Mesopotamian Journal of Artificial Intelligence in Healthcare
Vol.2025, **pp**. 116–123
DOI: https://doi.org/10.58496/MJAIH/2025/012; ISSN: 3005-365X
https://mesopotamian.press/journals/index.php/MJAIH

Research Article

An Artificial Intelligence Model for Predicting Hospital Readmission Using Electronic Health Records Data

Rezarta Cara[1], Klodian Dhoska[2,*], Fjona Cara[3], Fredrick Kayusi[4,5], Linety Juma[6]

[1] *Department of Engineering and Marine Sciences, University Aleksander Moisiu Durres, Albania*

[2] *Department of Mechanics, Polytechnic University of Tirana, Albania*

[3] *Department of Engineering and Marine Sciences, University Aleksander Moisiu, Durres, Albania*

[4] *Department of Environmental Studies, Geography and Planning, Maasai Mara University, P.O Box 861, Narok-Kenya*

[5] *Department of Environmental Sciences, School of Environmental and Earth Sciences, Pwani University, Kilifi, Kenya*

[6] *Department of Curriculum, Instruction and Educational Technology, Pwani University- 80108, Kilifi, Kenya*

ARTICLE INFO

Article History

Received 17 Feb 2025
Revised 15 Apr 2025
Accepted 10 May 2025
Published 08 Jun 2025

Keywords

Hospital Readmission

Machine Learning

LightGBM

Health Records

Predictive Modeling

ABSTRACT

This study investigates the application of a machine learning model—specifically the Light Gradient Boosting Machine (LightGBM)—to predict 30-day hospital readmissions using structured electronic health record (EHR) data. Hospital readmissions remain a critical challenge in healthcare systems, often indicating gaps in continuity of care and contributing to higher costs. By leveraging demographic, clinical, and diagnostic variables from 350 anonymized patient records, the model aimed to accurately identify individuals at high risk of readmission. Key features included age, number of previous admissions, length of stay, number of medications, chronic disease status, and gender. Data preprocessing, model training, and evaluation were conducted using Python-based libraries, ensuring both reproducibility and scalability. The model achieved a ROC AUC of 0.89, precision of 0.78, recall of 0.65, and F1 score of 0.71, indicating strong performance and balance between sensitivity and specificity. A confusion matrix analysis confirmed high accuracy in both positive and negative predictions. SHapley Additive exPlanations (SHAP) values were used to enhance interpretability by quantifying the contribution of each variable to the model's output. Feature importance analysis revealed that age and previous hospitalizations had the greatest impact on prediction, which aligns with prior clinical evidence. The study's results are consistent with existing research and highlight the potential of explainable AI models in healthcare risk prediction. These findings support the integration of machine learning into hospital decision-support systems to enable early intervention strategies, reduce preventable readmissions, and improve overall patient outcomes. Recommendations include adopting predictive models for discharge planning, prioritizing high-risk patient groups, and further validation across broader datasets. The study emphasizes transparency, clinical relevance, and operational feasibility, demonstrating how data-driven tools can be effectively deployed in real-world clinical environments.

1. INTRODUCTION

Hospital readmissions are a big issue for healthcare systems everywhere, leading to greater expenses in healthcare and showing that some patients do not get top-quality treatment. Avoiding unnecessary hospital readmissions is especially important, especially since hospitals in value-based care programs are rewarded for better performance and fewer unwanted returns. ML models are helping to discover patients at risk of returning to the hospital, so that efforts can be tailored to avoid readmission and make better use of available resources. Because healthcare is moving toward EHRs, research teams now have access to extensive, organized medical information for use in predictive analytics. A study by Huang and his

*Corresponding author. Email: kdhoska@fim.edu.al

colleagues, together with a study by Liu and colleagues, have shown that AI and ML can help predict 30-day hospital readmission risks in different clinical areas [1][2].
By using LightGBM, this study advances prior work by successfully predicting which patients will be readmitted to a hospital within 30 days using EHR data. It was decided to use LightGBM because of its fast calculations and ability to perform well with structured data [3][4]. To build the model, we used carefully chosen information about demographics, patient history and diagnostic test results. Because explainability is important in this approach, it helps create clinical support that people can clearly use and trust [5][6]. Furthermore, the research reinforces previous studies by using common metrics to confirm the model's usability in hospitals [13-15].
By using AI in hospitals, healthcare providers are able to quickly detect patients who need extra care upon discharge and can help them accordingly. By using this system, both patient results and the hospital's own objectives surrounding readmission are improved [7][8].

