International Conference on Social Science and Humanity
Homepage: https://proceedings.antispublisher.org/index.php/icossh
Email: admin@antispublisher.com

e-ISSN: 3063-9611
ICOSSH, Vol. 3, No. 5, 2026
Page 9-23
© 2026 ICOSSH

A Comparison of Random Forest and XGBoost Performance in Predicting 30-Day Readmission of Inpatients using the MIMIC-IV Dataset

Satriyo Kristanto[1], Yustyzia Priyatmaja Putri[2]
[1,2]Baptis Hospital Health Sciences College of Kediri, Indonesia

Check for updates
OPEN ACCESS
BY-SA

DOI: https://doi.org/10.61796/icossh.v3i5.645

Sections Info

Article history:
Submitted: September 30, 2026
Final Revised: October 15, 2026
Accepted: November 05, 2026
Published: November 23, 2026

Keywords:
Machine learning
Patient readmission
Random forest
XGBoost
SHAP

ABSTRACT

Objective: *Based on the results of a study using the MIMIC-IV dataset – which consists of 425 patient records after data cleaning – the objective of this study was to build predictive models for 30-day patient readmission using the Random Forest and XGBoost algorithms.* ***Method:*** *The variables used included patient age, gender, length of stay (LOS), admission type, insurance status, marital status, and race. The research stages included data preprocessing, creating a 30-day readmission label, encoding categorical variables, splitting the data into training and test sets at an 80:20 ratio, and evaluating the models using the metrics Accuracy, Precision, Recall, F1-Score, and Area Under the Curve (AUC-ROC).* ***Results:*** *The findings confirm that Random Forest delivered the best performance with an accuracy of 78.82%, precision of 55.00%, recall of 55.00%, an F1-score of 55.00%, and an AUC-ROC of 0.8354. Meanwhile, XGBoost yielded an accuracy of 77.65%, precision of 52.63%, recall of 50.00%, an F1-score of 51.28%, and an AUC-ROC of 0.8012. Analysis of the confusion matrix showed that Random Forest correctly identified 56 non-readmission cases and 11 readmission cases. Additionally, feature importance analysis indicates that length of stay (LOS) is the most influential factor in readmission prediction, followed by patient age and admission type.* ***Novelty:*** *To improve model interpretability, an Explainable Artificial Intelligence (XAI) analysis using SHAP was conducted, confirming the significant contribution of the LOS variable to readmission risk. The results of this study indicate that Random Forest outperforms XGBoost in predicting patient readmissions and has the potential to support clinical decision-making in efforts to prevent hospital readmissions.*

INTRODUCTION

Hospital readmission within 30 days after a patient's discharge is a key indicator in evaluating the quality of healthcare services, as it reflects the effectiveness of patient care during hospitalization, the quality of discharge planning, and the success of post-discharge follow-up. High readmission rates not only increase the cost burden of healthcare services but also lead to an increased risk of complications, a decline in patients' quality of life, and inefficient use of hospital resources. In the era of digital transformation, the implementation *of Electronic Health Records* (EHRs) enables hospitals to collect patient data in an integrated manner, which can then be utilized to support data-driven decision-making. In the context of Health Information Management, the use of EHR data through a machine learning approach has emerged as a growing strategy to identify patients at risk of readmission at an earlier stage, allowing clinical interventions to be implemented before the patient is discharged. One of the most widely used EHR datasets in research is *the Medical Information Mart for Intensive Care* IV (MIMIC-IV) because it provides comprehensive clinical and administrative patient data that is openly accessible for research purposes. Therefore, the development of machine learning-based

readmission prediction models represents a promising effort to support improvements in data-driven healthcare quality [1]–[4].

Research developments regarding readmission prediction over the past five years indicate that *machine learning* algorithms outperform conventional statistical methods in identifying complex patterns in electronic health record data. Panchangam [5] developed *an XGBoost-based* readmission prediction model using EMR data from approximately 310,000 patient admissions with cardiovascular and cerebrovascular conditions, demonstrating that the model outperformed the conventional LACE Plus clinical risk score, with a 30-day readmission identification accuracy of 78.39%. Dafrallah and Akhloufi [6] developed an *Explainable Artificial Intelligence* (XAI)-based readmission prediction model using SHAP on patient discharge records from the MIMIC-III dataset, demonstrating that an interpretability-based approach enhances transparency in clinical decision-making. Furthermore, Davis and Greiner [7] combined a survival analysis model with longitudinal data representation to predict time to readmission in a cohort of over 420,000 patients in Alberta, Canada, and demonstrated that this approach is effective when combined with patient representations that are not dependent on specific disease cohorts. Zhang and Kuo [8] utilized medication usage information for ICU patients from the MIMIC-IV dataset to predict prolonged hospital stays during the second admission of readmitted patients, with the best Random Forest model achieving an AUC of 0.716. In the field of respiratory diseases, Wu et al. [9] compared the performance of Lasso regression, *Random Forest, XGBoost*, and neural networks in predicting 30-day readmission and early mortality among patients with exacerbations of chronic obstructive pulmonary disease using a national database of more than 100,000 admissions, and found that XGBoost provided the best discrimination with an AUROC of 0.721 on the validation data, while also confirming the significant contribution of prior hospitalization history through SHAP analysis. These research findings indicate that the application of machine learning—particularly ensemble algorithms such as Random Forest and XGBoost—to electronic health record data holds great potential for improving the quality of readmission risk predictions, although relative performance among algorithms still varies depending on the characteristics of the population and the dataset used.

