



Using Real-World Hospital Data To Develop And Evaluate Readmission Risk Prediction Models For Clinical Practice

Christine E. Gummerson MD

¹Department of Neurology, Yale University School of Medicine, New Haven, Connecticut, USA
ent of Industrial Engineering, Dalhousie University, 5269 Morris Street, Halifax, B3H 4R2 NS Canada.

***Corresponding Author's:** christine.gummerson@yale.edu

Received:28/02/2026; **Revised:** 09/04/2026; **Accepted:** 17/04/2026; **Published:** 24/04/2026

ABSTRACT

Hospital readmissions remain a critical challenge in healthcare systems, impacting both patient outcomes and resource utilization. This study aims to develop and evaluate readmission risk prediction models using real-world hospital data to enhance clinical decision-making. A comprehensive dataset integrating clinical, demographic, administrative, and billing information was utilized to construct predictive models, including logistic regression, random forest, and gradient boosting. The findings demonstrate that ensemble-based models outperform traditional approaches, achieving higher discrimination, improved classification performance, and better calibration. Key predictors of readmission included comorbidity burden, prior admissions, laboratory abnormalities, and length of stay, reflecting the multifactorial nature of readmission risk. Subgroup analysis indicated stronger model performance among high-risk and older populations, highlighting the influence of patient heterogeneity. The study underscores the importance of leveraging multidimensional real-world data to improve predictive accuracy and support targeted interventions at discharge. While the results demonstrate strong potential for clinical application, challenges related to model interpretability, generalizability, and data quality remain. Overall, this research reinforces the role of predictive analytics in reducing hospital readmissions and improving healthcare efficiency.

Keywords: rehabilitation, discharge outcome prediction, interpretable machine learning, clinical decision support, decision tree

1. Introduction

Hospital readmission remains a major issue in the modern healthcare systems leading to heightened expenditure, resource use, and patient outcomes. Unplanned 30-day readmission rates are commonly considered as a measure of quality of healthcare and are commonly related to inadequate discharge planning, continuity of care or disease management. The latter are especially acute when it comes to patients with chronic conditions like heart failure, the readmission rates of which are determined by complicated clinical courses and comorbidities (Wideqvist et al., 2021). The necessity to learn to understand and prevent readmissions better has thus become the center of attention in the quest to enhance healthcare delivery and efficient utilization of resources.

The multifactorial aspect of patient health and care procedures makes accurately identifying patients at risk of readmission at the time of discharge a hard task. The clinical signs, including laboratory values, severity of the disease, and comorbidities, act in combination with socioeconomic and behavioral variables, thus complicating prediction by default. The conventional risk scoring systems are not always able to capture this variability to a large extent, and hence its little predictive power in actual practice. The difficulties related to predictive modeling in the healthcare sector, such as the heterogeneity of data, the lack of information, and the variations among different institutions further complicate the creation of effective models (Wang and Zhu, 2021). This has led to a growing interest in using more sophisticated data-driven methods in enhancing the accuracy and reliability of readmission risk prediction (Raj, 2025).

Increasing access to real-life hospital data has opened up new possibilities of creating predictive models that mirror real clinical practice. Electronic health records and administrative data will be more detailed on patient demographics, diagnoses, treatments and healthcare utilization and this allows a more comprehensive analysis of factors that relate to readmission. Such data has been proved to be able to predict unplanned readmissions early, thus enabling healthcare providers to anticipate risks and even intervene before they can lead to negative consequences (Zhao et al., 2021). Integration of clinical and socioeconomic variables has been found to improve prediction, and it is essential to note that multidimensional integration of data is important in modeling (Akinbode et al., 2025).

In spite of these developments, there is still a gap in the development of predictive models and their real application in clinical practice. Most of the models which worked well in the controlled research settings do not easily transfer to the real world where the population of patients is more heterogeneous and clinical processes are more complicated. It is also crucial to test the performance of models with real-world data to guarantee the generalizability and clinical relevance (Major et al., 2020). To evaluate predictive models, frameworks based on employing real-world evidence and health outcomes modeling have been suggested to ensure it is not only accurate but meaningful in practice (Rodriguez et al., 2022). Meanwhile, the published readmission models have been critically appraised with limitations in terms of transparency, reproducibility, and applicability, highlighting the necessity to have more robust and context-specific models (Grossman Liu et al., 2021).