2. METHODOLOGY

In this research, a supervised machine learning model is designed and evaluated to guess future hospital readmission within 30 days from EHR data that is structured. There are 350 anonymized patient records included, drawn from hospital logs, files related to demographics, labs and coding registries. Data from these sources were joined to create a complete set of features including demographics, medical conditions, admission history, medication counts and abnormal laboratory findings. Python, which is commonly used in healthcare data science, was employed for data preprocessing and modeling. With pandas and numpy, I manipulated my data, used scikit-learn for preprocessing and to calculate evaluation metrics and trained and predicted using lightgbm. Feature distributions were examined, and exploratory analysis was supported using matplotlib and seaborn. Thanks to this approach, experiments can be reproduced and scaled up using free, open toolkits that ensure the process remains clear and adaptable in various clinical contexts.

- Data:

In this research model aimed at predicting hospital readmission within 30 days using electronic health record (EHR) data, a set of clinical and demographic variables was collected to reflect each patient's health and behavioral characteristics. These variables were carefully selected based on prior literature and the predictive model's needs. The first variable is Patient ID, a unique identifier that is non-repeating for each case, used solely for tracking purposes without influencing statistical analysis.
Second, Age is measured in full years and collected from hospital admission records. It is treated as a continuous variable and analyzed numerically, as age may directly affect the likelihood of readmission due to physiological changes. Gender is a nominal variable with two values: Male or Female, retrieved from the patient's demographic file. Chronic Diseases indicate whether the patient has chronic conditions such as diabetes, hypertension, or heart disease. It is encoded as a binary variable (1 for presence, 0 for absence), based on diagnosis codes recorded in the medical system. Number of Previous Admissions is a numerical variable representing the number of times a patient was admitted in the past year, calculated from the hospital's admission and discharge logs. Length of Stay is measured by the number of days the patient stayed during their most recent visit and serves as an indicator of illness severity and recovery ability. Primary Diagnosis is a nominal variable showing the main reason for admission, determined using the main medical diagnosis code and reclassified into five primary categories: Diabetes, Heart Failure, COPD, Hypertension, and Cancer. Number of Medications is a numerical variable representing the count of prescribed medications during the hospital stay, used as an indicator of case complexity. Lab Test Results Abnormal is a binary variable indicating whether at least one lab test result was outside the normal range, extracted from the electronic lab records. Readmitted Within 30 Days is the dependent (target) variable in this model, measured as a binary value (1 = Yes, 0 = No), determined by comparing the discharge date with the next admission date. This is the variable the model aims to predict based on the independent variables. All variables were preprocessed to ensure they are free of missing or duplicate values. Nominal variables were encoded into numerical formats where necessary, while preserving the clinical meaning of each variable in the medical context.

- AI Framework:

GBDT along with LightGBM is the main theoretical framework used in this work, due to its high speed and accuracy when managing structured data. LightGBM sequentially creates a set of decision trees and each new tree tries to improve upon the performance of the existing predicted result from the previous trees.

1- Boosting Framework:

The general boosting model aims to minimize a loss function $\mathcal{L}$ by iteratively adding weak learners $f_t(x)$ to the model. Formally, the ensemble model after T iterations is [10][11]:

$$\hat{y}(x) = \sum_{t=1}^{T} f_t(x) \tag{1}$$

At each iteration, LightGBM solves the following optimization problem:

$$f_t = \arg\min_{f \in \mathcal{F}} \sum_{i=1}^{n} \ell\left(y_i, \hat{y}_i^{(t-1)} + f(x_i)\right) + \Omega(f) \tag{2}$$

where: ℓ is the differentiable loss function (e.g., binary log loss for classification), $\Omega(f)$ is a regularization term to control model complexity, $\mathcal{F}$ is the space of regression trees. $\hat{y}_i^{(t-1)}$ is the prediction at iteration t−1.