Nevertheless, there are still several research gaps that require attention. The majority of previous studies have focused on developing a single algorithm or integrating highly specific clinical data, making it difficult to directly compare results to determine the most suitable algorithm for datasets with different characteristics. Furthermore, studies specifically comparing the performance *of Random Forest* and *XGBoost* using the MIMIC-IV dataset—which contains demographic, administrative, and diagnostic variables commonly available in hospital information systems—remain relatively limited. On the other hand, the interpretability of models has not been extensively examined alongside performance evaluations, even though the transparency of prediction results is a critical factor in the implementation of clinical decision support systems. Therefore, this study offers a novel approach by conducting a comparative

evaluation of *Random Forest* and *XGBoost* for 30-day readmission prediction using the MIMIC-IV dataset, as well as integrating Feature Importance and *SHapley Additive exPlanations* (SHAP) analyses to explain the factors influencing the prediction results [6], [10], [11].

Based on this background, the objective of this study is to compare the performance of the *Random Forest* and *XGBoost* algorithms in predicting 30-day readmission among inpatients using the MIMIC-IV dataset, as well as to evaluate the performance of both models based on the metrics *Accuracy, Precision, Recall, F1-score*, and *Area Under the Curve* (AUC-ROC). Furthermore, this study aims to identify the variables that most significantly influence readmission predictions through Feature Importance analysis and *SHapley Additive exPlanations* (SHAP). The research results are expected to contribute scientifically to the development of accurate and interpretable readmission prediction models based on the approach, as well as serve as a foundation for developing *machine learning-based* decision support systems to improve the quality of healthcare services and health information management in hospitals [1], [5], [6].

RESEARCH METHOD

This study employs a quantitative approach using an experimental design to compare the performance of the *Random Forest* and *Extreme Gradient Boosting* (XGBoost) algorithms in predicting readmission of inpatients within 30 days. All stages of the analysis were conducted using the Python programming language in the Google Colaboratory environment, utilizing the *Pandas*, *NumPy*, *Scikit-learn*, XGBoost, *Matplotlib*, *Seaborn*, and SHAP libraries. The research stages included data acquisition, preprocessing, target variable selection, model building, performance evaluation, and interpretation of model results using the *Explainable Artificial Intelligence* (XAI) approach.

The research data source comes from the Medical Information Mart for Intensive Care IV (MIMIC-IV), an electronic health record (EHR) database developed by *the Massachusetts Institute of Technology* (MIT) in collaboration with *Beth Israel Deaconess Medical Center* [1]. This dataset contains administrative and clinical information on patients who received inpatient care and has undergone *de-identification* to protect patient privacy. This study utilized four main tables: *patients*, *admissions, diagnoses_icd, and d_icd_diagnosis*. These four tables were combined using the *subject_id* and *hadm_id* attributes so that each inpatient episode contains patient information, admission characteristics, and disease diagnoses within a single integrated dataset.

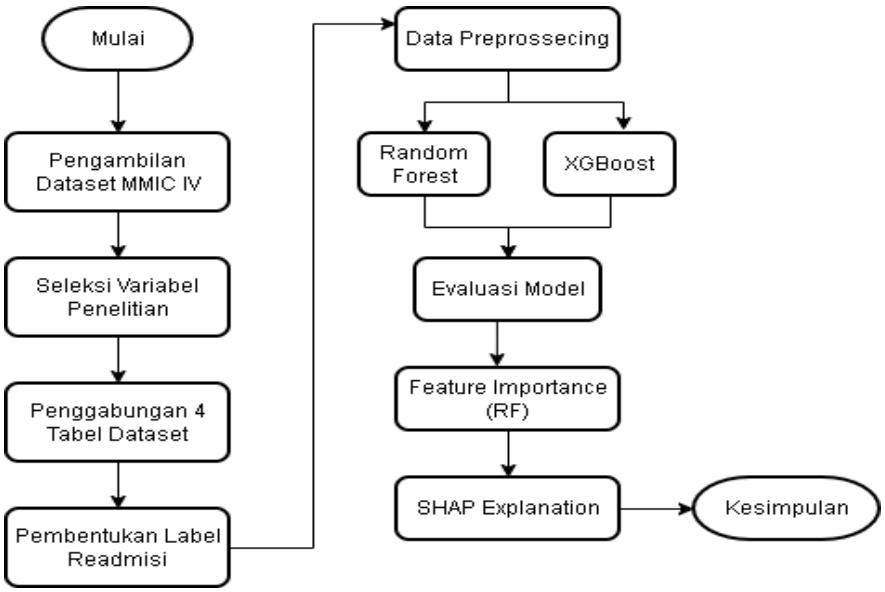

Figure 1. Flowchart

The next step is data *preprocessing* to produce a dataset ready for the modeling process. This stage begins with the selection of relevant variables, including age (*anchor_age*), gender (*gender*), admission type (*admission_type*), length *of* stay (LOS), insurance status (*insurance*), marital status (*marital_status*), race (*race*), and primary diagnosis based on ICD codes. The *Length of Stay* (LOS) value is calculated as the difference between the time the patient was admitted to the hospital (*admittime*) and the time the patient was discharged (*dischtime*), as shown in Equation (1).

$$LOS = Dischtime - Admittime \quad (1)$$

Next, a target variable (label) for 30-day readmission was created by calculating the time difference between the discharge date of the current care episode and the admission date of the subsequent hospitalization episode for the same patient. Patients were categorized as readmitted if they were readmitted within 30 days or less after discharge, while patients with an interval of more than 30 days or who did not have a subsequent hospitalization episode were categorized as non-readmitted. Patients who died during their hospitalization were excluded from the dataset because they could not be readmitted. Categorical variables were then transformed using *label encoding*, while numerical data were retained in their original form. The *preprocessed* dataset was subsequently split into training and testing data using an 80:20 ratio with the *stratified sampling* method to maintain the class proportions in both data groups, with a fixed `*random_state*` to ensure the reproducibility of the data split results.

