World Journal of Advanced Research and Reviews

eISSN: 2581-9615 CODEN (USA): WJARAI

Cross Ref DOI: 10.30574/wjarr

Journal homepage: https://wjarr.com/

WJARR World Journal of Advanced Research and Reviews

(RESEARCH ARTICLE)

Check for updates

Machine Learning Approaches for Predicting 30-Day Hospital Readmissions: Evidence from Massachusetts Healthcare Data

Awele Okolie [1,*], Dumebi Okolie [2], Callistus Obunadike [3], Darlington Ekweli [4], Pearlrose Nwuke [5], Bello Abdul-Waliyyu [3] and Paschal Alumona [6]

[1] *School of Computing and Data Science, Wentworth Institute of Technology, USA.*
[2] *Department of Finance and Economics, Faculty of Business and Law, Manchester Metropolitan University, UK.*
[3] *Department of Computer Science and Quantitative Methods, Austin Peay State University, Tennessee, USA.*
[4] *Department of Health Care Administration, University of the Potomac, Washington, DC, USA.*
[5] *Department of Computer Science, Metropolitan College, Boston University, USA.*
[6] *Booth School of Business, University of Chicago, USA.*

World Journal of Advanced Research and Reviews, 2025, 28(01), 914-929

Publication history: Received on 31 August 2025; revised on 11 October 2025; accepted on 13 October 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.28.1.3457

Abstract

Thirty-day hospital readmissions represent a critical challenge in healthcare, contributing to significant financial burdens, increased patient morbidity, and reflecting gaps in care continuity. This study aimed to develop and evaluate machine learning models for predicting 30-day hospital readmissions using a comprehensive, statewide healthcare dataset from Massachusetts. Employing a quantitative, predictive modeling design, this research compared the performance of Ridge regression with two advanced ensemble methods: Random Forest and Gradient Boosting. The models were trained and tested on a hospital-year panel dataset derived from the Massachusetts readmissions data book. Performance was evaluated using Root Mean Squared Error (RMSE) and the coefficient of determination (R^2). The results demonstrated the superior predictive power of the ensemble methods over the traditional linear model. Gradient Boosting emerged as the top-performing model, achieving the lowest RMSE of 1.48 and the highest R^2 of 0.81, followed closely by Random Forest (RMSE = 1.52, R^2 = 0.80). In contrast, Ridge regression showed limited predictive capability (RMSE = 2.54, R^2 = 0.43). Feature importance analysis from the Gradient Boosting model identified the number of deaths/readmissions and the number of cases as the most influential predictors, with hospital quality ratings and geographic factors also contributing significantly. The findings indicate that machine learning, particularly Gradient Boosting, provides a robust and accurate tool for identifying patients at high risk of readmission. Implementing such models can enable healthcare systems to better allocate resources, tailor discharge planning, and ultimately improve patient outcomes by reducing costly and disruptive readmissions.

Keywords: Machine Learning; Predictive Modeling; Gradient Boosting; Random Forest; Healthcare Analytics; Risk Stratification; Data-Driven Healthcare; Transitional Care

1. Introduction

A premature hospital readmission after discharge represents a serious adverse event in a patient's care. It represents a severe disruption during recovery, strongly associated with unnecessary patient morbidity, functional deterioration, and far-reaching psychological distress for patients and families (Halac et al., 2025). Unplanned and premature hospital readmissions are considered a major patient safety problem and a strong predictor of care fragmentation (Yang et al., 2025). This unplanned and premature rehospitalization suggests that a break has occurred within the care continuum,

* Corresponding author: Okolie Awele

casting doubt on the assumption that the patient has been effectively transitioned from the acute care setting back into their community (Sherman et al., 2022).

Due to their frequency and expense, readmission rates are a universal metric for hospital quality and continuity of care. Publicly reporting rates of readmission for numerous diseases, including acute myocardial infarction, heart failure, and pneumonia, with historic rates of 20-25%, is where the Centers for Medicare & Medicaid Services (CMS) began (Davis et al., 2022). A high level of readmission will most likely reflect more systemic issues related to discharge practices, such as medication reconciliation problems or a lack of coordination with post-acute providers (Michailidis et al., 2022). Readmission is a useful, data-driven metric that patients, regulators, and payers can use to compare and evaluate the performance of healthcare organizations.

The financial burden on hospitals and payers resulting from hospital readmissions is high and a significant area of focus for wasteful spending in the U.S. healthcare system. The total annual estimated amount of preventable rehospitalizations, as reported by payers, is approximately $41.3 billion in the U.S. (Lv et al., 2023). The Medicare program, which covers a high-risk population with a higher likelihood of readmission, has a disproportionate cost burden, and exploratory research has estimated the annual costs to Medicare at $17.4 billion (Subasi, 2024). A majority of these costs can be avoided with more coordinated care. This enormous, largely preventable waste is a significant burden on an already overtaxed system, and it's an even greater economic imperative for politicians and hospitals to develop preventive, evidence-based solutions that optimize hospital and patient outcomes in the long run (Artetxe et al., 2018).

To address the rising expenses and declining quality, U.S. policymakers implemented a landmark intervention. The 2010 Affordable Care Act established the Hospital Readmissions Reduction Program (HRRP), which significantly altered the economic incentives for healthcare providers (Michailidis et al., 2022). Penalties from the Centers for Medicare & Medicaid Services (CMS) are imposed on hospitals with an increased 30-day readmission rate of clinically related target conditions, such as pneumonia and heart failure (Almeida et al., 2025). These penalties are not trivial, with hospitals being penalized to the tune of up to 3% of their overall base Medicare payments. This sudden and acute financial impact began to transform the readmission issue from a reactive quality measure into a high-stakes, operational, and strategic priority among hospital administrators across the U.S., resulting in significant investments in care transitions efforts and predictive analytics to prevent financial peril (Zhang et al., 2024).