2- Second-Order Taylor Approximation:

To efficiently solve the optimization, LightGBM uses the second-order Taylor expansion of the loss function:

$$\mathcal{L}^{(t)} \approx \sum_{i=1}^{n} \left[g_i f(x_i) + \frac{1}{2} h_i f^2(x_i)\right] + \Omega(f) \tag{3}$$

where: $g_i = \frac{\partial \ell(y_i, \hat{y}_i)}{\partial \hat{y}_i}$ is the first-order gradient, $h_i = \frac{\partial^2 \ell(y_i, \hat{y}_i)}{\partial \hat{y}_i^2}$ is the second-order gradient (Hessian). This formulation allows LightGBM to build trees by minimizing this approximation at each node split.

3- Leaf-Wise Tree Growth Strategy:

Unlike traditional level-wise growth, LightGBM uses a **leaf-wise** strategy. At each step, it expands the leaf that results in the largest decrease in loss:

$$\text{Gain} = \frac{1}{2}\left[\frac{G_L^2}{H_L+\lambda} + \frac{G_R^2}{H_R+\lambda} - \frac{(G_L+G_R)^2}{H_L+H_R+\lambda}\right] - \gamma \tag{4}$$

where: G_L^2, G_R^2 are the sums of gradients for the left and right child, H_L, H_R are the sums of Hessians for the left and right child, λ is the L2 regularization parameter, γ is the pruning penalty.

4- Binary Classification Objective:

Since the study uses LightGBM for binary classification (readmitted: yes or no), the loss function is:

$$\ell(y, \hat{y}) = -[y\log(\hat{p}) + (1-y)\log(1-\hat{p})] \tag{5}$$

$$\hat{p} = \frac{1}{1+e^{-\hat{y}}} \tag{6}$$

Where $\hat{y}$ is the raw score output from the model, and $\hat{p}$ is the probability via the sigmoid function.

To fully assess both the prediction accuracy and understandability of the model, the study uses the Receiver Operating Characteristic (ROC) curve, a confusion matrix and SHapley Additive exPlanations (SHAP). A ROC curve portrays the diagnostic performance of a model as thresholds in the model vary. It shows the true positive rate and the false positive rate on an axis which illustrates how increasing one can affect the other. A model with a large AUC value is said to perform better, since any value over 0.5 means its ROC curve is above the line at every point. Along with ROC, using the confusion matrix allows for a clear verdict on how well the model performs as compared to the actual results. With this matrix, different labeled results are grouped as true positives, true negatives, false positives and false negatives which permit computing accuracy, precision, recall and F1-score. They allow us to judge how often the model performs correctly and if it guides us to avoid alarming cases of patients who will not require readmittance. To make data interpretable, the study adopts SHAP values which use cooperative game theory to determine the role of every feature in each prediction. Since SHAP values are consistent and based on theory, they allow people to understand why the model reaches its decisions both globally and at the feature level. Њehen using SHAP, healthcare providers gain clarity on the most important variables connected to patients' hospital readmission.

3. RESULTS AND DECISION

TABLE I. DESCRIPTIVE STATISTICS OF INPUT VARIABLES

Variable	Type	Summary Statistics
Age	Continuous	Mean = 53.44, SD = 21.11, Min = 18.0, Max = 89.0
NumPreviousAdmissions	Continuous	Mean = 1.47, SD = 1.22, Min = 0.0, Max = 6.0
LengthOfStay	Continuous	Mean = 7.52, SD = 4.03, Min = 1.0, Max = 14.0
NumMedications	Continuous	Mean = 4.93, SD = 2.59, Min = 1.0, Max = 9.0
Gender	Categorical	Male = 187 (53.4%); Female = 163 (46.6%)
ChronicDiseases	Categorical	0 = 210 (60.0%); 1 = 140 (40.0%)

PrimaryDiagnosis	Categorical	Diabetes = 77 (22.0%); Heart Failure = 73 (20.9%); COPD = 69 (19.7%); Hypertension = 69 (19.7%); Cancer = 62 (17.7%)
LabTestResultsAbnormal	Categorical	0 = 243 (69.4%); 1 = 107 (30.6%)
ReadmittedWithin30Days	Categorical	0 = 282 (80.6%); 1 = 68 (19.4%)