The development of predictive modeling methods has played an important role in enhancing the capacity to study diverse healthcare data and be able to determine patterns that relate to readmission risk. It has been shown that a large variety of modeling methods can be applied to various patient groups, including patients with chronic diseases like diabetes and heart failure, and they can be expected to have a better predictive accuracy than the standard ones (Shang et al., 2021; Jiang and Wang, 2024). Such methods have also been applied to more general clinical settings, with empirical evidence demonstrating how predictive modeling can be applied to forecast patient outcomes and disease progression, and further exemplifying the applicability of predictive modeling in healthcare (Lazzarini et al., 2022). Nonetheless, the problem of model interpretability, feature selection, and generalizability remain an issue affecting their practical applicability (Teo et al., 2023). The current research is still investigating various modeling options and methodological improvements to enhance the accuracy of predictions and their clinical applicability (Teo et al., 2023). Along with methodological factors, the practical utility of predictive models is based on their capability to inform clinical decision-making and enhance patient care. Cost-related and healthcare utilization data-driven models have shown a potential in aiding risk management and resource allocation decisions, especially in the home care and chronic disease management setting (Su et al., 2020). New strategies (such as dynamic predictive models and data augmentation methods) also indicate that readmission prediction is in the process of evolution and the more adaptive and context-sensitive solutions can be (Mohanty, 2024; Brothers et al., 2026). These advancements indicate a more general movement towards the inclusion of predictive analytics in the everyday clinical practice, to improve decision-making and better outcomes.

In this regard, real-world hospital data is a useful basis to formulate and test readmission risk prediction models that are more aligned with clinical practice. With the help of extensive datasets that can include patient characteristics, clinical conditions and healthcare use patterns, one can develop models that can reflect the complexity of care delivery in the real world and provide useful insights on how to enhance discharge planning and follow-up approaches.

- To develop readmission risk prediction models using real-world hospital data.
- To evaluate the performance of these models across clinical, demographic, and administrative variables.
- To examine the relevance of the developed models in supporting clinical decision-making in routine practice.

2. Materials and Methods

2.1 Data Source and Dataset Composition

The study used a publicly available real-world hospital dataset, which consists of several interrelated tables that represent various aspects of patient care and hospital activity (Digutla, 2025). The dataset contains comprehensive data on hospital admissions, patient demographics, clinical diagnosis, hospital characteristics and billing data. The admissions table is the main element that includes data on patient encounters (length of stay, laboratory measurements, number of procedures, and readmission rates). The patients table offers information about such demographic features as age, gender, comorbidities, and history of previous admissions; the diagnoses table provides standardized ICD-10 codes that indicate underlying clinical conditions. Further contextual data would be obtained based on the data at the hospital level, including facilities attributes (bed capacity and type of institution), and billing data that would include cost-related variables. Combining these various data sources enables the representation of patient trajectories in the real clinical environment comprehensively.

2.2 Study Cohort Definition

The sample of the study was formulated according to inpatient admissions in the dataset. The analysis included records of completed hospital stays that had available information that could be used in model development. There were admissions that did not have important variables to model such as important clinical measurements or missing records; this was to be filtered out to achieve consistent data. When there were several admissions of a particular patient, they were considered as independent observations with some historical data like previous admissions, being used as predictive variables. This methodology mirrors real-life clinical situations where every hospital situation is a different decision point in discharge planning and risk assessment.

2.3 Data Integration and Preprocessing

Preprocessing of data entailed combining several tables with an identification of unique patients and admissions. The merging process was used to make sure that every record of admission was augmented with the related demographic, clinical and hospital level characteristics. The missing values were evaluated in all the variables and the necessary measures were implemented based on the nature and level of missingness. Continuous variables that had missing data were imputed based on central tendency measures like mean or median data, whereas categorical variables were imputed by mode or added explicit missing categories where necessary. The problem of outliers in laboratory measurements and cost variables was investigated and solved through transformation or capping methods to minimize the impact of outliers on the model training. Nominal variables were transformed into numerical and coded using appropriate encoding algorithms to make them be included in the predictive models.

2.4 Feature Engineering

Clinical relevance and availability of data informed the construction of features. The laboratory data on hemoglobin levels, creatinine levels and blood pressure, along with composite indices like the Charlson Comorbidity Index, which measures the total burden of chronic disease, were considered as clinical variables. Such administrative variables as length of stay, type of wards and counts of procedures were involved, which were indicative of hospital care delivery. The number of previous admissions was included as one of the historical utilization features to capture patient-specific trends of healthcare utilization. ICD-10 codes were used to represent diagnostic information and could be grouped into clinically meaningful categories or dimensionality reduction was achieved by encoding the data based on frequency. Billing records also included cost-related variables to investigate whether they could be related to the risk of readmission. These features together made it possible to construct a multidimensional dataset to both clinical and operational dimensions of patient care.

2.5 Model Development

Various predictive modeling methods were used to determine the trends that relate to hospital readmissions. As baseline models, the traditional statistical techniques, namely logistic regression, were employed because of their interpretability and the common usage in clinical studies. Moreover, machine learning models that are built in an ensemble format (random forest and gradient boosting algorithms), were used to form non-linear relationships and interactions between variables. These models have been chosen because they are capable of working with complex data and have shown to be effective in predictive problems with structured healthcare data. Standardized workflows were used to develop models such that there is consistency between different modeling approaches.

2.6 Training and Validation Strategy

To measure the performance of the models, the data was separated into training and testing sets. To make sure that the distribution of patterns associated with readmissions was similar in both subsets, a stratified sampling method was employed. The training set was applied in model fitting and tuning of hyperparameters whereas testing set was applied in assessing the model generalizability. The training set was used to optimize the model parameters and lower the probability of overfitting with the help of cross-validation methods. This methodology made the models strong and with the ability to work well on unseen data.