The biggest challenge to decreasing readmission rates is the ability to identify patients at risk promptly and appropriately when they are being discharged from the hospital (Halac et al., 2025). Readmission is a multifaceted outcome, resulting from a combination of clinical comorbidities, social determinants of health, functional status, and the strength of the support system after discharge, as asserted by Gao et al. (2023). Identifying patients by primary diagnosis has failed. Because intensive transitional interventions are expensive, it is not feasible to implement them in large quantities. Patient mapping to identify the type of population with a high probability of readmission is a crucial element of any reliable and replicable solution to reduce readmissions (Artetxe et al., 2018). Selecting the population to risk-stratify remains the first operational hurdle for organizations willing to improve transitions of care and qualify for exemption from penalties under the Hospital Readmissions Reduction Program (HRRP) (Sherman et al., 2022).

Historically, responses to this predictive issue have relied on parsimonious, rule-based rating systems, such as the LACE index, which utilizes several readily available variables, including Length of Stay and patient comorbidities (Huang et al., 2022). While groundbreaking, such early models have only limited predictive power because they were created in a time when data was not freely available. The environment has undergone significant changes since then, primarily due to the widespread adoption of Electronic Health Records (EHRs), which has ushered in an era of unprecedented data richness (Davis et al., 2022). Hospitals now possess vast longitudinal data sets of thousands of potential predictors per patient, ranging from fine-grained laboratory results and medication lists to unstructured physician notes. This explosion of big data has rendered the naive assumptions of earlier models inadequate, while also presenting a tempting challenge to apply more sophisticated analysis techniques to harness the complexity and achieve a higher degree of predictive accuracy (Subasi, 2024).

The confluence of widespread data availability and the limitations of traditional methods requires a more powerful analytical framework. This study, therefore, utilizes machine learning to address this critical prediction task. The primary objective is to develop, train, and compare a collection of advanced machine learning models based on a rich, statewide Massachusetts healthcare dataset. By determining the best-performing predictive model, this paper aims to provide a more precise and reliable tool for stratifying patient risk, enabling healthcare systems to target resources better, improve transitional care quality, and ultimately reduce the incidence of costly 30-day hospital readmissions.

2. Literature Reviews

2.1. Conceptualizing Hospital Readmission

For over 30 years, hospital readmissions following discharge have been a key topic in health services research and policy. Artetxe et al. (2018) highlighted the clinical and economic burden of unwanted and unplanned readmissions, particularly for cardiovascular patients. Jencks, Alnazari et al. (2025) reported that nearly 20% of Medicare patients were rehospitalized within 30 days of their initial discharge at a cost to the U.S. healthcare system of over $17 billion per year. They presented the burden both in terms of dollars and confirmed readmission rates as an acceptable surrogate for quality of care.

Clinical trials have consistently demonstrated that comorbidities (e.g., heart failure, chronic obstructive pulmonary disease, diabetes) are robust predictors of subsequent early rehospitalization (Lv et al., 2023; Michailidis et al., 2022). Others have emphasized the importance of demographic and socioeconomic determinants. Pons-Suner et al. (2023) demonstrated that hospitals caring for disadvantaged populations have high rehospitalization rates, which raises doubts about whether these rates accurately reflect provider quality. This issue has also prompted further academic discussion regarding the equity and usefulness of readmission rates as accountability measures overall (Ashfaq et al., 2019).

At the policy level, one of the keys turning points was the implementation of the Hospital Readmissions Reduction Program (HRRP) by the Centers for Medicare & Medicaid Services in 2012. Baselines for penalties based on discharge rates for readmissions of pneumonia and acute myocardial infarction were established by HRRP, which institutionalized readmissions as a specific quality-improvement goal (Shang et al., 2021). Some evidence suggests that readmission rates have declined but improved slightly since the implementation of HRRP (Afrash et al., 2022). However, Zhang, et al. (2024) discussed possible unintended consequences of the policy, including detrimental effects on mortality and disparities. These readmission issues, however, are not exclusive to the U.S. Studies in the U.K., Canada, and the Nordic countries also document similar issues, albeit with institution-specific arrangements and varying mixes of patients determining outcomes (Gao et al., 2023; Huang et al., 2022). This project highlights that out-of-hospital readmission is not the sole issue in the U.S., but rather an ongoing and complex issue for the health system.

2.2. Determinants of Hospital Readmission

Determinants of hospital readmission have been explored extensively within clinical, demographic, and systems domains, revealing the multifactorial nature of the problem. Tabak et al. (2016) demonstrated that clinical factors at the patient level, which account for readmission risk, offer an important but incomplete explanation for variability. This was a significant indication that a more comprehensive conceptual framework is required. Henderson et al. (2023) and Gandra (2024) built on this observation, highlighting the interaction between comorbidities, socio-demographics, and levels of healthcare system dynamics.

Clinical diagnoses are always the strongest proximal predictors of hospital readmission. Comorbidities, such as heart failure, diabetes, chronic obstructive pulmonary disease (COPD), and kidney disease, have been shown to increase the risk of readmission within 30 days (Morel, et al., 2020; Xiong et al., 2022). Liu et al. (2024) found that heart failure patients had among the highest risks for readmission. Among the risk factors for readmission are the severity of illness, the length of the index hospital stays, and the complexity of the discharge diagnoses (Davis et al., 2022). Medication management, specifically the issue of polypharmacy, is also important, and Shang et al. (2021) found that a lack of reconciliation of discharge medications was one of the most significant preventable causes of readmissions.