The modeling phase was carried out using two *ensemble learning* algorithms: *Random Forest* and *XGBoost*. Random Forest builds a number of decision trees independently using the *bootstrap aggregation (*bagging*)* technique, then generates a final prediction based on a *majority voting* mechanism. In contrast, *XGBoost* builds decision trees incrementally (gradient boosting), where each new tree serves to correct the prediction errors produced by the previous tree through optimization of the objective function [12], [13]. The selection of these two algorithms was based on their ability to handle nonlinear

relationships, minimize the risk *of overfitting*, and deliver high classification performance on electronic health record data [14], [15].

The performance of both models was evaluated using five key metrics: *Accuracy, Precision, Recall, F1-score*, and *Area Under the Receiver Operating Characteristic Curve* (AUC-ROC).

Accuracy is used to measure the proportion of correct predictions relative to the entire test dataset, as shown in Equation (2).

$$Accuracy = \frac{TP + TN}{TP+TN+FP+FN} \quad (2)$$

Precision indicates the proportion of true positive predictions out of all positive predictions and is calculated using Equation (3).

$$Precicion = \frac{TP}{TP+FP} \quad (3)$$

Recall, or sensitivity, indicates the model's ability to identify all patients who were actually readmitted and is expressed in Equation (4).

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

The F1-score, as the harmonic mean of *Precision* and *Recall*, is used to evaluate model performance in situations with an imbalanced class distribution, as shown in Equation (5).

$$f1 - Score = 2\ x\ \frac{Precicion\ x\ recall}{precicion+recall} \quad (5)$$

In addition, the model's discriminative ability is evaluated using the AUC-ROC, which describes the model's ability to distinguish between readmitted and non-readmitted patients at various threshold values. An AUC value closer to 1 indicates better discriminative capability [16].

To improve the model's interpretability, this study applies two *Explainable Artificial Intelligence (XAI)* approaches: *Feature Importance* and SHAP. *Feature Importance* analysis is used to identify the variables that contribute most significantly to the classification process in the *Random Forest* model. Meanwhile, SHAP was adopted to explain the contribution of each variable to the prediction results at both the global and individual levels. This approach allows for the identification of factors that increase or decrease the probability of readmission, thereby making the model's prediction results more transparent and supporting the implementation of clinical decision support systems in hospitals [6], [10], [11].

Overall, the research process began with the acquisition and integration of MIMIC-IV data, followed by *preprocessing*, the creation of readmission labels, the development of Random Forest and XGBoost models, performance evaluation using classification metrics, and interpretation of results using *Feature Importance* and SHAP. This research workflow was designed to produce predictive models that not only perform well but are also scientifically explainable, thereby supporting the development of an accurate, transparent, and easily implementable readmission prediction system in healthcare settings.

RESULTS AND DISCUSSION

Results

Dataset Preprocessing Results

The research dataset was obtained from *the Medical Information Mart for Intensive Care IV* (MIMIC-IV), which consists of four main tables: *patients*, *admissions*, *diagnoses_icd*, and *d_icd_diagnoses*. These four tables were integrated using the *subject_id* and *hadm_id* attributes, resulting in a dataset containing patient characteristics, hospitalization information, and disease diagnoses.

The *preprocessing* stage included selecting research variables, calculating *Length of Stay* **(**LOS**)**, creating a 30-day readmission label, handling missing values, transforming categorical variables using *label encoding***,** and splitting the data into training and testing sets at an 80:20 ratio**.** The variables used in the modeling process consist of age, gender, length of stay, admission type, insurance status, marital status, race, and primary diagnosis.

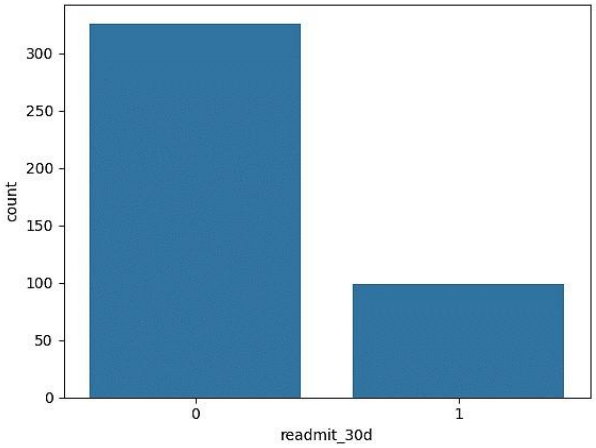

Figure 2. Readmit_30d Diagram Results

Model Performance Comparison Results

After the *preprocessing* was completed, the data distribution consisted of 99 readmitted patients (23%) and 326 non-readmitted patients (77%). This class distribution served as the basis for dividing the training and test data using the *stratified sampling* method, ensuring that the proportions of each class were maintained.

