Explainable AI Models for Predicting Postoperative Readmission: Enhancing Surgical Aftercare with Interpretable Intelligence

Tarasovskaya Nataliya[1], Siska Nia Irasanti[2], Myat Thida Win[3, *]

[1] Margulan University, Pavlodar, Kazakhstan
[2] Department of Public Health, Faculty of Medicine, Universitas Islam Bandung, Indonesia

[3] Internal Medicine Department, Faculty of Medicine, University of Cyberjaya, Malaysia

[*]Email: myat@cyberjaya.edu.my

Abstract

This study explored the application of machine learning models in predicting the readmission rate of postoperative patients. We used random forests and Gradient Boosting Machines (GBM) to evaluate their performance in predicting the risk of readmission. The results showed that although these models showed good recall and were able to effectively identify high-risk patients, there was still a trade-off between recall and precision. Key clinical characteristics such as postoperative complications, age, type of surgery, and length of stay were identified as important predictors of readmission risk. Although these models have achieved encouraging results, they still face challenges in clinical application, especially in terms of interpretability and fairness. Future research should focus on enhancing the interpretability of the models and expanding the dataset and introducing new algorithms that can enhance the prediction performance, ultimately contributing to better management of patients and optimizing the allocation of medical resources.

Keywords: Machine Learning, Post-Surgical Readmission, Random Forest, Gradient Boosting Machines

Article History:
Received January 05, 2025
Revised March 20, 2025
Accepted April 01, 2025
Available Online April 12, 2025

Explainable AI Models for Predicting Postoperative Readmission: Enhancing Surgical Aftercare with Interpretable Intelligence

1. Introduction

Postoperative readmission, defined as an unplanned return to the hospital within a specified time after discharge, has become a key measure of healthcare quality, patient safety, and system efficiency. In many healthcare systems around the world, 30-day readmission rates are increasingly used as a performance benchmark and reimbursement determinant [Kristensen, et al., 2015]. The Centers for Medicare and Medicaid Services in the United States imposes fines on hospitals with excessively high readmission rates under the Hospital Readmission Reduction Program (HRRP) [McIlvennan, et al., 2015]. In addition to the economic impact, readmissions impose significant burdens on patients.

Surgical patients are a special group for whom readmissions are particularly problematic. Despite advances in perioperative care, enhanced recovery protocols, and surgical techniques, readmission rates after surgeries such as cardiac surgery, orthopedic joint replacements, and abdominal surgery remain high, with reported readmission rates ranging from 5% to more than 20%. Accurately identifying patients at higher risk of readmission before or immediately after discharge is critical to inform post-discharge planning, early intervention, and resource allocation [Coffey, et al., 2019; Soucier, et al., 2018].

Predicted reading is complex because it is influenced by multiple factors. Risk factors cover a wide range of areas. This complexity poses a challenge to traditional statistical methods because they may not capture nonlinear interactions and higher-order relationships between variables.

Traditional methods for predicting postoperative readmissions—such as logistic regression, Cox proportional hazards models, or scoring systems such as the LACE index—have limited capabilities for modeling complex, high-dimensional clinical data [Thottakkara, et al., 2016]. These models often rely on manually selected features and assume a linear relationship between predictors and outcomes, which may underestimate the true risk in vulnerable populations.

In recent years, Machine Learning (ML) and Artificial Intelligence (AI) techniques have demonstrated superior predictive performance in many healthcare applications [Ahmed, et al., 2020; Javaid, et al., 2022]. Algorithms such as random forests, gradient boosting machines, and deep neural networks can effectively learn from large datasets containing many features and interactions. These methods have performed well in predicting clinical outcomes.

However, many AI models in healthcare face a key barrier to clinical application: lack of interpretability [Band, et al., 2023]. Complex models may have high accuracy, but little understanding of the prediction mechanism. This opacity may undermine clinical trust, hinder regulatory approval, and raise ethical concerns, especially in high-risk scenarios. In order to successfully integrate AI into the postoperative care process, it must not only be accurate, but also explainable and actionable [Gordon, et al., 2019; Guni, et al., 2024].

This study aims to develop and validate an interpretable machine learning framework for predicting postoperative readmission rates, thereby improving the safety and efficiency of postoperative care. The approach strikes a balance between predictive performance and

interpretability, which enables healthcare professionals to understand, trust, and act on model outputs.

The main contributions of this study are as follows:

(1) Comprehensive data integration: We constructed a robust patient cohort using high-quality Electronic Health Record (EHR) data, intraoperative, and postoperative variables. Data sources such as the MIMIC-III clinical database and institutional hospital records were used to ensure the generalizability and reproducibility of the model.

(2) Advanced prediction models: Multiple machine learning algorithms, such as logistic regression, random forest, XGBoost, and LightGBM, were benchmarked to determine the most effective methods. Emphasis was placed on capturing nonlinear relationships and temporal trends inherent in clinical pathways.

(3) Explainable AI Implementation: We apply post hoc interpretability techniques such as SHAP to elucidate model behavior and reveal key predictors of readmission. Visualizations and decision explanations are presented in a clinically meaningful form.

(4) Clinical Relevance and Usability: The model is evaluated not only in terms of accuracy, but also in terms of importance consistency and subgroup fairness. We specifically focus on surgical specialties with the highest risk of readmission.