In addition to clinical comorbidity, demographic and socioeconomic factors are also important. Wang & Zhu (2022) found that readmission was associated with older patient age, male gender, and reduced socioeconomic status, after adjusting for clinical comorbidity. Insurance status and income are crucial, as readmission rates are higher among patients with Medicaid or no insurance (Teo et al., 2020). Geographic disparities also persist; hospitals that treat patients from rural or disadvantaged socioeconomic backgrounds generally have higher readmission rates, highlighting the more prevalent disparities in post-acute and outpatient access for these populations (Tavakolian et al., 2023). This adds to the argument that readmissions are not merely clinical problems, but a reflection of social determinants of health.

An additional characteristic of direct role determinants includes system and provider factors. Transitions in care, such as inadequate discharge planning, insufficient timely follow-up, and poor coordination of care between hospital and community providers, are strongly associated with a high risk of readmission (Almeida et al., 2025; Gao et al., 2023).

Michailidis et al. (2022) found that patients who did not have a follow-up visit within seven days of discharge were more likely to be readmitted. Additionally, inadequate communication between the hospital team and the primary care physician leads to a lack of continuity of care (Zhang et al., 2024). Staffing ratios, the availability of hospital resources, and institutionalized quality measures are also factors that influence direct role determinants, with "higher-performing hospitals" experiencing lower healthcare utilization rates (Lv et al., 2023).

A more recent wave of studies examines community-level and environmental risk factors as part of the overall risk assessment. Neighborhood disadvantage, social support, and relative scarcity of post-acute care services, including skilled nursing or rehabilitation facilities, have each been predictive of higher readmission rates (Artetxe et al., 2018; Pons-Suner et al., 2023). These results reflect the growing recognition that structural health determinants extend beyond hospitals and are integral to population health outcomes. However, despite these likely findings, researchers suggest that these factors alone are insufficient to predict readmission risk. Subasi (2024) found that most predictive models based on only a limited number of variables (e.g., demographics and comorbidities) were poorly performing, with C-statistics generally below 0.70. This implies the determinants of readmission are multifactorial, and predictive analytics is beset with challenges. Due to this, more recent studies recommend incorporating stronger datasets, including clinical, demographic, psychosocial, and contextual data, into predictive analytic models.

2.3. The Evolution of Readmission Risk Prediction

The first generation of predictive risk models for readmission was driven by a parsimony paradigm based on the realities of the pre-digital health era. The goal was to create easy-to-calculable scores from sparse, representative variables in accounts and standard clinical databases. An example is the LACE index (Ryu et al., 2021) and hospital score (Hogan et al., 2019), both of which were developed based on straightforward statistical methodologies that generally relied on logistic regression, which assumes linear relationships between predictors and outcomes.

In general, such designs were functional because their simplicity communicated significance, enabled easy and inexpensive application by the clinician at the bedside, and could be ascertained within a short but significant encounter (Gandra, 2024). Nevertheless, it did stand to reason that the potential to produce meaningful, predictive metrics of readmission was constrained from the beginning, as the prescriptive nature of any scoring tool could never hope to account for the complexity of myriad non-linear interactions that occur between myriad disparate factors across the post-discharge continuum (Ashfaq et al., 2019). These risk models are viewed as progressive, albeit basic, efforts to substantively link risk, or compromised risk in this instance, into a codified risk knowledge framework founded on limited yet efficacious indicators.

The advent of Electronic Health Records (EHRs) has ushered in a drastic transformation to the predictive era, from a previous era of data scarcity to an era of data abundance (Huang et al., 2022). This metamorphosis rendered the underpinning upon which early models were established less precise than it had been. Researchers now have access not just to a dozen administrative variables, but rather to huge, high-dimensional data sets, with thousands of features from which to select, including detailed longitudinal data from laboratory tests, vital signs, and medication administration data (Morel et al., 2020).

One of the main innovations was the ability to access unstructured data, i.e., discharge summaries and clinical notes from NLP, which we also happen to know are "baked," so to speak, into EHRs that record that domain and content (Davis et al., 2022). It is worth repeating that the richness of the data brought one challenge for analytics: the richness and number of variables would frequently be incompatible with the usual regression analyses. All this led to a drift toward advanced data-driven approaches: an opportunity to apply machine learning algorithms to uncover hidden patterns in such complexities (Ryu et al., 2021).

2.4. Traditional Approaches to Predicting Hospital Readmissions

Historically, logistic regression has been most commonly employed because of its ability to handle the analysis of binary responses, such as whether a patient is readmitted within 30 days. Hogan et al. (2019) and Shang et al. (2021) have, in their previous work, demonstrated that logistic regression can be used to combine demographic and clinical predictors to provide estimates of the probabilities of readmission. Cox proportional hazard models have also been applied in cases when time-to-readmission has been expressed as a survival response (Wang & Zhu, 2022). Statistical robustness and interpretability are the strengths associated with logistic regression and Cox models; however, the linear assumptions employed cannot explain the complexity or the non-linear relationships between predictors that might be present (Hosseinzadeh et al., 2013).