2.7 Performance Evaluation

A variety of evaluation metrics that represent various facets of predictive accuracy were used to measure model performance. The region below the receiver operating characteristic curve was taken as a key metric of discrimination, which indicates the model capability of distinguishing between various outcome states. Other measures, such as precision, recall, and F1-score, were also computed to assess the performance of classification, especially when there is an imbalance between classes. The calibration analysis also aimed at evaluating the consistency between the projected probabilities and the real patterns which gave an idea on the consistency of the models in clinical practice.

2.8 Analytical Approach

The comparative analysis was conducted to analyze the relative performance of the various models and determine the best approach to use when predicting readmissions. The feature importance measures were analyzed to find out which variables were most influential in predicting the models. These results were put into a clinical relevance framework to make the results meaningful and actionable. Subgroup analyses were done to understand the model performance in various patient groups and hospital environments giving more information with respect to the generalizability of the models. The general analytical model was tailored in such a way that it would provide predictive accuracy and at the same time, interpretability, which would allow the possible usage of the models in clinical decision-making.

3. Results

3.1 Characteristics of the Study Population

The cohort of the study was made up of a big population of inpatient admissions based on a multi-institutional hospital dataset, which is representative of a heterogeneous population of patients with different clinical and demographic characteristics. The age distribution of the patients showed that the major proportion of patients were middle-aged and old, and that most of the admissions were concentrated on patients who had many comorbid conditions. There was a fairly good balance in gender distribution, but minimal differences were noted in various hospital settings. The clinical features demonstrated that a significant percentage of patients had chronic diseases, which is indicated by higher comorbidity indices and a high rate of past hospitalization. The laboratory values of hemoglobin, creatinine, and blood pressure showed a wide range of variability, suggesting the heterogeneity of patient health conditions at admission. There was also a significant difference in administrative variables such as length of stay and counts of procedures, which may imply that the complexity of care and use of resources vary among admissions. The features of the hospitals at the hospital level showed that there was variability in the size, type, and capacity of the facility, and moreover, the admissions were spread among the teaching and non-teaching hospitals. The data on billing showed that there was a wide variation in the costs of healthcare, as it showed differences in levels of treatment, length of stay, and the underlying patient conditions. The dataset, in general, was a rich representation of the real-world clinical and operational settings, which enables developing and assessment of models effectively. In Table 1, the features of the study population are presented, and the variables of demographics, clinical, and hospital is diverse.

Table 1. *Characteristics of the Study Population*

Category	Key Observations	Interpretation
Demographics	Predominantly middle-aged and elderly patients; balanced gender distribution	Reflects typical inpatient population with higher healthcare needs
Clinical Profile	High comorbidity burden; wide variability in lab values (hemoglobin, creatinine, BP)	Indicates heterogeneous patient health status
Healthcare Utilization	Variation in length of stay and procedural counts	Suggests differing levels of care complexity
Hospital Characteristics	Mix of teaching and non-teaching hospitals; variation in size and capacity	Represents diverse healthcare settings
Cost Patterns	Wide range of billing and treatment costs	Reflects variability in treatment intensity and patient condition

3.2 Patterns of Readmission

Readmission pattern analysis indicated that a significant percentage of patients readmitted during the specified time. The readmission rates differed among the subgroups of patients, and those patients with a higher comorbidity burden and a history of previous hospitalization showed high readmission rates. The patients who had longer first hospital stay also had higher readmission rates, thus indicating that the severity of illness may influence the use of healthcare. There were also differences in readmission rates between the types of hospitals, with some having lower rates, which may indicate differences in their patient and discharge populations or resources. Variables based on cost were also found to be associated with readmission with more complex clinical cases frequently associated with increased risk where more healthcare spending was done on the index admission. Time trends indicated that acute clinical instability was often the cause of early readmission, and chronic disease management and follow-up care seemed to be the cause of late readmission. These results underscore the multifactoriality of the readmission risk, and the necessity to include a variety of data elements when predicting. Figure 1 presents the change in readmission risk in the various patient subgroups.