Model performance was evaluated using the metrics *Accuracy***,** *Precision***,** *Recall***,** *F1-score*, and *AUC*. The evaluation results for both algorithms are presented in Table 1.

Table 1. Model Evaluation Results

Metrics	Random Forest	XGBoost
Accuracy	78.82%	77.65%
Precision	55.00%	52.63%
Recall	55.00%	50.00%
F1-score	55.00%	51.28%
AUC	0.8354	0.8012

Based on the evaluation results in Table 1, both models were able to classify the test data. *Random Forest* achieved higher evaluation scores than XGBoost across all specified metrics.

Confusion Matrix Results

The confusion matrix is used to illustrate the distribution of classification results based on the number of *True Positives* (TP), *True Negatives* (TN), *False Positives* (FP), and *False Negatives* (FN) produced by each model.

The visualizations in Figures 3 and 4 show that both models were able to identify the readmission and non-readmission classes in the test data. The difference in prediction distribution between the two models is evident in the number of correct predictions as well as predictions that resulted in misclassification.

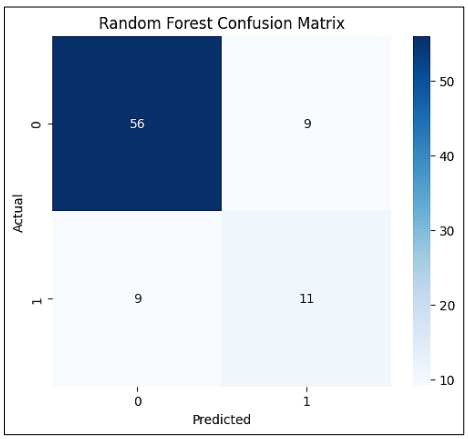

Figure 3. Visualization of the First Model's Prediction Results

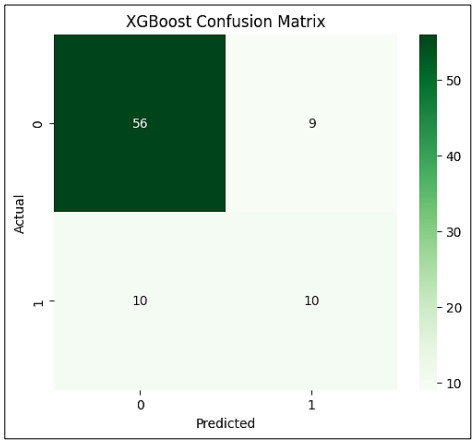

Figure 4. Visualization of the Second Model's Prediction Results

ROC Curve Evaluation Results

The ROC curve in Figure 5 is used to evaluate the model's ability to distinguish between readmission and non-readmission classes at various classification thresholds.

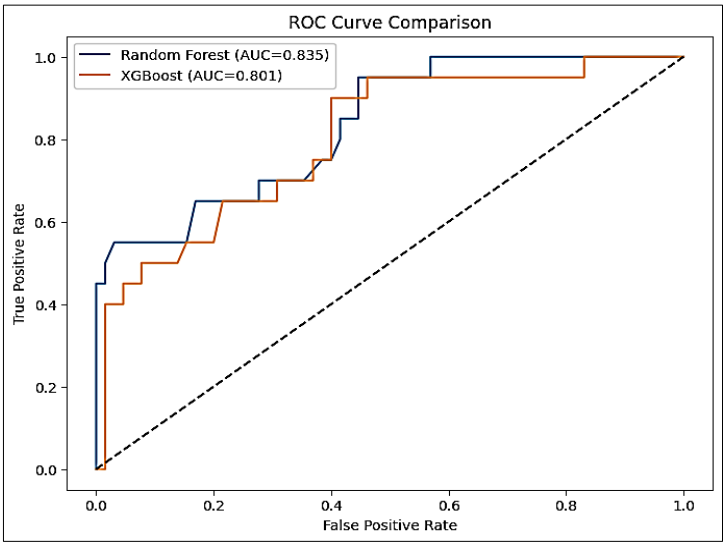

Figure 5. ROC Curve Results

The AUC values obtained for each model are summarized in Table 2.

Table 2. Model AUC Values

Model	AUC
Random Forest	0.8354
XGBoost	0.8012

The AUC values for both models are above 0.80, indicating good discriminatory power on the test data.

Feature Importance Results

Feature importance analysis was performed on the Random Forest model to determine the level of contribution of each variable to the classification process.

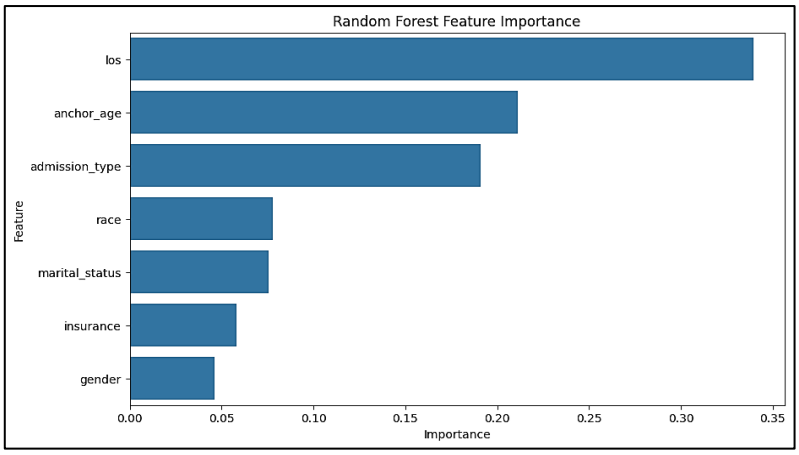

Figure 6. Random Forest Feature Importance

The visualization shows that each variable has a different level of contribution to the model. Variables with the highest *feature importance* values have a greater influence on the decision tree formation process compared to other variables.