Risk assessment tools for clinicians have been developed and validated to enable the quick assessment of readmission risk. Van Walraven and his colleagues developed the LACE index, which remains a popular measure. It uses four factors namely length of stay, Acuity of admissions, Comorbidities, and Emergency department visits to predict early readmission and death (Linzey et al., 2020). Likewise, the Charlson Comorbidity Index has been commonly applied in estimating the comorbidity burden for a patient to predict poor outcomes (Gao et al., 2023). The newer hospital score was also cross validated in various cohorts worldwide and proved scalable in discriminating between low, intermediate, and high-risk patients for readmission (Halac et al., 2025). These scoring mechanisms are appealing because they are parsimonious, clinically understandable, and enable practitioners to conduct risk assessments at the bedside.

Henderson et al. (2023) reported that most readmission prediction models, both regression and simple risk scores, had C-statistics of 0.55-0.65, corresponding to a poor to fair level of discriminative ability. Classical models are limited by their reliance on only a small subset of predictors, their inability to process high-dimensional clinical data, and the dynamics of interaction among predictor variables. Additionally, classical models lack cross-setting generalizability, as evidenced by varying performance depending on either the hospital or patient population of the model (Artetxe et al., 2018). Traditional methods helped shape current predictive practice and established essential methodological and conceptual benchmarks. Their transparency and interpretability, especially in a clinical setting, will always be valuable where accountability is a key concern in decision-making.

2.5. Application of Machine Learning Approaches

What drove the trend towards ML is the repeated failures of classical models, not new-algorithm fever. Gandra (2024) reviewed systematically regression-based models and concluded that the majority performed at best moderate discrimination. This presented an empirical challenge for computationally demanding models that can manage nonspecific coefficients of non-linear, high-dimensional data. Tree-based techniques were among the first lines of investigation. Afrash et al. (2022) demonstrated that decision trees were on par with, and in some cases identified as superior to, logistic regression when classifying the risk of readmission in heart failure. In repeat investigations, Rajaguru et al. (2022) extended methodology by improving random forests and support vector machines to develop improvements in the accuracy of prediction over related risk scores. These congruent, but not identical, findings suggested that there may be a conceptually significant trend: ML methods tended to boost meaningful predictive accuracy, with gains in that area being modest unless additional datasets were acquired.

The inclusion of deep learning has also turned the equation around. Liu et al. (2020) applied millions of EHR records to train DNNs and demonstrated evidence of improved performance over simpler standard approaches, albeit at a transparency cost. Notably, the predictive performance was further enhanced with the addition of unstructured clinical notes, via natural language processing, by Rumshisky et al. (2016) and Miotto et al. (2016), suggesting that predictive hints were buried in narrative reporting outside of structured fields only. The following explanation of the literature above is not fully compatible here.

Linzey et al. (2020) caution against overhyping ML applications, providing examples of problems in ML studies, including overfitting, external validity issues, and the lack of transparency in "black-boxed" patient decisions. Attention to any of the above-listed problems has led to the construction of initiatives around explainability frameworks, which include SHAP values, with LIME being the best known (Lundberg & Lee, 2017). There remain uncertainties about whether interpretability is possible as the complexity of algorithms grows.

2.6. Theoretical Review

2.6.1. Behavioral Model of Health Services Use

This theory was proposed by Andersen in 1995 and has been extensively used in healthcare investigations to provide insights into the underlying motives behind people's access to and use of healthcare services. The model suggests that the consumption of health services, such as making a surprise return to the hospital, results from three dynamic and interactive groups of factors: Predisposing Factors, Enabling Resources, and Need Factors (Alkhawaldeh, et al., 2023). This conceptual model will enable us to develop a more systematic and integrative method of classifying and conceptualizing readmission determinants, considering not only a list of risk factors but also the broader patient context (Radhamony et al., 2024).

The term "predisposing factors" refers to existing factors that influence an individual's tendency or predisposition to use a service. Predisposing factors include demographic factors (gender, age, and race/ethnicity), social structure (education, occupation, and marital status), and health beliefs (Alkhawaldeh, et al., 2023). Health beliefs influence a person's perception of health and their interaction with healthcare. Enabling resources are the material conditions that enable individuals to access the healthcare system (Lederle et al., 2021). Enabling resources consist of what is implied in terms of patient condition to comply with the post-discharge care plan, i.e., personal and family enabling resources (social and health support, income) or community enabling resources (transportation, number of primary care physicians available). Need factors are the most direct determinants of service use, translating the patient's professionally rated and perceived health. Need factors include the severity of the index admission, the burden of comorbidities, functional status, and impairments in cognitive or mental health (Krzyz et al., 2023).

2.6.2. Transitional Care Model

This is a prominent systems-level strategy developed and firmly established by Naylor and colleagues. It is an evidence-based, nurse-implemented strategy for preventing avoidable readmissions of high-risk patients from the hospital into home. The theoretical basis of the TCM is that readmissions are often preventable iatrogenic complications resulting from fragmentation in the healthcare system. It targets the 30th day after discharge as a time of significant vulnerability in which poor communication, inadequate patient preparation, and uncoordinated follow-up can lead to adverse events. The model's key intervention is a Transitional Care Nurse specifically assigned to follow the patient from the inpatient setting to home, providing coordinated services such as in-hospital planning, intensive patient and caregiver education, home visits, medication self-management therapy, and communication with primary and specialty care providers. The TCM's focus on bridging this system gap highlights the need to identify patients who would benefit most from such resource-draining support.