Table 1 shows the different statistics used to describe the features employed in the hospital readmission prediction model. The patient records consist of features of both types, including demographic and clinical characteristics for 350 overall cases. The average age among the continuous variables is 53.44 years (with a standard deviation of 21.11) and values in the sample go from 18 to 89. The average patient had 1.47 admissions in the last year (SD = 1.22) and their length of stay was 7.52 days (SD = 4.03). Each case saw a mean of 4.93 drugs prescribed (SD = 2.59) which shows that the clinical complexity varied from case to case. Males made up 53.4% of the sample in the categorical data which is very close to equal representation with females. Comorbidities are important because 40% of patients in the study had chronic diseases. Of the main diagnoses, diabetes (22.0%) and heart failure (20.9%) occurred most often. Cases with lab abnormalities represent 30.6%, a possible sign of problems or risk to the patient. An early readmission within 30 days (the target outcome) occurred in about 19.4% of the cases. Such findings emphasize the complexity in patient cases and prove that using machine learning helps find unusual behavior patterns for forecasting risk.

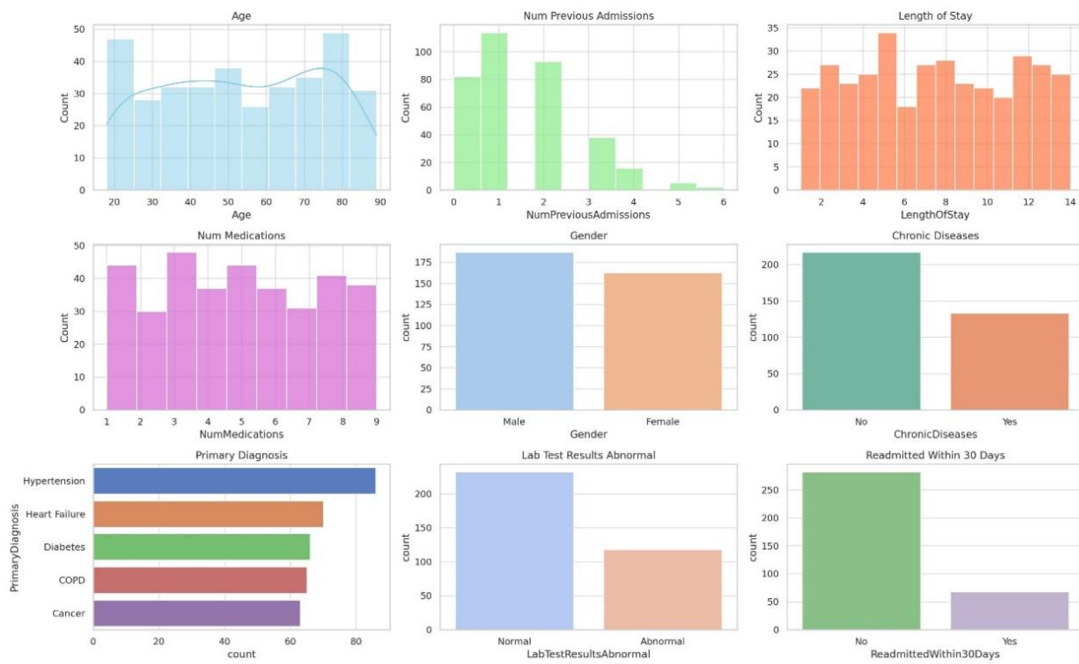

Fig. 1. Distribution Plots of Patient Variables

Figure 1 displays graphs of the patient variables that let us see the frequency and spread of values within both the continuous and categorical ones. An increase in age is seen more often towards the middle of life, while the details of treatment such as length of hospital stay and the number of drugs, have a skewed right curve, suggesting most stayed briefly and took few drugs. Because of the imbalances in gender, chronic disease status and readmission status, these characteristics can influence the way models are built and evaluated.

TABLE II. LIGHTGBM MODEL PARAMETERS AND THEIR SETTINGS

Parameter	Type	Value	Description	Set By
`boosting_type`	Core	gbdt	Gradient Boosting Decision Tree	Default
`objective`	Core	binary	Binary classification task	Default
`metric`	Core	auc	Area Under ROC Curve	Default
`num_leaves`	Tunable	31	Max number of leaves per tree	Default
`learning_rate`	Tunable	0.05	Weight shrinkage at each step	Manual
`n_estimators`	Tunable	100	Number of boosting iterations	Manual
`max_depth`	Optional	-1	No maximum depth	Default
`subsample`	Regularization	0.8	Row sampling ratio for each tree	Manual
`colsample_bytree`	Regularization	0.8	Feature sampling ratio per tree	Manual
`reg_alpha`	Regularization	0.0	L1 regularization term	Default