(5) Foundations for Real-World Deployment: We discuss the implications of deploying the model in a Hospital Information System (HIS). The framework is designed to support future extensions to intelligent discharge planning systems and personalized follow-up strategies.

By advancing a predictive tool that is both accurate and interpretable, this study contributes to the burgeoning field of human-centered healthcare AI. It directly responds to the need for a reliable, understandable, and practical decision support system in postoperative clinics. Our findings may inform policy formulation to reduce preventable readmissions and ultimately improve outcomes for postoperative patients.

2. Related Work

2.1 Traditional Methods in Readmission Risk Prediction

Hospital readmission rates, especially those following surgery, have long been recognized as a major challenge facing healthcare systems [Merkow, et al., 2015]. Initial attempts to address this problem relied heavily on traditional statistical methods, especially logistic regression and related linear modeling techniques [Stoltzfus, 2011]. These models are built on structured datasets and typically use manually selected variables to predict the likelihood of readmission within a specific time.

The advantage of these traditional models is that they are highly interpretable and easy to incorporate into clinical settings. They can directly test the relationship between independent variables and outcomes and provide clinicians with easy-to-understand decision rules or risk scores. However, their applicability is often limited by several key flaws.

Traditional statistical models are based on strict assumptions, which are rarely met in complex clinical datasets [Wu, et al., 2021; Lee & Yoon, 2017]. Surgical outcomes are influenced by multiple factors—from physiological conditions to social determinants—that interact in nonlinear and often difficult-to-predict ways. Linear models often fail to capture the full range of interactions between clinical characteristics.

Feature selection in these models is often based on domain knowledge and univariate statistical tests, which can miss important but non-significant predictors [Staartjes, et al., 2022; Chicco, et al., 2019]. This approach can also introduce bias and reduce the adaptability of the model when applied to different patient populations or healthcare settings.

Traditional models have issues with scalability and performance when applied to large, high-dimensional datasets, such as those found in modern Electronic Health Records (EHRs) [Xie, et al., 2022; Shickel, et al., 2017]. Their performance metrics, such as sensitivity and specificity, are often lacking compared to more advanced data-driven approaches.

Generalization remains an ongoing issue. Models trained in a specific hospital setting may not transfer effectively to other hospital settings due to differences in care protocols, patient demographics, or data structure [Wiens, et al., 2014; Zhang, et al., 2022]. This limits their usefulness in creating scalable, system-level readmission risk management solutions.

2.2 AI-Based Models and Their Advantages

The rise of artificial intelligence and machine learning offers powerful alternatives to traditional statistical methods [Gupta, et al., 2021]. These methods can process large amounts of clinical data and reveal complex, nonlinear relationships between variables without pre-specified assumptions.

Machine learning algorithms have demonstrated strong predictive performance in healthcare applications [Abdollahi, et al., 2021; Nwanosike, et al., 2022]. These models can automatically learn feature interactions, handle missing data more robustly, and provide personalized risk scores based on rich and diverse clinical inputs.

One of the main advantages of AI models is their ability to incorporate a wider range of features—from structured data such as lab values and vital signs to unstructured data such as clinical notes and imaging metadata [Tayefi, et al., 2021; Mohsen, et al., 2022]. This multimodal integration enables more comprehensive risk analysis, which is particularly important in postoperative care, where recovery trajectories can be affected by multiple concurrent factors.

By integrating real-time postoperative metrics such as wound healing status, early complications, mobility progression, and pain control, AI models can dynamically update predictions of readmission risk during hospitalization and after discharge. This adaptability and sensitivity to clinical changes far exceeds that of static time-point regression models [Jiang, et al., 2019; Yu, et al., 2015].

Advanced models such as gradient boosting and deep learning have significantly improved prediction accuracy [Ebrahimi, et al., 2019]. Their ability to model higher-order interactions makes it possible to identify complex clinical patterns that may not be obvious to clinicians or undetectable by traditional methods. In benchmark tests, these models often demonstrate significantly higher area under the receiver operating characteristic curve, precision, and recall them more reliable in identifying high-risk populations [Carrington, et al., 2022; Huang, et al., 2021].

Despite these advantages, AI-based models have not yet been widely adapted into routine clinical workflows. One of the main obstacles is their "black box" nature - while they can provide accurate predictions, they often cannot explain how and why specific decisions were made, which limit clinicians' trust and accountability in decisions.

2.3 The Role of Explainable AI in Healthcare

To address concerns about transparency and trustworthiness, explainable artificial intelligence has emerged as an important area of research. XAI aims to bridge the gap between model accuracy and interpretability by providing tools to help clinicians understand the logic behind AI predictions [Rane, et al., 2023].

Techniques such as feature attribution models, alternative model approximations, and model-independent explanation frameworks are increasingly being used in the medical field [Jin, et al., 2022; Wani, et al., 2024]. Among them, the method of assigning importance scores to individual features for each prediction is popular because of its intuitiveness. These scores help clarify which clinical variables have the greatest impact on the risk of patient readmission, allow doctors to verify the model's reasoning and take appropriate actions.

In addition to feature-based explanations, visualization tools are also being explored to improve model transparency. Heat maps, partial dependency plots, and decision paths can help clinicians more intuitively understand how different combinations of clinical indicators lead to specific outcomes [Linhares, et al., 2022]. These visualization tools enhance usability and make it easier for medical professionals to incorporate AI insights into their daily practice.