SHAP Explanation Results

The model was interpreted using the *SHapley Additive exPlanations (*SHAP) method**.** The SHAP visualization results are presented in Figure 7 as a *Summary Plot* to show the distribution of each variable's contribution to the predicted outcome.

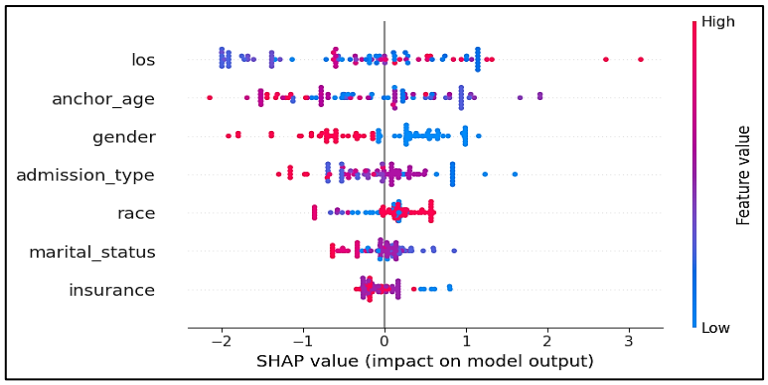

Figure 7. SHAP Summary Plot

If available, the relationship between a variable's value and its contribution to the prediction can be displayed using Figure 8, *the* SHAP *Dependence Plot*. The SHAP visualization shows the distribution of each variable's contribution to the model's prediction across the entire test dataset. The magnitude of the SHAP value illustrates the relative contribution of each variable to the prediction results generated by the model.

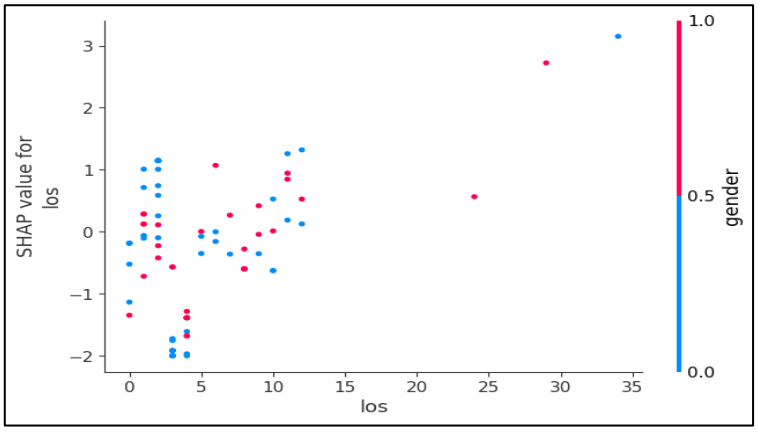

Figure 8. SHAP Dependence Plot

Discussion

Comparison of Random Forest and XGBoost Performance

The findings confirm that the *Random Forest* algorithm performs better than *Extreme Gradient Boosting* (XGBoost) in predicting 30-day readmission of inpatients using the MIMIC-IV dataset. This superiority is demonstrated by an *accuracy* of 78.82% and an AUC of 0.8354, which are higher than those of XGBoost, with *an accuracy* of 77.65% and an AUC of 0.8012. AUC values above 0.80 for both models indicate good discriminatory capability in distinguishing patients at risk of readmission from those who are not [5], [7], [12].

These results indicate that the *bagging* approach used by Random Forest is better suited to the characteristics of the study dataset than the *gradient boosting* approach used by XGBoost. Random Forest builds many decision trees independently using *bootstrap sampling*, then combines the prediction results through a *majority voting* mechanism. This approach produces a model that is more stable against data variations and more resistant to *overfitting*, especially when the data contains a combination of numerical and categorical variables, as in the MIMIC-IV dataset [14], [15]. In contrast, XGBoost is highly dependent on the *hyperparameter* optimization process, so its performance may decline if the parameter configuration is not optimal [12], [13].

The findings of this study are consistent with the results of Panchangam et al., which showed that *ensemble learning*-based algorithms are capable of achieving high readmission prediction performance on electronic health record data. However, this study provides an additional contribution through a direct evaluation of *Random Forest* and *XGBoost* using the MIMIC-IV dataset, thereby offering a clearer picture of which algorithm is more suitable for the characteristics of this data [5], [7], [8].

Comparison with Similar Studies in Different Contexts

The findings of this study can also be compared with similar studies that evaluate the performance of ensemble algorithms in different patient populations. Emi-Johnson and Nkrumah [17] evaluated the performance of *Logistic Regression, Random Forest, XGBoost,* and Deep Neural Networks in predicting 30-day readmission across 101,766 inpatient episodes for diabetic patients using *the Diabetes 130-US Hospitals dataset*, and found that *XGBoost* actually yielded the highest AUC-ROC (0.667) compared to *Random Forest* (0.630). This discrepancy with the findings of this study—in which *Random Forest* demonstrated superior performance—suggests that the relative advantage between the two algorithms is specific to the characteristics of the dataset, sample size, and the *data preprocessing* methods applied, rather than representing a universal advantage across all readmission prediction contexts. The significant difference in sample size—425 episodes in this study compared to more than 100,000 episodes in that study—further supports the hypothesis that the performance of *Random Forest* in this study needs to be retested on a larger sample before it can be broadly generalized.