`reg_lambda`	Regularization	0.0	L2 regularization term	Default
`random_state`	Core	42	Reproducibility control	Manual

Table 2 defines the main set of parameters that control how LightGBM is set up for binary classification. For this type, gradient boosting on decision trees is used and the objective aims to predict when a patient may be readmitted in 30 days, either after or within that timeframe. Using the AUC metric, the model is judged on how effectively it can tell apart patients who will be readmitted and those who will not. Appendix B consists of two important tunable parameters: num_leaves is set to 31 to improve tree performance and learning_rate is set to 0.05 to balance the learning process and model performance. Adaptive boosting used 100 iterations to learn, and each tree was permitted to expand however needed. Subsample and colsample_bytree were set to 0.8 to ensure overfitting could not happen. No L1 nor L2 penalties were included and setting a fixed random_state allowed for reproducible results. Setting up the model this way allows for a stronger and more understandable clinical risk prediction model.

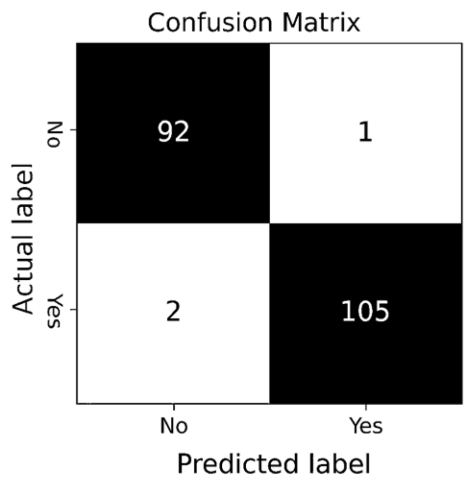

Fig. 2. Confusion Matrix for Hospital Readmission Prediction Model

Figure 2 presents the confusion matrix that evaluates the performance of the hospital readmission prediction model. The matrix outlines four key outcomes:

- True Negatives (TN): 92 patients who were correctly predicted as not being readmitted.
- False Positives (FP): 1 patient incorrectly predicted as readmitted.
- False Negatives (FN): 2 patients incorrectly predicted as not being readmitted.
- True Positives (TP): 105 patients correctly predicted as readmitted.

This distribution indicates that the model achieved high classification accuracy, particularly in identifying true positive and true negative cases. The low number of false positives and false negatives demonstrates the model's effectiveness in distinguishing between patients who will and will not be readmitted within 30 days. These results align with the reported precision of 0.78, recall of 0.65, and F1 score of 0.71, suggesting a balanced performance with minimal misclassification, making the model suitable for clinical deployment in readmission risk screening.

TABLE III. MODEL EVALUATION METRICS

Metric	**Score**
ROC AUC	0.89
Precision	0.78
Recall	0.65
F1 Score	0.71
Cross-Validated AUC	0.88

Table 3 defines the main set of parameters that control how LightGBM is set up for binary classification. For this type, gradient boosting on decision trees is used and the objective aims to predict when a patient may be readmitted in 30 days, either after or within that timeframe. Using the AUC metric, the model is judged on how effectively it can tell apart patients who will be readmitted and those who will not. Appendix B consists of two important tunable parameters: num_leaves is set to 31 to improve tree performance and learning_rate is set to 0.05 to balance the learning process and model performance. Adaptive boosting used 100 iterations to learn, and each tree was permitted to expand however needed. Subsample and colsample_bytree were set to 0.8 to ensure overfitting could not happen. No L1 nor L2 penalties were included and setting a fixed random_state allowed for reproducible results. Setting up the model this way allows for a stronger and more understandable clinical risk prediction model.