Explainability is not only a matter of clinical usefulness, but also a legal and ethical requirement in many jurisdictions. Regulators are increasingly emphasizing that AI systems in healthcare must be explainable, fair, and auditable [Díaz-Rodríguez, et al., 2023]. Explainability is no longer a nice-to-have, it is a prerequisite for responsible and compliant deployment of AI technologies in clinical settings.

From a user experience perspective, explainable models also facilitate better communication with patients. When clinicians can clearly articulate why certain patients are considered at risk for readmission, it can lead to more informed shared decision making, improved patient engagement, and greater adherence to post-discharge care plans [Gledhill, et al., 2023; Greysen, et al., 2017].

However, a delicate balance must be found. Oversimplify models for the sake of explainability can result in degraded performance. Focus solely on predictive power without focusing on explainability can render models unusable in high-stakes healthcare settings. Developing hybrid approaches, adding a powerful layer of explanation to accurate AI models, represents a promising direction for future research.

2.4 Summary

Traditional readmission prediction methods provide a foundation for understanding risk, but their assumptions and ability to handle complex heterogeneous clinical data are limited. AI-based models have achieved significant improvements in predictive performance and feature integration, but their application faces obstacles due to interpretability issues. In recent years, the rise of explainable AI has provided a solution that can bridge these gaps by balancing accuracy and transparency [Albahri, et al., 2023].

This study aims to build on this evolving landscape and propose a prediction framework that combines the advantages of advanced machine learning techniques with powerful interpretability tools. It aims to provide a practical and reliable tool to support postoperative care and reduce unnecessary readmissions.

3. Data and Feature Engineering

The performance of any AI-based predictive system in healthcare is highly contingent upon the quality, structure, and relevance of the underlying data. In the context of post-surgical readmission risk prediction, effective data collection and robust feature engineering are essential to capture both clinical complexity and patient variability [Zhang, et al., 2022]. This section outlines the datasets used, the methods applied for data preprocessing, and the strategy for feature selection and construction to ensure optimal model inputs.

3.1 Data Sources

This study utilized high-quality, large-scale Electronic Health Record (EHR) datasets to develop and validate predictive models. Specifically, two data sources were considered:

(1) Public hospital databases: Resources such as the Medical Information Market for Intensive Care III (MIMIC-III) provide detailed, de-identified clinical data of more than 40,000 patients, including surgical records, vital signs, medication records, laboratory results, and clinical notes. These data provide a rich environment for model training and benchmarking.

(2) Institutional hospital data: In actual deployments, localized hospital datasets provide specific contextual data about surgical interventions, patient demographics, readmission history, and clinical follow-up records. These datasets enable models to be fine-tuned for specific populations and institutional practices.

The integration of multiple data sources enhances the generalization ability of the model and enables the model to have more robust performance in different clinical settings.

3.2 Data Preprocessing

Before feeding data into machine learning models, rigorous preprocessing is critical. Raw EHR data often contains missing values, outliers, inconsistent formats, and redundant variables that can affect model performance if not handled properly.

All continuous variables are screened for physiological plausibility. Outliers are limited using clinical thresholds or interquartile ranges. Categorical variables are standardized across sources.

Different imputation techniques are used depending on the nature of the variable. For time-invariant clinical attributes such as sex or comorbidity index, missing values are treated as new categories. For dynamic or continuous measurements, mean substitution or K-nearest neighbors are used for imputation depending on the missingness pattern.

Continuous features are standardized to a standard scale. Categorical variables are one-hot encoded to avoid ordinal assumptions. For high-cardinality features such as ICD codes, we consider using embedding layers or frequency-based aggregation methods.

Since patient conditions change over time, temporal data are aggregated into statistical summaries within relevant time windows to reduce dimensionality and preserve predictive signals.

3.3 Feature Engineering Strategy

One of the key innovations in building effective readmission prediction models is the design of relevant high-value features that reflect the patient's postoperative risk profile

[Santos, et al., 2024]. The selection and construction of these features covers multiple topic areas:

(1) Demographic and socioeconomic factors: Variables such as age, sex, insurance status, and source of admission are included as baseline predictors. These variables are often associated with access to post-discharge care and long-term rehabilitation outcomes.

(2) Clinical history: Past medical history, previous readmissions, and comorbidity indices provide a longitudinal perspective on patient vulnerability.

(3) Surgical characteristics: Type of surgery, duration of surgery, method of anesthesia, and blood loss are used to assess surgical risk. These characteristics are often closely associated with the likelihood of complications and subsequent readmission.

(4) Postoperative vital signs and laboratory tests: Extract and count early postoperative indicators such as blood pressure trends, oxygen saturation, creatinine levels, and white blood cell counts. These indicators are strong indicators of early complications.

(5) Medication and treatment regimen: Combine specific drug categories, ICU admissions, or postoperative interventions to reflect the intensity and complexity of treatment.

(6) Discharge conditions and plans: Integrate discharge arrangements and follow-up instructions to reflect the level of support after discharge, which may affect the risk of readmission.

3.4 Feature Selection and Dimensionality Reduction

Although a rich feature set can enhance predictive power, it can also lead to multicollinearity and overfitting if not managed properly [Garg & Tai, 2013]. To address this, we used a multi-step feature selection process:

(1) Univariate filtering: The initial filter removes features with small variance or weak univariate correlation with the readmission label. This eliminates uninformative or redundant variables at an early stage.