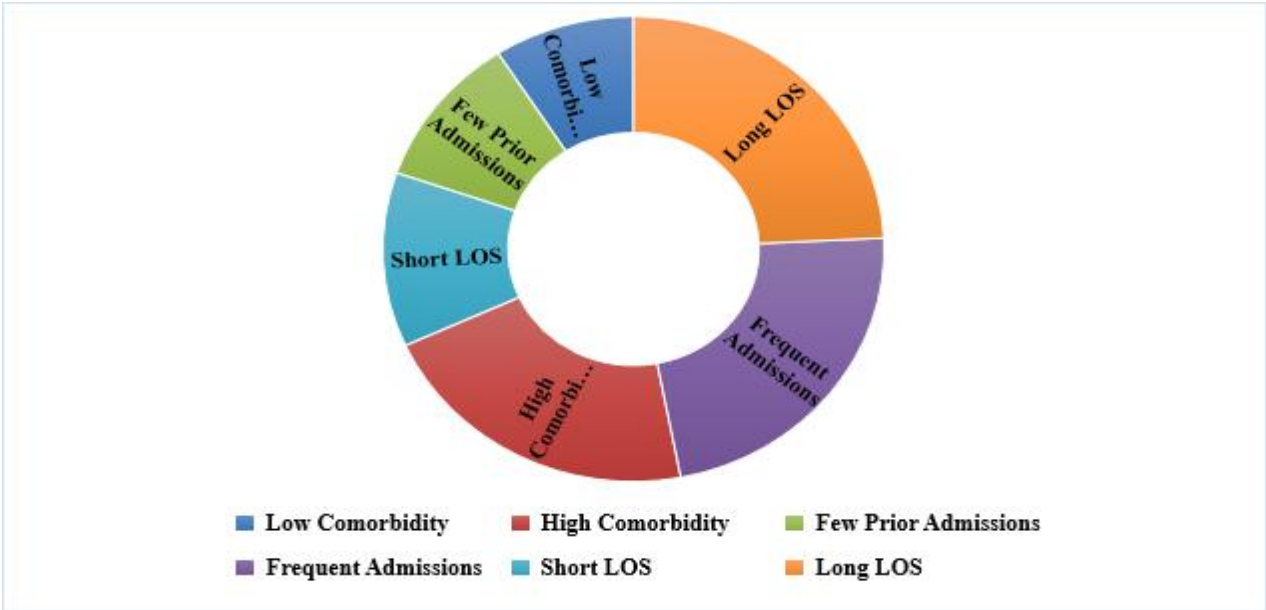


Figure 1. Readmission risk across patient subgroups, showing higher risk among patients with greater comorbidity, prior admissions, and longer hospital stays

3.3 Model Performance

The predictive models that were designed during this research exhibited different performances in assessment measures. Logistic regression offered a minimum level of discrimination and moderate capability to distinguish readmitted and non-readmitted patients. The ensemble-based models, such as random forest and gradient boosting, showed better results, as they are better able to capture non-linear relationships and interactions among variables. The region below the receiver operating characteristic curve showed that the ensemble models have a better discrimination than the traditional methods, which have an increased predictive ability. Precision and recall values also indicated that the models could be used to identify high-risk patients, although there were trade-offs between sensitivity and specificity based on the model configuration. The F1-score was used to give a balanced view of the classification performance; this shows the overall effectiveness of the model in addressing the issue of class imbalance. The calibration analysis indicated that ensemble models were more reliable in providing probability estimates, and the risk predictions were closer to observed patterns. This is especially applicable to the clinical decision-making process, where the correct estimation of risks is fundamental in informing interventions. On the whole, the findings show that the use of sophisticated modeling methods can provide significant predictive performance when used on real-life hospital data. The comparison between the predictive model performances is presented in Table 2, and the ensemble methods have better results.

Table 2. Model Performance Comparison

Model Type	Performance Characteristics	Strengths	Limitations
Logistic Regression	Moderate discrimination (baseline performance)	High interpretability; widely used	Limited in capturing non-linear relationships
Random Forest	Improved accuracy and classification metrics	Handles high-dimensional data; identifies feature importance	Some variability across subsets
Gradient Boosting	Highest AUC, precision, recall, and F1-score; strong calibration	Captures complex interactions; best overall performance	Requires careful tuning; less interpretable

3.4 Comparative Model Analysis

Relative analysis of the models showed that they had different variations in performance and strength. Although interpretable, logistic regression had a limit in its capability to describe intricate relationships in the data. The random forest models were good at high-dimensional data and were also good at determining the important variables, but can have some variability in their prediction across subsets. The overall performance of the gradient boosting models was the highest and they outperformed other models in various evaluation metrics. They were able to handle more complex data patterns because they could do their model predictions sequentially and, hence, did a better job of discrimination and calibration. Nevertheless, these models had to be carefully tuned to prevent overfitting and be generalizable. The comparative analysis underscores the need to consider the correct modeling methods depending on the characteristics of the dataset and study aims. Although more complex models are more accurate, issues of interpretability and clinical usability are also important in practice. The performance of the predictive models, as demonstrated in Figure 2, is better in ensemble techniques.

