFace. https://huggingface.co/emilyalsentzer/Bio_ClinicalBERT

[2] Beauvais, B., Whitaker, Z., Kim, F., & Anderson, B. (2022). *Is the hospital value-based purchasing program associated with reduced hospital readmissions?* Journal of Multidisciplinary Healthcare, 15, 1089–1099. https://doi.org/10.2147/JMDH.S358733

[3] Blinder, Y. (n.d.). *all_data.sql* [SQL script]. GitHub. https://github.com/YaronBlinder/MIMIC-III_readmission/blob/master/all_data.sql

[4] Deo, R. C. (2015). Machine learning in medicine. *Circulation, 132*(20), 1920–1930. https://doi.org/10.1161/CIRCULATIONAHA.115.001593

[5] Golas, S. B., Shibahara, T., Agboola, S., Otaki, H., Sato, J., Nakae, T., Hisamitsu, T., Kojima, G., Felsted, J., Kakarmath, S., Kvedar, J., & Jethwani, K. (2018). A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: A retrospective analysis of electronic medical records data. *BMC Medical Informatics and Decision Making, 18*(1), 44. https://doi.org/10.1186/s12911-018-0620-z

[6] Inan, O. T., Tenaerts, P., Prindiville, S. A., Reynolds, H. R., Dizon, D. S., Cooper-Arnold, K., Turakhia, M. P., Pletcher, M. J., Preston, K. L., Krumholz, H. M., Marlin, B. M., Mandl, K. D., & Califf, R. M. (2020). Digitizing clinical trials. *Journal of Personalized Medicine, 10*(2), 21. https://doi.org/10.3390/jpm10020021

[7] McNeill, H., & Khairat, S. (2020). Impact of intensive care unit readmissions on patient outcomes and the evaluation of the national early warning score to prevent readmissions: Literature review. *JMIR Perioperative Medicine, 3*(1), e13782. https://doi.org/10.2196/13782

[8] Moerschbacher, A., & He, Z. (2023). Building prediction models for 30-day readmissions among ICU patients using both structured and unstructured data in electronic health records. In *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine* (pp. 4368–4373). https://doi.org/10.1109/bibm58861.2023.10385612

[9] PhysioNet. (2023). *MIMIC-IV (version 3.1)*. https://physionet.org/content/mimiciv/3.1

[10] Toma, M., & Wei, O. C. (2023). Predictive modeling in medicine. *Encyclopedia, 3*(2), 590–601. https://doi.org/10.3390/encyclopedia3020042