(2) Correlation analysis: A correlation matrix is used to evaluate highly correlated features. To minimize redundancy, one of each correlated feature pair is removed.

(3) Recursive Feature Elimination (RFE): For algorithms such as logistic regression and support vector machines, RFE is used to iteratively remove the least important features based on the model coefficients.

(4) Tree-based feature importance: Feature importance scores are calculated using random forest and gradient boosting models. Features that are consistently ranked low across multiple models are pruned.

(5) Domain knowledge management: Clinically relevant features are retained even if they are statistically weak in the training data to maintain interpretability and clinical consistency.

In some models, dimensionality reduction techniques are also tested to further compress high-dimensional data while preserving informative variance, especially for unstructured or high-frequency time series inputs.

3.5 Label Definition and Outcome Variable

The binary classification task defines the outcome variable as readmission within 30 days of discharge. This window is consistent with most healthcare quality benchmarks and

reimbursement frameworks. To maintain label specificity, readmissions due to elective surgery or unrelated medical issues were excluded.

When data allowed, multi-class expansion of the labels was explored in auxiliary experiments to assess the temporal sensitivity of the model.

4. Model Development

In developing a model to predict postoperative readmission, we went through several key stages, each designed to ensure that the model was both effective and practical for clinical applications. This section provides a comprehensive overview of the model selection and architecture, the training and validation strategies used to develop the model, and the methods used to ensure model interpretability and explainability.

4.1 Model Selection and Architecture

The first and most critical step in model development is to choose an algorithm that best fits the structure and complexity of the data. In the domain of healthcare data, especially in predicting postoperative readmission, it is critical to choose a model that can handle both structured and unstructured data and the potential class imbalance inherent in the problem. We explored various models.

Logistic regression was initially chosen as a baseline model because of its simplicity and ease of understanding [Shipe, et al., 2019]. Although it is generally inferior to more complex models in terms of predictive accuracy, its coefficients provide valuable insights into the relationship between various features and the probability of readmission. It is also an excellent baseline model for more complex models.

Random forest and gradient boosting machines were chosen because they can handle nonlinear relationships, interactions between features, and are robust to overfitting [Kavzoglu & Teke, 2022]. These three-based models are particularly well suited for healthcare datasets because linear models have difficulty capturing relationships between variables. Gradient boosting machines are popular in predictive modeling due to their superior performance on structured datasets, especially when dealing with missing values and complex feature interactions.

Support Vector Machines (SVMs) were also considered, especially because of their effectiveness in high-dimensional spaces [Ghaddar & Naoum-Sawaya, 2018]. We chose SVM with a Radial Basis Function (RBF) kernel to explore the model's ability to capture the nonlinear boundary between readmission and non-readmission cases.

We introduced artificial neural networks to evaluate the feasibility of deep learning in healthcare prediction. While deep learning models generally require larger datasets and more computing power, they show promise in capturing complex patterns and interactions between features, which may be critical for predicting postoperative outcomes [Xue, et al., 2021; Fritz, et al., 2019].

We tested the performance of each model and compared them to each other to determine the best approach for this specific healthcare problem. While some models showed clear advantages in terms of accuracy and AUC scores, others provided better interpretability or were more robust to overfitting.

4.2 Training and Validation Strategy

Once a model is selected, a structured training and validation strategy is critical to ensure its robustness and generalization ability. We implemented a five-fold cross-validation strategy to partition the dataset into five subsets. In each iteration, four subsets were used for training and one subset was used for validation. This process was repeated five times, ensuring that every data point had a chance to be used in both the training and validation sets. Cross-validation reduces the risk of overfitting by providing a more reliable assessment of model performance.

During the training phase, the data was divided into training, validation, and test sets. The training set accounted for 70% of the data, the validation set accounted for 15%, and the test set accounted for the remaining 15%. The training set was used to optimize the model parameters, while the validation set helped monitor the performance of the model during the training process and avoid overfitting. The test set was used to evaluate the final model performance after training.

A key challenge of medical datasets, especially in the context of postoperative readmissions, is class imbalance. Readmission cases are usually much less common than non-readmission cases, which can lead to biased predictions. We employed techniques to generate synthetic samples for the minority class. We incorporate class weights into models such as logistic regression and neural networks to penalize misclassification of the minority class and ensure that the model can adequately focus on the correct identification of readmission cases.

Hyperparameter tuning plays a vital role in model optimization. We use a combination of grid search and random search methods to find the best hyperparameters for each model. For tree-based models, we fine-tune hyperparameters to improve performance. In neural networks, we optimize parameters such as the number of layers, the number of units per layer, and the dropout rate to prevent overfitting and enhance the generalization ability of the model.

4.3 Explainability and Interpretation Methods

One of the biggest challenges in deploying machine learning models in healthcare is transparency and interpretability. Clinicians must trust and understand the model's decisions, especially when those decisions directly impact patient care. Ensuring the interpretability of the model is a critical step in the development process.

SHAP values are used to interpret the output of complex models. SHAP values provide a unified measure of feature importance, allowing for both global and local interpretation. Global SHAP values show which features are most important across the entire dataset, while local SHAP values allow for detailed interpretation of individual predictions. This approach can help clinicians understand which variables are driving predictions and whether the decisions made by the model are consistent with clinical knowledge.