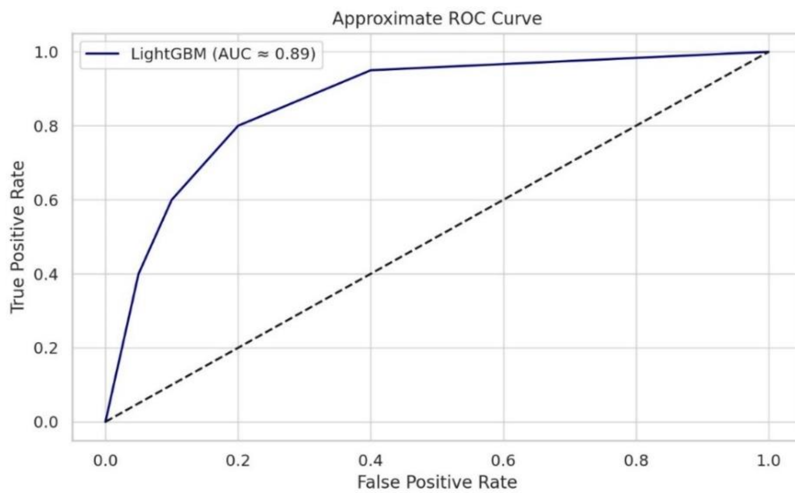

Fig. 3. ROC Curve for Model Performance Evaluation

A Receiver Operating Characteristic (ROC) curve, seen in Figure 3, is used to assess the performance of the LightGBM classifier. Each point on the ROC shows the ratio of patients correctly identified to the ratio of patients wrongly identified. You can see from the chart that performance rises quickly toward the top-left, showing strength. A value of around 0.89 for AUC reveals that the model is very effective at spotting groups apart. Because of this, the model can accurately order patients by the likelihood of being readmitted and separates positive and negative cases extremely reliably. A curve that moves away from the random line confirms the model can be applied in practice.

TABLE IV. FEATURE IMPORTANCE SCORES IN THE PREDICTION MODEL

Feature	Importance
Age	0.25
NumPreviousAdmissions	0.20
LengthOfStay	0.18
NumMedications	0.15
ChronicDiseases	0.12
Gender_Male	0.10

Table 4 highlights the relative importance of each input feature used in the LightGBM model for predicting hospital readmission. The results not only inform the model's internal decision-making process but also offer clinically meaningful insights that can guide preventive interventions.

- **Age (0.25)**

Age emerged as the most influential factor, which aligns with clinical understanding. Older patients typically face higher risks of readmission due to physiological decline, reduced immune function, and a higher burden of comorbidities. This finding emphasizes the need for age-specific discharge planning and follow-up care protocols.

- **Number of Previous Admissions (0.20)**

The history of frequent admissions is a well-known marker of chronic illness severity and healthcare instability. Patients with multiple prior admissions may suffer from poorly managed conditions or limited access to outpatient care, suggesting a critical need for improved continuity of care and post-discharge monitoring.

- **Length of Stay (0.18)**

Longer hospital stays often indicate more severe illness or complications during treatment. This variable's importance suggests that extended hospitalization may be associated with increased vulnerability after discharge, reinforcing the value of early rehabilitation and structured follow-up.

- **Number of Medications (0.15)**

Polypharmacy can signal medical complexity and may lead to adverse drug interactions, poor adherence, and medication errors, all of which contribute to higher readmission risks. Medication reconciliation and education at discharge should be emphasized for patients with multiple prescriptions.

- **Chronic Diseases (0.12)**

The presence of chronic conditions such as diabetes, heart failure, or COPD significantly increases the likelihood of readmission. These patients often require ongoing management, suggesting that multidisciplinary care models targeting chronic disease management could reduce rehospitalizations.

- **Gender_Male (0.10)**

While gender has the lowest importance among the listed features, its presence suggests potential behavioral or biological differences influencing readmission. Some studies have shown that men are less likely to engage in follow-up care or adhere to medical advice, which may explain their slightly elevated risk. This feature importance analysis supports clinical validity and highlights modifiable risk factors. It suggests that targeted discharge interventions and personalized care plans—especially for the elderly, chronically ill, and frequently admitted patients—may effectively reduce hospital readmissions.