3. Methodology

3.1. Research Design

This study employs a quantitative research design with a focus on predictive modeling using machine learning approaches. The design emphasizes the systematic application of computational techniques to healthcare data, enabling the generation of reliable predictions. A comparative framework is adopted, where traditional statistical methods are benchmarked against advanced machine learning algorithms to assess performance and applicability. The choice of a predictive, quantitative design is motivated by its ability to handle large datasets, uncover hidden patterns, and enhance decision-making in healthcare. This approach provides a structured pathway to evaluate methodological rigor while ensuring replicability.

3.2. Data Source and Description

The primary data source is the Massachusetts readmissions data book covering State Fiscal Years 2011–2023. The file is organized as a multi-sheet report rather than a single flat table; it contains statewide yearly summaries, hospital-level cross-sections, hospital-by-year panels, and demographic breakdowns. Key items available across sheets include counts of discharges and readmissions, observed readmission rates, and risk-adjusted rates reported at both the hospital and aggregate levels. Hospital identifiers and names are listed in the hospital tables, allowing for the creation of a hospital–year panel.

Because the data book is aggregate, all analyses will operate at the hospital level. Typical data quality issues present in the raw file include nonstandard header rows, mixed formats across sheets, occasional missing cells, and minor name inconsistencies that require normalization before merging. To strengthen analysis, the data book can be augmented with external county-level covariates if needed. The final dataset for modelling will be a cleaned and harmonized hospital-year table derived from the relevant sheets of the data book.

3.3. Data Preprocessing and Feature Engineering

Data preprocessing is a critical step to ensure consistency, accuracy, and usability of the dataset. The raw Massachusetts readmission data book contains multiple sheets with varying formats; therefore, the first step involves extracting relevant tables and harmonizing them into a structured hospital-year dataset. Non-data rows, duplicate entries, and irregular text formats are removed to achieve uniformity. Missing values are addressed through appropriate strategies such as imputation, exclusion, or aggregation, depending on the extent and importance of the gaps.

Categorical variables, such as hospital names and locations, are encoded into numerical formats suitable for machine learning algorithms. Continuous features, including counts and rates, are standardized to reduce scale disparities. Feature engineering is applied to create new predictors such as ratios, trends across years, or grouped regional indicators. Geographic coordinates are retained to capture spatial influences. The final preprocessed dataset is split into training and testing subsets, ensuring that evaluation reflects out-of-sample predictive performance. This stage provides a clean, representative, and analytically ready dataset for subsequent modeling and analysis.

3.4. Model Development

Model development focuses on constructing, training, and comparing predictive algorithms to estimate hospital readmission outcomes. The process begins by selecting suitable machine learning models that strike a balance between interpretability and predictive strength. Baseline models, such as linear regression, are included to provide a point of reference. More advanced algorithms, such as random forests, gradient boosting machines, and support vector regression, are employed to capture non-linear relationships and complex interactions within the data.

Each model is trained on the prepared dataset using systematic parameter tuning. Hyperparameter optimization, achieved through methods such as grid search or cross-validation, ensures that models achieve optimal performance while avoiding overfitting. Regularization techniques may also be applied as needed to control complexity and enhance generalizability. To maintain rigor, the development phase follows a structured pipeline, comprising data partitioning, model training, validation, and iterative refinement. Models are implemented using standard machine learning libraries, ensuring reproducibility. The outcome of this stage is a suite of candidate models ready for formal evaluation and performance benchmarking.

3.5. Model Evaluation and Validation

Model evaluation and validation are crucial for determining the reliability and generalizability of predictive outcomes. Performance is assessed using established regression metrics, including Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2). These measures provide insight into both the accuracy of predictions and the proportion of variance that the models explain. To guard against overfitting, k-fold cross-validation is employed, ensuring that performance is not dependent on a single train-test split. Comparisons are made across models to identify which algorithm best balances predictive accuracy and interpretability. Residual analysis is also conducted to examine systematic biases or deviations. Validation further includes testing models on a withheld dataset to simulate real-world performance. Emphasis is placed on transparency by reporting both strengths and limitations of the chosen models. This rigorous evaluation process ensures that findings are credible, reproducible, and helpful in informing healthcare decision-making.

3.6. Ethical Considerations

Ethical integrity is central to research involving healthcare data. Although the dataset used in this study is aggregated and publicly available, precautions are taken to ensure that no identifiable patient information is disclosed. The use of hospital-level summaries helps mitigate the risk of breaching patient confidentiality, while care is exercised in reporting results to ensure that individual institutions are not unfairly stigmatized. Findings are presented in a balanced manner, avoiding overgeneralization or misuse that could misinform stakeholders. Additionally, transparency is maintained by documenting all preprocessing, modeling, and evaluation steps to allow reproducibility. The study acknowledges the potential biases inherent in administrative data, such as uneven reporting standards across hospitals, and addresses these through methodological rigor. By adhering to principles of fairness, transparency, and respect for stakeholders, the research upholds ethical standards consistent with responsible data science practice.

4. Results

Table 1 Summary Statistics of Numerical Variables

Variable	Mean	Median	Std. Dev.	Min	Max
Risk Adjusted Rate	12.102	12.120	3.390	0.000	26.140
Number of Deaths/Readmissions	84.538	35.000	131.053	0.000	879.000
Number of Cases	187.590	132.000	172.349	1.000	821.000

Table 1 presents the summary statistics of the numerical variables used in the analysis. The risk-adjusted rate had a mean of 12.10 (SD = 3.39), with values ranging from 0.00 to 26.14, and a median of 12.12, suggesting that the distribution is fairly centered around its mean. The number of deaths/readmissions showed a mean of 84.54 (SD = 131.05), a wide range between 0.00 and 879.00, and a median of 35.00, indicating a positively skewed distribution driven by a few high values. Similarly, the number of cases had a mean of 187.59 (SD = 172.35), with values ranging from 1.00 to 821.00, and a median of 132.00, indicating variability in patient volume across hospitals. These results suggest considerable dispersion in outcomes and case volumes across the sample.