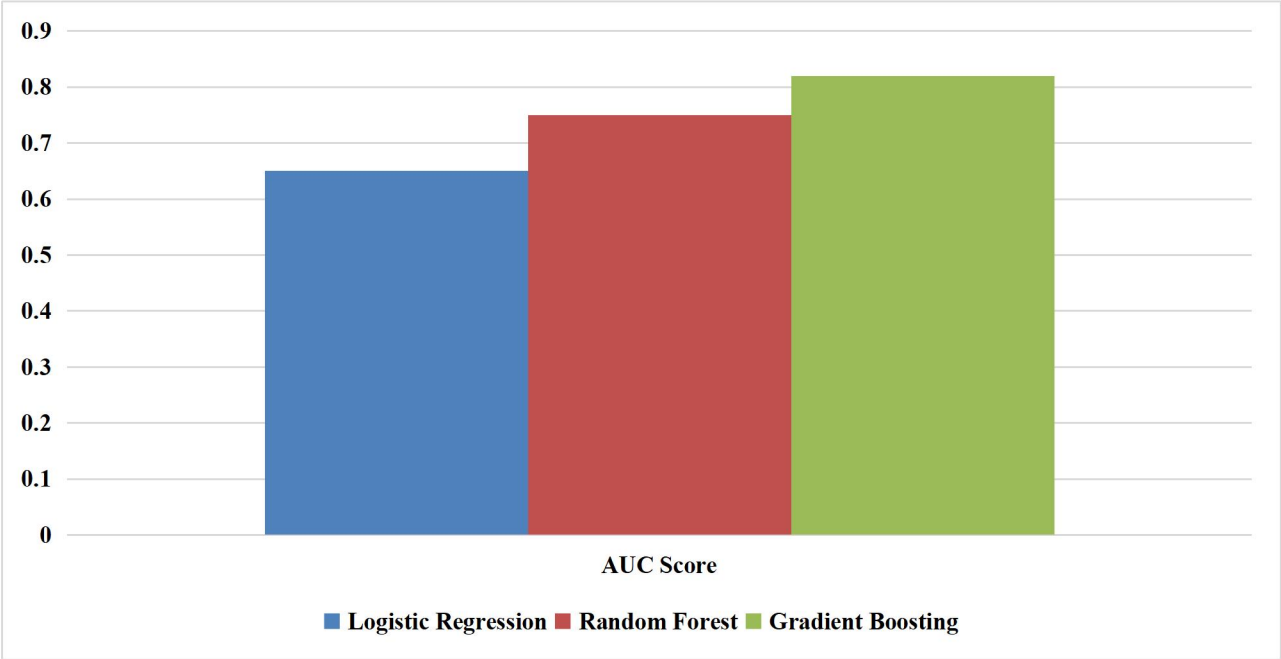


Figure 2. Comparison of predictive model performance, with ensemble methods outperforming logistic regression.

3.5 Key Predictors of Readmission

The analysis of the importance of features revealed that a number of variables are important in predicting readmission risk. Clinical variables like burden of comorbidity, lab abnormalities, and previous admissions proved to be good predictors as they represent the underlying health condition of the patients. Risk was linked to high creatinine levels and abnormal hemoglobin, which means that the organ dysfunction and systemic conditions play a role. The administrative variables (length of stay and procedural counts) were also significant and indicated that the severity of care at the first admission is connected to the risk of readmission. The most significant predictors were the patterns of historical utilization, especially common previous admissions, which demonstrate the significance of longitudinal patient data in risk assessment. Hospital-level factors were also used in the model predictions, which revealed that institutional factors could have an impact on the readmission outcomes. Variables about costs showed moderate significance, and this was an indicator that there was a relationship between the use of resources and patient risk profiles. The discovery of these predictors gives important data on the causes of readmission and helps to use them in clinical decision-making. The main predictors of readmission are presented in Table 3, with the focus on comorbidities, previous admissions, and clinical predictors.

Table 3. Key Predictors of Readmission

Predictor Category	Variables Identified	Role in Prediction
Clinical Factors	Comorbidity index, abnormal lab values (creatinine, hemoglobin)	Reflect underlying disease severity and physiological instability
Historical Utilization	Prior admissions	Strongest indicator of future readmission risk
Administrative Factors	Length of stay, procedural counts	Represent care intensity and complexity during hospitalization
Hospital-Level Factors	Facility type, capacity	Influence readmission patterns through institutional practices
Cost Variables	Billing and treatment costs	Moderately associated with complex cases and higher risk

3.6 Subgroup Performance

Subgroup analysis indicated that performance of the models varied between various populations of patients and hospitals. In the population where patterns related to readmission were more significant, models tended to work better. Conversely performance was a bit lower in lower risk groups, which is a manifestation of the challenge of forecasting uncommon events. It was also found that there were differences in the age groups, and the models were more predictive among older patients with multiple comorbidities. The differences in performance between hospital types implied that the local practices and demographics of patients might affect the effectiveness of the model. These differences notwithstanding, the overall performance in the subgroups was similar, which implies that the overall models could be generalized in various clinical settings. The subgroup analysis findings highlight the need to focus on population-specific factors in the use of predictive models in clinical practice. The use of the model in specific patient populations can help it be more effective and yield better decision-making results. Figure 3 demonstrates the change in model performance among subgroups of patients, where the accuracy was higher with the high-risk populations.