We explored other interpretability techniques. PDPs are used to show the relationship between features and predicted outcomes, helping to visually demonstrate the impact of a single variable on model predictions. ICE plots provide a more nuanced view by showing how changes in features affect the prediction of a single instance, providing deeper insight into the model's behavior.

For neural networks, we applied techniques such as Layer-Wise Relevance Propagation (LRP) to determine which parts of the input data contribute most to the final prediction.

LRP works by backpropagating relevance scores through the network, clearly demonstrating the contributions of different layers and nodes to the prediction.

By integrating these interpretability methods into the model development process, we can ensure that the models are not only accurate, but also interpretable. This transparency is critical to fostering clinician trust and ensuring that predictive models can be effectively applied in real-world clinical settings.

5. Results

In this section, we present the results of the model evaluation, focus on the overall performance comparison of the models, feature importance analysis, and insights gained from subgroup and sensitivity analyses. These analyses provide a deeper understanding of the performance of the models, the factors that affect their predictions, and how the models perform under different conditions.

5.1 Performance Comparison of Models

The first step in evaluating the postoperative readmission prediction models is to compare their overall performance using multiple metrics, such as accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve. Given the class imbalance in the dataset, we specifically focus on metrics such as precision, recall, and F1 score, which provide a more balanced assessment of model performance, especially in predicting the minority class.

The logistic regression model performed well, but not as well as more complex models such as random forest and Gradient Boosting Machine (GBM). Logistic regression has an accuracy of approximately 75%, but a relatively low recall of only 60%, meaning it is less effective in identifying patients who are readmitted. The precision is relatively high at 80%, indicating that the model is generally correct in predicting readmissions, but still misses many readmission cases.

Random forest and GBM performed well, with both models having an accuracy of approximately 85%. The models achieved significantly higher recall, reaching 72% for random forest and 75% for GBM, which indicated that they performed better in identifying patients who were readmitted. The models also achieved higher precision, reaching 83% for random forest and 85% for GBM. The F1 score, which measures precision and recall, also reflected their stronger performance, reaching 77% for random forest and 79% for GBM.

Support Vector Machines (SVMs) with RBF kernels also performed well, achieving 83% accuracy and 76% F1 score. However, the interpretability of this model was slightly inferior to that of the tree-based model, which makes it less suitable for deployment in clinical settings where interpretability is critical.

Artificial Neural Networks (ANNs) performed well in terms of accuracy, achieving 87%, but their performance depends heavily on the proper tuning of the network architecture and training parameters. Although ANNs performed well in detecting readmissions, their interpretability was low compared to other models, which could be a major barrier to their clinical application.

To further evaluate the model performance, other evaluation metrics such as AUROC and AUPRC were calculated. Both random forests and GBMs performed well on these metrics, with AUROC values exceeding 0.85, which indicates that they have high

discriminative power for readmitted and non-readmitted cases. AUPRC is particularly useful in imbalanced datasets, further demonstrating the superior performance of these models, especially GBM, which achieved an AUPRC score of 0.80.

5.2 Feature Importance Analysis

An important aspect of understanding predictive models, especially in healthcare, is to determine which features drive the predictions. This allows healthcare professionals to interpret the behavior of the model and gain insight into the factors that influence the risk of readmission. We analyzed feature importance using SHapley Additive Explanations values for random forests and gradient boosting machines, as well as Partial Dependence Plots (PDPs) for other models.

SHAP values revealed several key features that significantly impacted the prediction of postoperative readmission. The most important features across all models included postoperative complications, patient age, surgery type, and length of stay. Postoperative complications were the strongest predictors of readmission, which is consistent with clinical intuition. These complications significantly increased the risk of patients needing to be hospitalized again.

Patient age also plays a crucial role, with older patients being at a higher risk of readmission. This could be due to a variety of factors, such as slower recovery, the presence of comorbidities, and general frailty in older patients after surgery. Surgery type is another important feature, with high-risk surgeries, increasing the likelihood of readmission. The length of initial hospital stay is also a key factor, with patients who have a longer hospital stay due to complications having a higher risk of remission.

Other notable characteristics included pre-existing conditions, time since last surgery, and discharge plans. These factors suggest that not only the surgery itself, but also the patient's broader medical history and discharge process are critical in determining readmission risk.

Partial dependence plots further reveal how individual characteristics affect model predictions. The PDP for age shows that readmission risk clearly increases with age, which highlights that the model correctly identifies that older patients are at higher risk. The PDP for length of stay shows that readmission risk rises sharply with length of stay, especially after a stay of more than five days.

5.3 Subgroup and Sensitivity Analysis

Subgroup analysis aims to determine whether the models perform equally well in different subgroups of patients. This analysis aims to identify potential biases in model predictions and ensure that the models are applicable to all types of patients.

The results of the subgroup analysis show that the models performed similarly across age groups, but there were some differences in recall for older patients. Models performed slightly better in younger patients, but the differences were not significant. The models performed particularly well in patients undergoing high-risk surgeries, with higher recall and F1 scores compared to low-risk surgeries. This suggests that these models are better able to identify patients with a higher likelihood of complications and readmission.

We also performed sensitivity analysis to assess the robustness of the model to changes in input features. The results showed that the model was most sensitive to changes in postoperative complications and length of stay, which are known to be important predictors

of readmission. The model was less sensitive to changes in less influential features such as socioeconomic factors and outpatient follow-up, which suggests that the model focuses primarily on clinical variables.