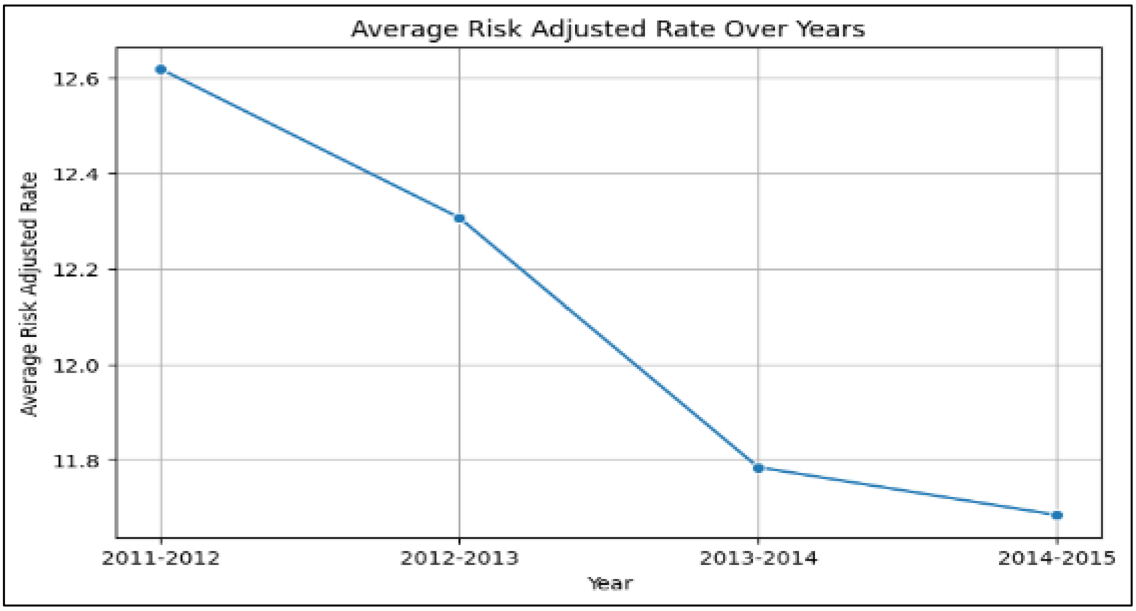

Figure 1 Trends analysis of Risk Adjusted Rate of Readmission from 2011 to 2015

Figure 1 illustrates the trend in average risk-adjusted readmission rates over the four-year period from 2011–2012 to 2014–2015. The results indicate a consistent decline over time, with the average rate decreasing from 12.61 in 2011–2012 to 12.30 in 2012–2013, and further dropping to 11.79 in 2013–2014 before reaching 11.69 in 2014–2015. This downward trend suggests that hospitals achieved gradual improvements in managing readmissions during the study period, reflecting potential gains in healthcare quality and effectiveness of interventions targeted at reducing hospital readmission rates.