This comparison also underscores the importance of caution when interpreting *the data preprocessing* in this study, particularly the use of *Label Encoding* for nominal categorical variables such as race, admission type, insurance status, and marital status. Unlike the approach in the comparative study, which used *one-hot encoding*, *Label Encoding* has the potential to imply ordinal relationships that the nominal data do not actually possess. Although decision tree-based models are relatively tolerant of this representation, future research is recommended to evaluate the impact of different *encoding* strategies on the stability and interpretability of prediction results.

Interpretation of Factors Influencing Readmission Predictions

Feature importance analysis shows that the model not only produces accurate predictions but is also capable of identifying the variables that most strongly influence the risk of readmission. The variables with the greatest contribution—such as *length of stay*, age, type of admission, and primary diagnosis—indicate that the risk of readmission is influenced by a combination of clinical and administrative factors. A longer length of stay generally reflects the severity of the disease or the complexity of the patient's condition, while advanced age is associated with an increased prevalence of chronic diseases and comorbidities [6], [11].

These results are reinforced by a *SHapley Additive exPlanations* (SHAP) analysis, which provides information on the direction and magnitude of each variable's contribution to the model's predictions. Unlike conventional *feature importance*, which only indicates the level of a variable's importance, SHAP explains how changes in a variable's value can increase or decrease the probability of readmission at both the individual and population levels [10], [11]. Thus, the developed model not only performs well but also meets the interpretability requirement, which is one of the key criteria for implementing artificial intelligence in the healthcare field [6], [18].

Implications of the Research for Health Information Management

The research results show that the use of electronic medical record data through

machine learning algorithms can support the process of identifying patients at risk of readmission at an earlier stage. From a Health Information Management perspective, this predictive model has the potential to be integrated into *a Hospital Information System* (HIS) or *Clinical Decision Support System* (CDSS) to alert healthcare professionals before a patient is discharged. Information regarding the level of readmission risk can serve as a basis for *discharge planning*, scheduling follow-up visits, and prioritizing patients who require more intensive monitoring. In the context of the electronic health record interoperability policy currently being promoted by the government through the digital health transformation program, the availability of structured administrative and clinical data is a critical prerequisite for the widespread adoption of such *machine learning*-based predictive models by hospitals in Indonesia.

In addition to improving the quality of care, the implementation of predictive models also has the potential to reduce *preventable readmissions*, increase the efficiency of hospital bed utilization, and optimize the allocation of healthcare resources. In line with the development of digital transformation in the healthcare sector, the integration of accurate and explainable predictive models has become a strategic step in supporting *data-driven healthcare* and more objective clinical decision-making [1], [16], [18], [19].

Relevance to Healthcare Quality and Financing Policies in Indonesia

Although this study uses the MIMIC-IV dataset from the U.S. healthcare system, the findings regarding factors influencing readmissions remain relevant when contextualized within Indonesia's healthcare quality policies. The Ministry of Health of the Republic of Indonesia has established Hospital Accreditation Standards through Minister of Health Decree No. HK.01.07/MENKES/1596/2024, which designates Quality Improvement and Patient Safety (PMKP) as one of the main standard categories in hospital accreditation assessments [20]. High readmission rates in this context can be viewed as an indicator reflecting the effectiveness *of discharge planning* and hospital clinical management; thus, machine learning-based prediction models, such as the one developed in this study, have the potential to support hospitals' efforts to meet these quality standards.

In addition to quality considerations, unplanned readmissions also have implications for healthcare financing under the National Health Insurance (JKN) scheme. Based on Ministry of Health Regulation No. 3 of 2023 concerning Standard Rates for Healthcare Services in the Implementation of the Health Insurance Program, financing for inpatient care at advanced-level referral healthcare facilities is carried out through the INA-CBG scheme [21]. Readmissions that are actually preventable have the potential to generate additional financing claims, which ultimately burden the sustainability of the JKN program. Thus, a readmission prediction model capable of identifying high-risk patients prior to discharge has the potential to support the efficiency of healthcare financing, in line with the spirit of quality improvement and cost control that are the focus of this policy.

Nevertheless, it should be emphasized that the characteristics of the administrative variables in this study—such as *insurance* type and *race*—are specific to the U.S.

healthcare system and are not entirely relevant for direct application to the context of medical records in Indonesia. Implementing a similar model in the Hospital Management Information System (SIMRS) in Indonesia would require adjusting variables—such as JKN enrollment status, inpatient class, and referral type—as well as revalidation using electronic medical record data from Indonesian hospitals before the model can be widely adopted in clinical practice or within *Clinical Decision Support Systems*.

Opportunities for Methodological Strengthening

Despite the superior performance demonstrated by *the Random Forest model*, several methodological aspects require critical scrutiny to prevent overinterpretation of the results. First, the evaluation of both models was conducted using only *a single train-test split* with an 80:20 ratio. Given the relatively small sample size (425 patients, of whom only 99 were readmitted), performance metric estimates from a single split may exhibit high variance and may not reliably represent the model's stable generalization ability. A *stratified k-fold cross-validation* approach—for example, with 5 or 10 folds—would provide more reliable performance estimates because it accounts for variations in results across multiple different combinations of training and test data.