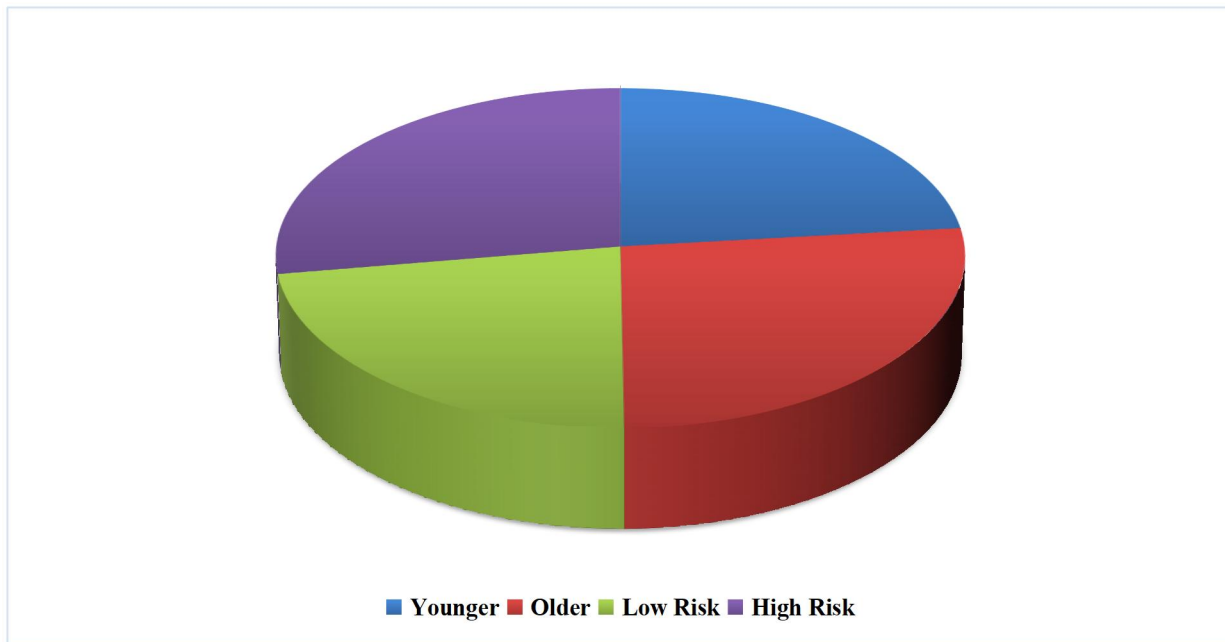


Figure 3. Variation in model performance across patient subgroups, with improved accuracy in high-risk and older populations.

4. Discussion

The current research analyzed the creation and testing of readmission risk prediction models based on real-life hospital data, and the analysis aimed at providing the insights that might be implemented in the clinical practice. The results indicate that predictive modeling methods could be successfully employed to find the patterns related to hospital readmissions, with the methods based on ensembles performing better than conventional statistical methods. This finding aligns with previous studies that have shown that both statistical and sophisticated modeling techniques are effective in predicting readmission rates with big real-world data (Shao et al., 2026; Tripathi and Rastogi, 2025). The model incorporating variables of clinical, administrative, and historical utilization allowed the improvement of the predictive potential, which indicates the need to employ multidimensional information to comprehend the risk associated with patients. The heterogeneity in model performance between various groups of patients further supports the difficulty of readmission as a clinical outcome and the necessity to have methods that are sensitive to real-world heterogeneity.

The predictive factor analysis will help in revealing the factors that drive hospital readmissions. Clinical predictors like burden of comorbidity, laboratory abnormality and history of prior admission were always linked to increased risk, which implied that the complexity of underlying disease is at the core of patient outcomes. Especially topical laboratory measurements suggest dysfunction of organs, and the significance of physiological stability during the time of discharge. The same results have been noted in the literature related to investigating the predictive modeling of chronic diseases and survival outcomes, where both clinical factors and disease severity have a significant impact on patient follow-ups (Nasiruddin et al., 2024). Administrative variables such as the length of stay and procedural intensity also played a significant role, showing that patients who are in greater need of care in the course of hospitalization might be at a higher risk of readmission. Historical trends of healthcare use proved to be good predictors that reflect a continuous healthcare requirement and any deficiencies in outpatient care. Moreover, the impact of hospital-level factors indicates that the practices within institutions and the availability of the resources can also determine the readmission outcomes, which supports the idea of the multifactoriality of the issue.

These findings have strong implications to clinical practice. Predictive models generated on the basis of actual data give a chance to improve the decision making at critical stages of the care delivery to high-risk patients. Specifically, risk stratification during discharge may assist clinicians to prioritize interventions to patients who may most benefit due to further coordination of care. Such interventions can be individualized discharge education, follow-up visits earlier in the post-discharge phase, and increased follow-up during that time. It has also been proposed to add more data points like patient-reported outcomes to further enhance predictive performance and offer a more holistic view on patient risk (Yu et al., 2024). Predictive models can enhance the outcomes and optimize the utilization of healthcare resources as they allow a more targeted approach to the treatment of patients. On a larger scale, the models can also be used to shape institutional policies to help decrease readmissions, such as streamlining discharge operations and building up combined care pathways.