Sensitivity analysis of model hyperparameters showed that slight changes in parameters such as learning rate, number of trees, and tree depth did not significantly affect model performance, which confirmed that the model is stable and not prone to overfitting.

6. Discussion and Conclusion

The results of this study highlight the great potential of machine learning models, especially random forests and gradient boosting machines, in predicting postoperative readmission. The models demonstrated high recall, a key metric in healthcare because they are effective in identifying patients at high risk for readmission. There is still a trade-off between recall and precision, which is a common challenge in healthcare applications. Although these models flagged many high-risk patients, they also generated false positives, which means that some patients identified as likely to be readmitted did not actually require follow-up. If this issue is not addressed, it may lead to unnecessary medical interventions and increased healthcare costs.

The clinical relevance of these models is clear because they align with known risk factors. These findings suggest that these models capture patterns that align with clinical practice, make them a valuable tool for healthcare professionals. However, there is still room for improvement. Incorporating additional characteristics such as socioeconomic status, outpatient care, and patient-reported outcomes could provide a more comprehensive view of patient recovery and potentially improve the accuracy of the models.

Despite the promising results of machine learning models, implementation in clinical settings is not without challenges. One major issue is model interpretability. While tools such as SHAP values and partial dependence plots help explain the model's predictions, the complexity of models such as random forests and GBMs still poses a challenge to clinicians who need to understand and trust these predictions. Clear and understandable explanations are essential to gain clinician buy-in and ensure models are effectively applied in practice. Efforts should be made to increase the transparency and usability of these interpretability techniques, especially in busy clinical settings where time and resources are limited.

Data privacy and ethical issues must be addressed when deploying machine learning models in healthcare. The use of sensitive patient data requires strict compliance with privacy regulations. Models must be tested for fairness to avoid biases that may disproportionately affect certain demographic groups. Diversity and representativeness of training data is critical to minimize the risk of biased predictions that could lead to discriminatory healthcare practices.

Research directions are diverse. Expanding datasets to include a wider range of characteristics could improve model performance and provide a more complete understanding of factors that influence readmission rates. More advanced machine learning techniques could further improve predictive accuracy, especially for large, unstructured datasets. Real-time data and adaptive learning systems can help models stay up to date with new patient conditions and evolving healthcare practices.

After these models are deployed validation is critical to ensure their continued accuracy and effectiveness in a variety of clinical settings. Only through rigorous testing and

continuous improvement can we ensure that these models provide reliable support to healthcare providers and help improve patient outcomes.

AKNOWLEDGEMENT

Reference

Abdollahi, J., Nouri-Moghaddam, B., & Ghazanfari, M. (2021). Deep Neural Network Based Ensemble learning Algorithms for the healthcare system (diagnosis of chronic diseases). arXiv preprint arXiv:2103.08182. DOI: 10.48550/arXiv.2103.08182.

Ahmed, Z., Mohamed, K., Zeeshan, S., & Dong, X. (2020). Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. Database, 2020, baaa010. DOI: 10.1093/database/baaa010.

Albahri, A. S., Duhaim, A. M., Fadhel, M. A., Alnoor, A., Baqer, N. S., Alzubaidi, L., ... & Deveci, M. (2023). A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion. Information Fusion, 96, 156-191. DOI: 10.1016/j.inffus.2023.03.008.

Band, S. S., Yarahmadi, A., Hsu, C. C., Biyari, M., Sookhak, M., Ameri, R., ... & Liang, H. W. (2023). Application of explainable artificial intelligence in medical health: A systematic review of interpretability methods. Informatics in Medicine Unlocked, 40, 101286. DOI: 10.1016/j.imu.2023.101286.

Carrington, A. M., Manuel, D. G., Fieguth, P. W., Ramsay, T., Osmani, V., Wernly, B., ... & Holzinger, A. (2022). Deep ROC analysis and AUC as balanced average accuracy, for improved classifier selection, audit and explanation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(1), 329-341. DOI: 10.1109/TPAMI.2022.3145392.

Chicco, D., & Rovelli, C. (2019). Computational prediction of diagnosis and feature selection on mesothelioma patient health records. PloS one, 14(1), e0208737. DOI: 10.1371/journal.pone.0208737.

Coffey, A., Leahy-Warren, P., Savage, E., Hegarty, J., Cornally, N., Day, M. R., ... & O'Caoimh, R. (2019). Interventions to promote early discharge and avoid inappropriate hospital (re) admission: a systematic review. International journal of environmental research and public health, 16(14), 2457. DOI: 10.3390/ijerph16142457.

Díaz-Rodríguez, N., Del Ser, J., Coeckelbergh, M., de Prado, M. L., Herrera-Viedma, E., & Herrera, F. (2023). Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. Information Fusion, 99, 101896. DOI: 10.1016/j.inffus.2023.101896

Ebrahimi, M., Mohammadi-Dehcheshmeh, M., Ebrahimie, E., & Petrovski, K. R. (2019). Comprehensive analysis of machine learning models for prediction of sub-clinical mastitis: Deep Learning and Gradient-Boosted Trees outperform other models. Computers in biology and medicine, 114, 103456. DOI: 10.1016/j.compbiomed.2019.103456.