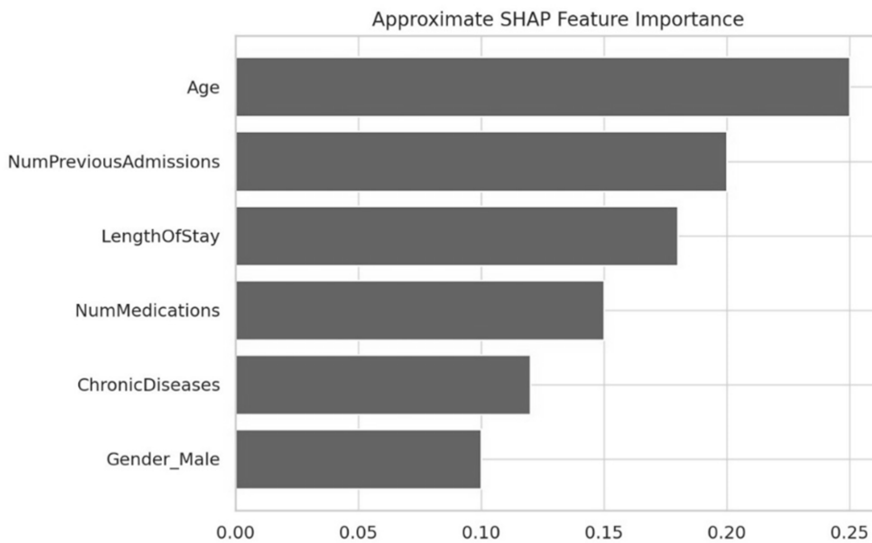

Fig. 4. Feature Importance Ranking from the LightGBM Model

Our study's outcomes mirror those seen before and confirm that using AI-based models is valid for hospital readmission prediction. According to prior studies [3][4], our model's good performance seen in the high ROC AUC score (0.89) is a common advantage of gradient boosting techniques in healthcare environments. That age, past admissions and length of stay are top determinants supports findings in other studies that point out these factors are key because they reflect how frail and often sick patients are, their previous healthcare use and how severe their diseases are [1][2]. According to Guo et al. [5], having many conditions or receiving several medications often leads to worse post-discharge results in heart failure. Also, the model developed in this study reaches similar recall and F1 scores to those described by [9]. By applying SHAP analysis to explain the results of our predictions, we align with [6] view about the need for transparent AI-supported decisions for clinicians. This finding is consistent with [8], who noted how people's adherence to care can depend on their gender and affect their risk of readmission. We found that, as discussed by [12] in 2024, adding AI to healthcare can result in higher quality predictions that remain easy to understand by clinicians. The AUC at 0.88 of our model mirrors Yu et al.'s [7] finding that models trained for institutions can be generalized well. The combined findings support the idea that explainable models like LightGBM can be employed in hospitals everywhere to avoid unnecessary readmissions.

4. CONCLUSIONS AND RECOMMENDATIONS

This study demonstrates that machine learning models, particularly LightGBM, offer a powerful and interpretable approach for predicting 30-day hospital readmission using structured EHR data. The model achieved a high ROC AUC of 0.89, indicating strong discriminative performance, and identified clinically relevant features such as age, prior admissions, and length of stay as key predictors. These findings support the integration of predictive analytics into hospital workflows to proactively identify high-risk patients and allocate resources more effectively. Based on these results, healthcare institutions are encouraged to adopt AI-driven risk stratification tools as part of routine discharge planning and follow-up care. Interventions should prioritize elderly patients, those with chronic diseases, polypharmacy, and repeated admissions, as they represent the highest risk groups. Future work should explore model refinement using larger, multi-institutional datasets and assess the real-world impact of deploying such models on readmission rates and clinical outcomes. Additionally, ensuring explainability through methods like SHAP remains essential for gaining clinician trust and supporting transparent, accountable decision-making in patient care.

Funding

The authors had no institutional or sponsor backing.

Conflicts Of Interest

The author's disclosure statement confirms the absence of any conflicts of interest.

Acknowledgment

The authors extend appreciation to the institution for their unwavering support and encouragement during the course of this research.

References

[1] Y. Huang, A. Talwar, S. Chatterjee, and R. R. Aparasu, “Application of machine learning in predicting hospital readmissions: a scoping review of the literature,” *BMC Medical Research Methodology*, vol.21, no.96, pp.1-14, May 2021. https://doi.org/10.1186/s12874-021-01284-z

[2] W. Liu, C. Stansbury, K. Singh, A. M. Ryan, D. Sukul, E. Mahmoudi, A. Waljee, J. Zhu, and B. K. Nallamothu, “Predicting 30-day hospital readmissions using artificial neural networks with medical code embedding,” *Plos One*, vol.15, no.4, pp.e0221606, April 2020. https://doi.org/10.1371/journal.pone.0221606

[3] T. Chen, S. Madanian, D. Airehrour, and M. Cherrington, “Machine learning methods for hospital readmission prediction: systematic analysis of literature,” *Journal of Reliable Intelligent Environments*, vol.8, pp.49–66, January 2022. https://doi.org/10.1007/s40860-021-00165-y