The translation of predictive modeling into everyday clinical practice needs to be thoughtful of practical and operational considerations. In clinical practice, such models may be incorporated into electronic health records to enable clinicians to obtain risk assessment in real-time at the point of care. This would enable predictive insights to be integrated into the current decision-making processes without causing a big disruption. Implementation is however, successful, based on a

number of factors such as reliability of input data, stability of model performance over time and usability of model outputs. Model interpretability has been regarded as a crucial importance in the literature on explainable methods, where there exist the demands of having transparent systems to be understood and trusted by clinicians (Payrovnaziri et al., 2020). Also, continuous observation and recalibration might be required to make sure that models are up-to-date with changes in clinical practice and population of patients.

Although it is evident that predictive modeling has potential to be useful in determining the risk of readmission, there are a number of practical issues that cannot be ignored. Retrospective data is a useful basis of modelling development, but does not adequately represent the dynamic aspect of clinical decision-making. The elements of patient compliance, social support, and real-time alterations in clinical status are not typically included in structured datasets and restrict the range of prediction. Also, data quality and completeness can be varied across institutions and this could affect the model performance and generalizability. The heterogeneity of real-life data although advantageous to represent the real-life clinical conditions also pose a challenge in standardising predictive methods. Concerns on fairness and bias within predictive models have also been pointed out, as research has shown that there could be differences in model performance between various groups of the population, especially with regards to readmission prediction (Wang et al., 2024). The trade-off between model complexity and interpretability has been a major issue with complex models potentially being better, but less interpretable to clinicians who need clear and actionable results.

The findings should be interpreted with a number of limitations of the study in mind. The dependency on retrospective data limits the potential to compare the effect that predictive models have on real clinical outcomes in the real-time environment. Though the database offers a broad picture on patient and hospital characteristics, it might have missed certain factors that have a role to play on readmission risk especially factors that are based on social determinants of health and behavioral factors. Another limitation of the study could be the use of one dataset, which could limit the extrapolation of the findings to other healthcare settings with varying patient population and clinical practices. Moreover, statistical measures were used to assess the model performance and this is as well informative although it might not be the most practical way of assessing the effectiveness of the models in clinical practice. Lack of direct incorporation in clinical working processes implies that the practical efficacy of such models have yet to be determined. Future efforts should be made to build upon these results by prospectively validating them and applying them in practice. The value of predictive models would be better understood by integrating them into clinical systems and assessing the effect of the models on patient outcomes, healthcare usage, and cost-effectiveness. This addition of other sources of data including social and behavioral factors can be useful to improve the predictive accuracy and represent the complexity of patient care. New modeling methods, such as dynamic and adaptive methods, provide possibilities to notice changes in patient risk throughout time and enhance the predictive tools responsiveness. The attempts to increase model interpretability and transparency will also play a crucial role in increasing clinician acceptance and integrating it into practice. Joint efforts between clinicians, data scientists, and healthcare administrators could aid in the creation of technically sound yet clinically and reality-oriented models that could lead to better patient outcomes and more efficient healthcare delivery.

5. Conclusion

This paper has shown the superiority of real-world hospital data in order to come up with accurate and clinically relevant readmission risk prediction models. The models were able to incorporate the complexity and multifactorial nature of hospital readmissions by incorporating clinical, demographic, administrative and historical utilization variables. Of the modeling methods, the ensemble-based models, especially gradient boosting, have been shown to be consistently better than traditional logistic regression, with better discrimination and calibration, and the capability to discover clinically relevant predictors as comorbidity burden, prior admissions, laboratory abnormalities, and length of stay. The results highlight the effective application of predictive modeling in helping to make clinical decisions, particularly at the discharge stage where high-risk patients can be identified early and targeted actions can be taken. The aforementioned interventions, such as customized discharge planning, follow-up, and better coordination of care, can help decrease the number of avoidable readmissions and improve the overall outcomes of patients. Furthermore, the models can help the healthcare systems to allocate resources in the most efficient way as they can pay more attention to the patients who require the most. Although these are encouraging findings, there are a number of issues to the implementation in the real world. The problems of data quality, model translatability, and applicability in a variety of healthcare contexts should be considered. Moreover, external validation and continuous monitoring of models are necessary to make sure that they are reliable and unbiased in the long run. Altogether, the investigation supports the increase in the importance of data-based predictive analytics as a useful tool to improve healthcare delivery and minimize readmissions to hospitals in clinical practice.

6. References

1. Akinbode, A. K., Olinmah, F. I., Chima, O. K., Okare, B. P., & Aduloju, T. D. (2025). Predictive Modelling For Hospital Readmission Using Socioeconomic And Clinical Data. *Engineering And Technology Journal*, 10(8), 6438-6465.
2. Brothers, T., Adhikari, K., Ramazani, M., Imtiaz, A., & Al-Mamun, M. (2026). Synthetic data-augmented machine learning for 30-day readmission prediction in patients with chronic conditions: a retrospective real-world study. *BMJ Open*, 16(4), e108273.