Fritz, B. A., Cui, Z., Zhang, M., He, Y., Chen, Y., Kronzer, A., ... & Avidan, M. S. (2019). Deep-learning model for predicting 30-day postoperative mortality. British journal of anaesthesia, 123(5), 688-695. DOI: 10.1016/j.bja.2019.07.025.

Garg, A., & Tai, K. (2013). Comparison of statistical and machine learning methods in modelling of data with multicollinearity. International Journal of Modelling, Identification and Control, 18(4), 295-312. DOI: 10.1504/IJMIC.2013.053535.
Ghaddar, B., & Naoum-Sawaya, J. (2018). High dimensional data classification and feature selection using support vector machines. European Journal of Operational Research, 265(3), 993-1004. DOI: 10.1016/j.ejor.2017.08.040.
Gledhill, K., Bucknall, T. K., Lannin, N. A., & Hanna, L. (2023). The role of collaborative decision-making in discharge planning: Perspectives from patients, family members and health professionals. Journal of clinical nursing, 32(19-20), 7519-7529. DOI: 10.1111/jocn.16820.
Gordon, L., Grantcharov, T., & Rudzicz, F. (2019). Explainable artificial intelligence for safe intraoperative decision support. JAMA surgery, 154(11), 1064-1065. DOI: 10.1001/jamasurg.2019.2821.
Greysen, S. R., Harrison, J. D., Kripalani, S., Vasilevskis, E., Robinson, E., Metlay, J., ... & Auerbach, A. D. (2017). Understanding patient-centred readmission factors: a multi-site, mixed-methods study. BMJ quality & safety, 26(1), 33-41. DOI: 10.1136/bmjqs-2015-004570.
Guni, A., Varma, P., Zhang, J., Fehervari, M., & Ashrafian, H. (2024). Artificial intelligence in surgery: the future is now. European Surgical Research, 65(1), 22-39. DOI: 10.1159/000536393.
Gupta, R., Srivastava, D., Sahu, M., Tiwari, S., Ambasta, R. K., & Kumar, P. (2021). Artificial intelligence to deep learning: machine intelligence approach for drug discovery. Molecular diversity, 25, 1315-1360. DOI: 10.1007/s11030-021-10217-3.
Huang, C., Li, S. X., Caraballo, C., Masoudi, F. A., Rumsfeld, J. S., Spertus, J. A., ... & Krumholz, H. M. (2021). Performance metrics for the comparative analysis of clinical risk prediction models employing machine learning. Circulation: Cardiovascular Quality and Outcomes, 14(10), e007526. DOI: 10.1161/CIRCOUTCOMES.120.007526.
Javaid, M., Haleem, A., Singh, R. P., Suman, R., & Rab, S. (2022). Significance of machine learning in healthcare: Features, pillars and applications. International Journal of Intelligent Networks, 3, 58-73. DOI: 10.1016/j.ijin.2022.05.002.
Jiang, W., Siddiqui, S., Barnes, S., Barouch, L. A., Korley, F., Martinez, D. A., ... & Levin, S. (2019). Readmission risk trajectories for patients with heart failure using a dynamic prediction approach: retrospective study. JMIR medical informatics, 7(4), e14756. DOI: 10.2196/14756.
Jin, D., Sergeeva, E., Weng, W. H., Chauhan, G., & Szolovits, P. (2022). Explainable deep learning in healthcare: A methodological survey from an attribution view. WIREs Mechanisms of Disease, 14(3), e1548. DOI: 10.1002/wsbm.1548.
Kavzoglu, T., & Teke, A. (2022). Predictive performances of ensemble machine learning algorithms in landslide susceptibility mapping using random forest, extreme gradient boosting (XGBoost) and natural gradient boosting (NGBoost). Arabian Journal for Science and Engineering, 47(6), 7367-7385. DOI: 10.1007/s13369-022-06560-8.

Kristensen, S. R., Bech, M., & Quentin, W. (2015). A roadmap for comparing readmission policies with application to Denmark, England, Germany and the United States. Health policy, 119(3), 264-273. DOI: 10.1016/j.healthpol.2014.12.009.

Lee, C. H., & Yoon, H. J. (2017). Medical big data: promise and challenges. Kidney research and clinical practice, 36(1), 3. DOI: 10.23876/j.krcp.2017.36.1.3.

Linhares, C. D., Lima, D. M., Ponciano, J. R., Olivatto, M. M., Gutierrez, M. A., Poco, J., ... & Traina, A. J. (2022). Clinicalpath: a visualization tool to improve the evaluation of electronic health records in clinical decision-making. IEEE transactions on visualization and computer graphics, 29(10), 4031-4046. DOI: 10.1109/TVCG.2022.3175626.

McIlvennan, C. K., Eapen, Z. J., & Allen, L. A. (2015). Hospital readmissions reduction program. Circulation, 131(20), 1796-1803. DOI: 10.1161/CIRCULATIONAHA.114.010270.

Merkow, R. P., Ju, M. H., Chung, J. W., Hall, B. L., Cohen, M. E., Williams, M. V., ... & Bilimoria, K. Y. (2015). Underlying reasons associated with hospital readmission following surgery in the United States. Jama, 313(5), 483-495. DOI: 10.1001/jama.2014.18614.

Mohsen, F., Ali, H., El Hajj, N., & Shah, Z. (2022). Artificial intelligence-based methods for fusion of electronic health records and imaging data. Scientific Reports, 12(1), 17981. DOI: 10.1038/s41598-022-22514-4.