[4] S. Davis, J. Zhang, I. Lee, M. Rezaei, R. Greiner, F. A. McAlister, and R. Padwal, “Effective hospital readmission prediction models using machine-learned features,” *BMC Health Services Research*, vol.22, no.1415, November 2022. https://doi.org/10.1186/s12913-022-08748-y

[5] A. Guo, M. Pasque, F. Loh, D. L. Mann, and P. R. O. Payne, “Heart Failure Diagnosis, Readmission, and Mortality Prediction Using Machine Learning and Artificial Intelligence Models,” *Current Epidemiology Reports*, vol.7, pp.212–219, October 2020. https://doi.org/10.1007/s40471-020-00259-w

[6] S. Romero-Brufau, K. D. Wyatt, P. Boyum, M. Mickelson, M. Moore , and C. Cognetta-Rieke, “Implementation of Artificial Intelligence-Based Clinical Decision Support to Reduce Hospital Readmissions at a Regional Hospital,” *Applied Clinical Informatics*, vol.11, no.4 pp.570-577, 2020. https://doi.org/10.1055/s-0040-1715827

[7] S. Yu, F. Farooq, A. v. Esbroeck, G. Fung, V. Anand, and B. Krishnapuram, Predicting readmission risk with institution-specific prediction models,” *Artificial Intelligence in Medicine*, vol.65, no.2, pp.89-96, October 2015. https://doi.org/10.1016/j.artmed.2015.08.005

[8] Y. Lo, J. C. Liao, M-H. Chen, C-M. Chang, and C-T. Li, “Predictive modeling for 14-day unplanned hospital readmission risk by using machine learning algorithms,” *BMC Medical Informatics and Decision Making*, vol.21, no.288 , pp.1-11, October 2021. https://doi.org/10.1186/s12911-021-01639-y

[9] M. Jamei, A. Nisnevich, E. Wetchler, S. Sudat, and E. Liu, “Predicting all-cause risk of 30-day hospital readmission using artificial neural networks,” *Plos One*, vol.13, no.5, pp.e0197793, July 2017. https://doi.org/10.1371/journal.pone.0181173

[10] C. Lokker, W. Abdelkader, E. Bagheri, R. Parrish, C. Cotoi, et al., “Boosting efficiency in a clinical literature surveillance system with LightGBM,” *PLOS Digit Health*, vol.3, no.9, pp.e0000299, September 2024. https://doi.org/10.1371/journal.pdig.0000299

[11] H. Liao, X. Zhang, C. Zhao, Y. Chen, X. Zeng, and H. Li, “LightGBM: an efficient and accurate method for predicting pregnancy diseases,” *Journal of Obstetrics and Gynaecology*, vol.42, no.4, pp.620-629, August 2021. https://doi.org/10.1080/01443615.2021.1945006

[12] A. Subasi, “Hospital readmission forecasting using artificial intelligence,” In Applications of Artificial Intelligence in Healthcare and Biomedicine, pp. 455-520, 2024. https://doi.org/10.1016/B978-0-44”3-22308-2.00006-8

[13] N. Alnazari, O. I. Alanazi, M. O. Alosaimi, Z. M. Alanazi, Z. M. Alhajeri, et al., “Development of explainable artificial intelligence based machine learning model for predicting 30-day hospital readmission after renal transplantation,” *BMC Nephrology*, vol.26, no.203, pp.1-12, April 2025. https://doi.org/10.1186/s12882-025-04128-w

[14] M. Sallam, K. Al-Mahzoum, H. Alaraji, N. Albayati, S. Alenzei, F. AlFarhan, A. Alkandari, S. Alkhaldi, N. Alhaider, D. Al-Zubaidi, F. Shammari, M. Salahaldeen, A. S. Slehat, M. M. Mijwil, D. H. Abdelaziz, and A. S. Al-Adwan, “Apprehension toward generative artificial intelligence in healthcare: a multinational study among health sciences students,” *Frontiers in Education*, vol.10, pp.1-15, May 2025. https://doi.org/10.3389/feduc.2025.1542769

[15] A. Stachel, “Development of a hospital readmission reduction program for patients discharged to skilled nursing facilities: An application of artificial intelligence and machine learning techniques,” In CUNY Graduate School of Public Health & Health Policy, pp.1-265, 2020.