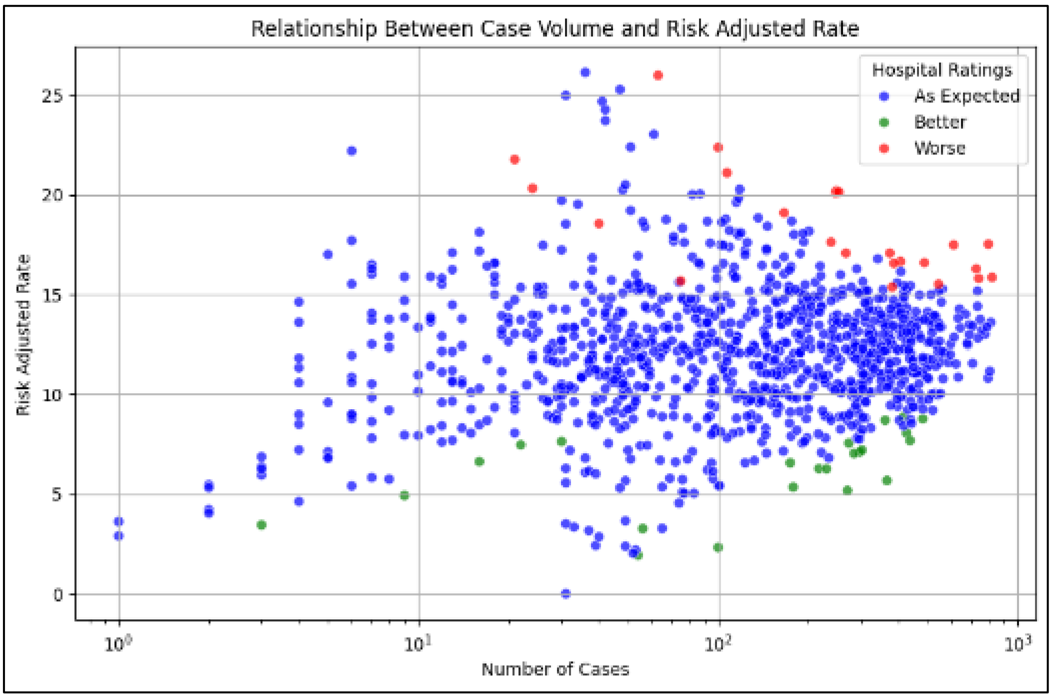

Figure 2 Relationship between Case Volume and Risk Adjusted Rate

Figure 2 illustrates the relationship between case volume and risk-adjusted readmission rates across hospitals, categorized by hospital rating. The scatter plot indicates that hospitals with higher patient volumes tend to cluster around moderate risk-adjusted rates, suggesting a degree of consistency in performance as volume increases. Conversely, hospitals with lower case volumes exhibit greater variability in outcomes, with both very low and very high readmission rates observed. Most hospitals were rated "as expected," while those rated "better" generally demonstrated lower risk-adjusted rates across varying case volumes. In contrast, hospitals rated "worse" tended to display higher readmission rates, often concentrated among facilities with larger case volumes. This pattern highlights how hospital performance may be influenced not only by patient volume but also by quality of care.

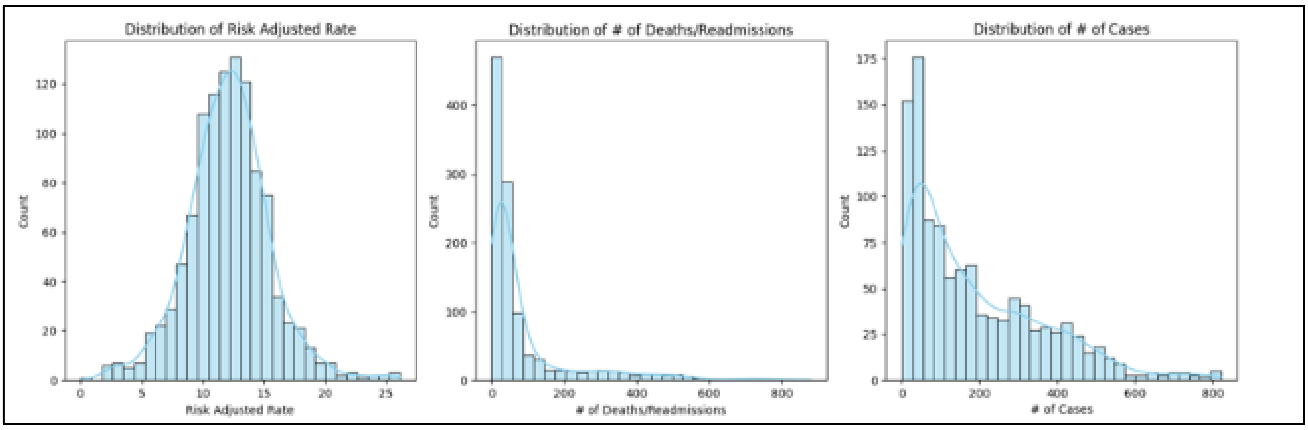

Figure 3 Distribution of the Features

The distributions of the factors associated with hospital readmission presented in Figure 3 revealed distinct patterns that highlight both concentration and variability across measures. The risk-adjusted rate was positively skewed, with most facilities clustered at lower values, suggesting that only a few exhibited unusually high rates. Similarly, the numbers of deaths and readmissions, as well as the number of cases, were heavily right-skewed, indicating that while most hospitals reported relatively few events, a small subset experienced disproportionately high counts.

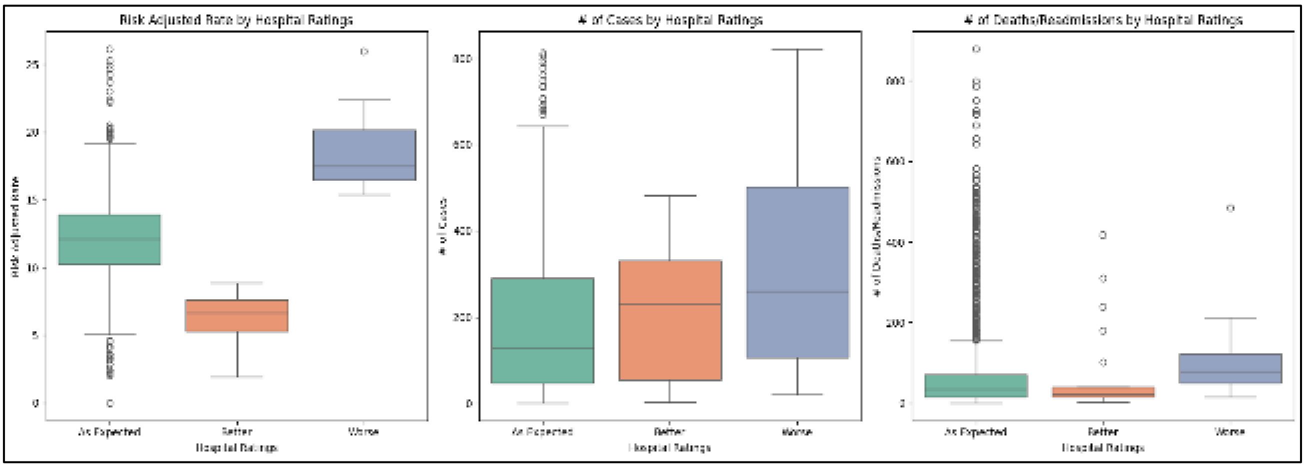

Figure 4 Hospital ratings by Risk Adjusted Rates, Number of Cases and Number OF Readmissions