Second, the performance comparison between *Random Forest* and *XGBoost* in this study was conducted without a comprehensive *hyperparameter tuning* process for either algorithm. Given that *XGBoost* is known to be highly sensitive to parameter configurations such as *learning rate*, tree depth, and the number of estimators [12], [13], the advantage of *Random Forest* found in this study may reflect suboptimal parameter configurations in *XGBoost*, rather than solely an inherent algorithmic advantage. Applying *hyperparameter* search techniques such as *grid search*, *random search*, or *Bayesian optimization* to both models would strengthen the validity of the claims regarding the performance comparison between the algorithms.

Third, the differences in *Accuracy* (78.82% versus 77.65%) and AUC-ROC (0.8354 versus 0.8012) between *Random Forest* and *XGBoost* are relatively small, while this study has not conducted statistical significance tests to confirm that these differences are not merely random variations. Tests such as *DeLong's test* for comparing two AUC curves or *McNemar's test* for comparing the classification error rates of two models could be used in future research to provide a stronger statistical basis for claims regarding the superiority of one algorithm over another.

Fourth, given that readmission labels are formed based on the history of inpatient visits for the same patient *(subject_id)*, it is important to ensure that the split between training and test data is performed at the patient level *(patient-level split)*, rather than at the inpatient episode level. If inpatient episodes from the same patient are distributed across both data groups, model performance risks being inflated due *to data leakage*, as warned in by various studies in health data science [22], [23]. Clarity regarding the unit of analysis for this data split must be explicitly stated in the research report to ensure the validity of the results.

Fifth, the class distribution in this study is imbalanced, with a readmission rate of 23% compared to a non-readmission rate of 77%. The relatively low *recall* values in both

models (55.00% for *Random Forest* and 50.00% for *XGBoost*) confirm that the models still struggle to optimally identify the minority class. The application of class imbalance handling techniques such as *the Synthetic Minority Oversampling Technique* (SMOTE) or *class weighting* from the modeling stage onward—rather than merely mentioning them as a direction for future research—would enhance the models' sensitivity in detecting patients at risk of readmission who truly require early intervention. These five methodological notes do not diminish the study's contribution in demonstrating the initial potential of both algorithms; however, they must be given primary consideration before these models are considered for further validation on a larger scale or implemented in real-world clinical practice. As an initial step that can be taken without altering the overall study design, the authors could report additional evaluation results in the form *of a range of* values from *multiple random splits* of the data with different `*random_state*` values, so that readers can assess for themselves how stable the performance differences between *Random Forest* and *XGBoost* reported in this study actually are.

Study Limitations and Directions for Future Research

This study has several limitations that should be considered when interpreting the results. First, the models were developed using the MIMIC-IV dataset, which originated from a single healthcare institution; therefore, the characteristics of the patient population may not necessarily represent conditions at other hospitals. Second, the variables considered are still limited to demographic, administrative, and primary diagnosis data, so they do not yet accommodate more detailed clinical information, such as laboratory test results, vital signs, medication prescriptions, or unstructured clinical notes. Additionally, this study has not comprehensively evaluated the impact of *hyperparameter* optimization or *class imbalance* handling techniques such as *the Synthetic Minority Oversampling Technique* (SMOTE).

Future research is recommended to integrate more comprehensive clinical data, compare other algorithms such as *LightGBM, CatBoost*, or *deep learning-based* models, and conduct external validation using data from different hospitals. Further model development utilizing multimodal data and *Explainable Artificial Intelligence* techniques is expected to improve predictive capabilities while strengthening healthcare professionals' confidence in the implementation of AI-based decision support systems in clinical practice [12], [19]. Future research is also recommended to explicitly apply data splitting at the patient level to avoid potential data leaks, conduct statistical significance tests between models, and systematically apply *hyperparameter tuning* and *cross-validation* as recommended in reviews on best practices for machine learning modeling in health data [21]–[23].

CONCLUSION

Fundamental Finding : This study successfully compared the performance of *the Random Forest* and *Extreme Gradient Boosting* (XGBoost) algorithms in predicting 30-day readmissions among hospitalized patients using the *Medical Information Mart for Intensive*

Care IV (MIMIC-IV) dataset. The evaluation results show that *Random Forest* performs better than *XGBoost*, with an *Accuracy* of 78.82%, *Precision* of 55.00%, *Recall* of 55.00%, *F1-score* of 55.00%, and AUC of 0.8354, while *XGBoost* achieved *an Accuracy* of 77.65%, *a Precision* of 52.63%, *a Recall* of 50.00%, *an F1-score* of 51.28%, and an AUC of 0.8012. **Implication:** These results indicate that *Random Forest* is more effective at classifying patients at risk of readmission based on the data characteristics used in this study, although the performance difference between the two models remains relatively small and requires further confirmation through statistical testing and validation on a larger sample. **Limitation:** In addition to producing good predictive performance, this study also shows that the application of *Feature Importance* and *SHapley Additive exPlanations* (SHAP) can improve model interpretability by identifying the variables that contribute to readmission predictions. **Future Research:** These findings contribute to the development of predictive models that are not only accurate but also transparent and easy to understand, thereby potentially supporting the implementation of *a Clinical Decision Support System* (CDSS) to help healthcare professionals identify high-risk patients and improve the quality of *discharge planning* and data-driven healthcare services.