3. Digutla, R. K. (2025). India hospital readmission dataset (2015–2024) [Data set]. Kaggle. <https://www.kaggle.com/datasets/digutlaranjithkumar/india-hospital-readmission-dataset-20152024>
4. Grossman Liu, L., Rogers, J. R., Reeder, R., Walsh, C. G., Kansagara, D., Vawdrey, D. K., & Salmasian, H. (2021). Published models that predict hospital readmission: a critical appraisal. *BMJ open*, 11(8), e044964.
5. Jiang, X., & Wang, B. (2024). Enhancing Clinical Decision Making by Predicting Readmission Risk in Patients With Heart Failure Using Machine Learning: Predictive Model Development Study. *JMIR Medical Informatics*, 12(1), e58812.
6. Lazzarini, N., Filippoupolitis, A., Manzione, P., & Eleftherohorinou, H. (2022). A machine learning model on Real World Data for predicting progression to Acute Respiratory Distress Syndrome (ARDS) among COVID-19 patients. *PLoS One*, 17(7), e0271227.
7. Major, V. J., Jethani, N., & Aphinyanaphongs, Y. (2020). Estimating real-world performance of a predictive model: a case-study in predicting mortality. *JAMIA open*, 3(2), 243-251.
8. Mohanty, A. (2024). Dynamic Predictive Models for Hospital Readmission Risk. *International Journal on Computational Modelling Applications*, 1(2), 10-19.
9. Nasiruddin, M., Dutta, S., Sikder, R., Islam, M. R., Mukaddim, A. A., & Hider, M. A. (2024). Predicting heart failure survival with machine learning: assessing my risk. *Journal of Computer Science and Technology Studies*, 6(3), 10-32996.
10. Payrovnaziri, S. N., Chen, Z., Rengifo-Moreno, P., Miller, T., Bian, J., Chen, J. H., ... & He, Z. (2020). Explainable artificial intelligence models using real-world electronic health record data: a systematic scoping review. *Journal of the American Medical Informatics Association*, 27(7), 1173-1185.
11. Raj, S. (2025). Developing AI Models for Predicting Hospital Readmission Rates. *Journal Publication of International Research for Engineering and Management (JOIREM)*, 5(06).
12. Rodriguez, P. J., Veenstra, D. L., Heagerty, P. J., Goss, C. H., Ramos, K. J., & Bansal, A. (2022). A framework for using real-world data and health outcomes modeling to evaluate machine learning-based risk prediction models. *Value in Health*, 25(3), 350-358.
13. Shang, Y., Jiang, K., Wang, L., Zhang, Z., Zhou, S., Liu, Y., ... & Wu, H. (2021). The 30-days hospital readmission risk in diabetic patients: predictive modeling with machine learning classifiers. *BMC medical informatics and decision making*, 21(Suppl 2), 57.
14. Shao, W., Shu, L., Wang, X., Yu, F., Zhou, T., & Han, D. (2026). Predicting 30-day readmission using DRG-based hospitalization data: a large real-world logistic regression model from a tertiary hospital. *Frontiers in Public Health*, 14, 1776878.
15. Su, M. C., Wang, Y. J., Chen, T. J., Chiu, S. H., Chang, H. T., Huang, M. S., ... & Chen, Y. C. (2020). Assess the performance and cost-effectiveness of LACE and hospital re-admission prediction models as a risk management tool for home care patients: An evaluation study of a medical center affiliated home care unit in Taiwan. *International journal of environmental research and public health*, 17(3), 927.
16. Teo, K., Yong, C. W., Chuah, J. H., Hum, Y. C., Tee, Y. K., Xia, K., & Lai, K. W. (2023). Current trends in readmission prediction: an overview of approaches. *Arabian journal for science and engineering*, 48(8), 11117-11134.
17. Tripathi, M., & Rastogi, D. D. (2025). Patient Readmission Risk Prediction: Machine Learning Classification Algorithms.
18. Wang, H. E., Weiner, J. P., Saria, S., Lehmann, H., & Kharrazi, H. (2024). Assessing racial bias in healthcare predictive models: Practical lessons from an empirical evaluation of 30-day hospital readmission models. *Journal of Biomedical Informatics*, 156, 104683.
19. Wang, S., & Zhu, X. (2021). Predictive modeling of hospital readmission: challenges and solutions. *IEEE/ACM transactions on computational biology and bioinformatics*, 19(5), 2975-2995.
20. Wideqvist, M., Cui, X., Magnusson, C., Schaufelberger, M., & Fu, M. (2021). Hospital readmissions of patients with heart failure from real world: timing and associated risk factors. *ESC heart failure*, 8(2), 1388-1397.
21. Yu, M., Harrison, M., & Bansack, N. (2024). Can prediction models for hospital readmission be improved by incorporating patient-reported outcome measures? A systematic review and narrative synthesis. *Quality of Life Research*, 33(7), 1767-1779.
22. Zhao, P., Yoo, I., & Naqvi, S. H. (2021). Early prediction of unplanned 30-day hospital readmission: model development and retrospective data analysis. *JMIR medical informatics*, 9(3), e16306.