Nwanosike, E. M., Conway, B. R., Merchant, H. A., & Hasan, S. S. (2022). Potential applications and performance of machine learning techniques and algorithms in clinical practice: a systematic review. International journal of medical informatics, 159, 104679. DOI: 10.1016/j.ijmedinf.2021.104679.

Rane, N., Choudhary, S., & Rane, J. (2023). Explainable artificial intelligence (XAI) in healthcare: Interpretable models for clinical decision support. Available at SSRN 4637897. DOI: 10.2139/ssrn.4637897.

Santos, R., Ribeiro, B., Sousa, I., Santos, J., Guede-Fernández, F., Dias, P., ... & Londral, A. (2024). Predicting post-discharge complications in cardiothoracic surgery: A clinical decision support system to optimize remote patient monitoring resources. International Journal of Medical Informatics, 182, 105307. DOI: 10.1016/j.ijmedinf.2023.105307.

Shickel, B., Tighe, P. J., Bihorac, A., & Rashidi, P. (2017). Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. IEEE journal of biomedical and health informatics, 22(5), 1589-1604. DOI: 10.1109/JBHI.2017.2767063.

Shipe, M. E., Deppen, S. A., Farjah, F., & Grogan, E. L. (2019). Developing prediction models for clinical use using logistic regression: an overview. Journal of thoracic disease, 11(Suppl 4), S574. DOI: 10.21037/jtd.2019.01.25.

Soucier, R. J., Miller, P. E., Ingrassia, J. J., Riello, R., Desai, N. R., & Ahmad, T. (2018). Essential elements of early post discharge care of patients with heart failure. Current heart failure reports, 15, 181-190. DOI: 10.1007/s11897-018-0393-9.

Staartjes, V. E., Kernbach, J. M., Stumpo, V., van Niftrik, C. H., Serra, C., & Regli, L. (2022). Foundations of feature selection in clinical prediction modeling. In Machine

Learning in Clinical Neuroscience: Foundations and Applications (pp. 51-57). Springer International Publishing. DOI: 10.1007/978-3-030-85292-4_7.
Stoltzfus, J. C. (2011). Logistic regression: a brief primer. Academic emergency medicine, 18(10), 1099-1104. DOI: 10.1111/j.1553-2712.2011.01185.x.
Tayefi, M., Ngo, P., Chomutare, T., Dalianis, H., Salvi, E., Budrionis, A., & Godtliebsen, F. (2021). Challenges and opportunities beyond structured data in analysis of electronic health records. Wiley Interdisciplinary Reviews: Computational Statistics, 13(6), e1549. DOI: 10.1002/wics.1549.
Thottakkara, P., Ozrazgat-Baslanti, T., Hupf, B. B., Rashidi, P., Pardalos, P., Momcilovic, P., & Bihorac, A. (2016). Application of machine learning techniques to high-dimensional clinical data to forecast postoperative complications. PloS one, 11(5), e0155705. DOI: 10.1371/journal.pone.0155705.
Wani, N. A., Kumar, R., Bedi, J., & Rida, I. (2024). Explainable AI-driven IoMT fusion: Unravelling techniques, opportunities, and challenges with Explainable AI in healthcare. Information Fusion, 102472. DOI: 10.1016/j.inffus.2024.102472.
Wiens, J., Guttag, J., & Horvitz, E. (2014). A study in transfer learning: leveraging data from multiple hospitals to enhance hospital-specific predictions. Journal of the American Medical Informatics Association, 21(4), 699-706. DOI: 10.1136/amiajnl-2013-002162.
Wu, W. T., Li, Y. J., Feng, A. Z., Li, L., Huang, T., Xu, A. D., & Lyu, J. (2021). Data mining in clinical big data: the frequently used databases, steps, and methodological models. Military Medical Research, 8, 1-12. DOI: 10.1186/s40779-021-00338-z.
Xie, F., Yuan, H., Ning, Y., Ong, M. E. H., Feng, M., Hsu, W., ... & Liu, N. (2022). Deep learning for temporal data representation in electronic health records: A systematic review of challenges and methodologies. Journal of biomedical informatics, 126, 103980. DOI: 10.1016/j.jbi.2021.103980.
Xue, B., Li, D., Lu, C., King, C. R., Wildes, T., Avidan, M. S., ... & Abraham, J. (2021). Use of machine learning to develop and evaluate models using preoperative and intraoperative data to identify risks of postoperative complications. JAMA network open, 4(3), e212240-e212240. DOI: 10.1001/jamanetworkopen.2021.2240.
Yu, S., Farooq, F., Van Esbroeck, A., Fung, G., Anand, V., & Krishnapuram, B. (2015). Predicting readmission risk with institution-specific prediction models. Artificial intelligence in medicine, 65(2), 89-96. DOI: 10.1016/j.artmed.2015.08.005.
Zhang, A., Xing, L., Zou, J., & Wu, J. C. (2022). Shifting machine learning for healthcare from development to deployment and from models to data. Nature biomedical engineering, 6(12), 1330-1345. DOI: 10.1038/s41551-022-00898-y.
Zhang, J., Li, D., Dai, R., Cos, H., Williams, G. A., Raper, L., ... & Lu, C. (2022). Predicting post-operative complications with wearables: a case study with patients undergoing pancreatic surgery. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 6(2), 1-27. DOI: 10.1145/3534578.