ACKNOWLEDGMENTS

The success of this study would not have been possible without the support of various parties; therefore, the authors would like to express their appreciation and gratitude to all those who contributed. Special thanks are also extended to the developers of *the Medical Information Mart for Intensive Care* IV (MIMIC-IV) for providing the open-access dataset that made this study possible. The author would also like to thank the institution where the research took place for the facilities and support provided, as well as all those who offered criticism, suggestions, and feedback throughout the writing process until this manuscript was completed.

REFERENCES

[1] A. E. W. Johnson, L. Bulgarelli, L. Shen, A. Gayles, A. Shammout, S. Horng, et al., "MIMIC-IV, a freely accessible electronic health record dataset," *Sci. Data*, vol. 10, no. 1, Art. no. 1, Jan. 2023.

[2] CMS, "Hospital Readmissions Reduction Program," 2023. [Online]. Available: https://www.cms.gov/medicare/payment/prospective-payment-systems/acute-inpatient-pps/hospital-readmissions-reduction-program-hrrp.

[3] P. Michailidis, A. Dimitriadou, T. Papadimitriou, and P. Gogas, "Forecasting hospital readmissions with machine learning," *Healthcare*, vol. 10, no. 6, Art. no. 981, May 2022.

[4] A. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," *N. Engl. J. Med.*, vol. 380, no. 14, pp. 1347–1358, Apr. 2019.

[5] P. V. R. Panchangam, A. T., B. U. T., and M. J. Maniaci, "Machine learning-based prediction of readmission risk in cardiovascular and cerebrovascular conditions using patient EMR data," *Healthcare*, vol. 12, no. 15, Art. no. 1497, Jul. 2024.

[6] S. Dafrallah and M. A. Akhloufi, "Hospital re-admission prediction using named entity recognition and explainable machine learning," *Diagnostics*, vol. 14, no. 19, Art. no. 2151, Sep. 2024.

[7] S. Davis and R. Greiner, "Survival models and longitudinal medical events for hospital

readmission forecasting," *BMC Health Serv. Res.*, vol. 24, no. 1, Art. no. 1394, Nov. 2024.
[8] M. Zhang and T.-T. Kuo, "Early prediction of long hospital stay for intensive care units readmission patients using medication information," *Comput. Biol. Med.*, vol. 174, Art. no. 108451, May 2024.
[9] C.-Y. Wu, C.-N. Hsu, C. Wang, J.-Y. Chien, C.-C. Wang, and F.-J. Lin, "Predicting outcomes after hospitalisation for COPD exacerbation using machine learning," *ERJ Open Res.*, vol. 11, no. 3, Art. no. 00651-2024, May 2025.
[10] C. Molnar, *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. Christoph Molnar, 2022.
[11] S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," Nov. 25, 2017.
[12] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A comparative analysis of gradient boosting algorithms," *Artif. Intell. Rev.*, vol. 54, no. 3, pp. 1937–1967, Mar. 2021.
[13] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, New York, NY, USA: ACM, 2016, pp. 785–794.
[14] G. Biau and E. Scornet, "A random forest guided tour," *TEST*, vol. 25, no. 2, pp. 197–227, Jun. 2016.
[15] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.
[16] M. M. Ahsan, S. A. Luna, and Z. Siddique, "Machine-learning-based disease diagnosis: A comprehensive review," *Healthcare*, vol. 10, no. 3, Art. no. 541, Mar. 2022.
[17] O. G. Emi-Johnson and K. J. Nkrumah, "Predicting 30-day hospital readmission in patients with diabetes using machine learning on electronic health record data," *Cureus*, vol. 17, no. 4, Art. no. e82437, 2025.
[18] E. J. Topol, "High-performance medicine: The convergence of human and artificial intelligence," *Nat. Med.*, vol. 25, no. 1, pp. 44–56, Jan. 2019.
[19] M. P. Sendak, J. D'Arcy, S. Kashyap, M. Gao, M. Nichols, K. M. Corey, et al., "A path for translation of machine learning products into healthcare delivery," *EMJ Innov.*, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:211837289.
[20] Kementerian Kesehatan Republik Indonesia, *Keputusan Menteri Kesehatan Republik Indonesia*. Jakarta, Indonesia, 2024. [Online]. Available: https://arissusanto.com/wp-content/uploads/2024/10/KMK-No.-HK.01.07-MENKES-1596-2024-ttg-Standar-Akreditasi-Rumah-Sakit-signed.pdf.
[21] Kementerian Kesehatan Republik Indonesia, *Peraturan Menteri Kesehatan Nomor 3 Tahun 2023*, 2023. [Online]. Available: https://peraturan.bpk.go.id/Details/275518/permenkes-no-3-tahun-2023.
[22] A. Chiavegatto Filho, A. F. De Moraes Batista, and H. Geremias dos Santos, "Data leakage in health outcomes prediction with machine learning. Comment on 'Prediction of incident hypertension within the next year: Prospective study using statewide electronic health records and machine learning,'" *J. Med. Internet Res.*, vol. 23, no. 2, Art. no. e10969, 2021. [Online]. Available: https://www.jmir.org/2021/2/e10969.
[23] S. Kaufman, S. Rosset, and C. Perlich, "Leakage in data mining," in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, New York, NY, USA: ACM, 2011, pp. 556–563.

***Satriyo Kristanto (Corresponding Author)**
Baptis Hospital Health Sciences College of Kediri, Indonesia
Email: satriyokristanto05@gmail.com

Yustyzia Priyatmaja Putri
Baptis Hospital Health Sciences College of Kediri, Indonesia