Figure 4 illustrates the relationship between hospital ratings and three key indicators: risk-adjusted rates, number of cases, and number of deaths/readmissions. Hospitals rated as Worse consistently demonstrated the highest median values across all three measures, with wider variability and numerous outliers, suggesting poorer performance and greater inconsistency in outcomes. In contrast, hospitals rated as "Better" exhibited the lowest median risk-adjusted rates and fewer deaths/readmissions, indicating a stronger quality of care and more favorable patient outcomes. The As Expected category fell between these two extremes, reflecting moderate performance. This highlights a clear gradient in which higher-rated hospitals are associated with lower risk-adjusted rates, fewer cases, and reduced readmissions, consistent with expectations of quality-based rating systems

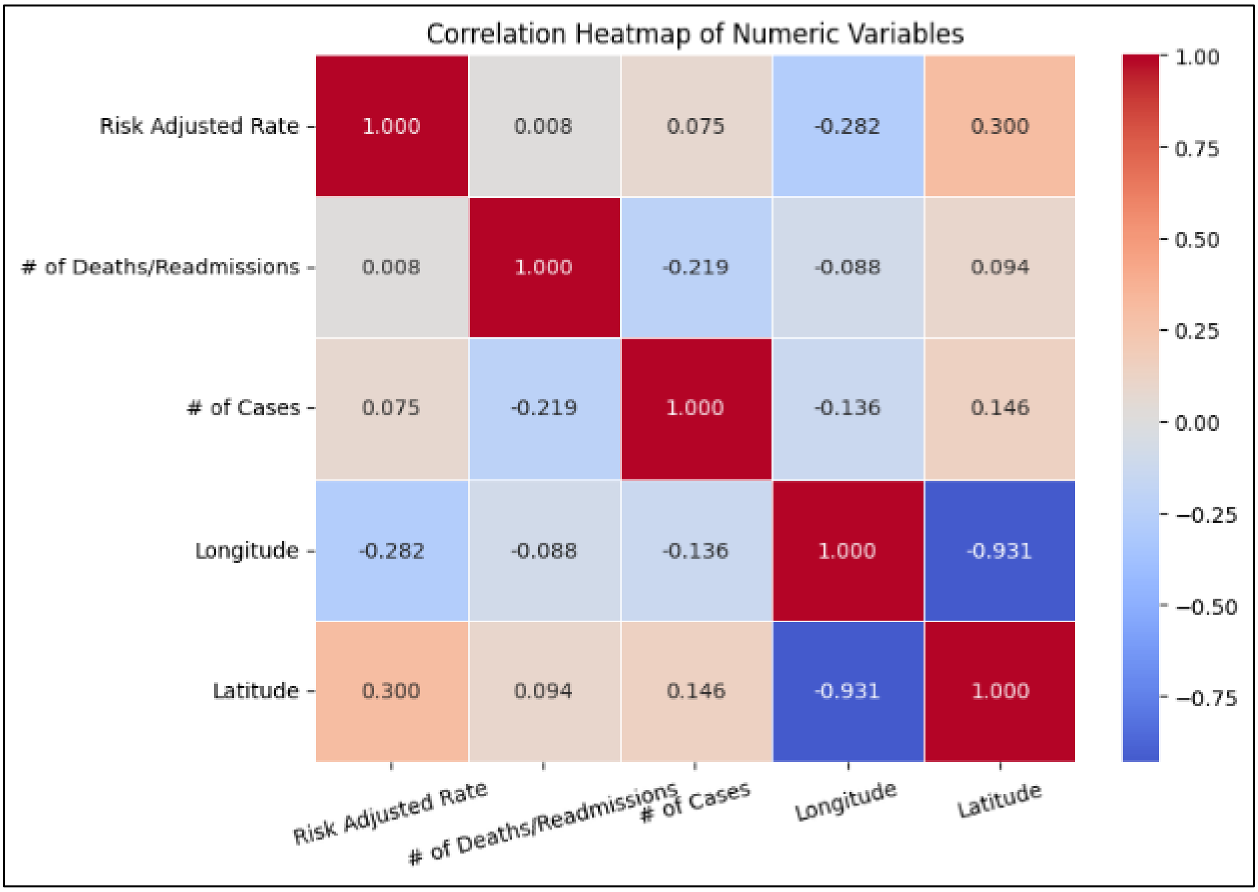

Figure 5 Heatmap for Correlation Analysis

The correlations analysis reported in Figure 5 showed several important patterns. The relationship between the risk-adjusted readmission rate and the number of deaths or readmissions was negligible (r = .01). At the same time, the association between the risk-adjusted readmission rate and the number of cases was also weak (r = .08). This suggests that neither hospital size nor raw readmission counts are strong predictors of adjusted outcomes. A small negative correlation was found between the number of cases and deaths/readmissions (r = –.22), indicating that larger hospitals do not necessarily experience proportionally higher readmission events. Geographic measures were highly correlated, with longitude and latitude demonstrating a very strong negative relationship (r = –.93), reflecting redundancy between the two coordinates. Latitude showed a modest positive association with the risk-adjusted readmission rate (r = .30), suggesting some degree of geographic variation in patient outcomes across hospitals.

4.1. Model Performance Evaluation

Table 2 Best Hyperparameters

Model	Key Hyperparameters
Ridge	α (alpha) = 1.0
Random Forest	max_depth = None min_samples_split = 2 n_estimators = 200
Gradient Boosting	learning_rate = 0.2 max_depth = 5 n_estimators = 200

The comparison of hyperparameters across the three machine learning models indicates that Ridge regression achieved its best performance with an alpha value of 1.0, reflecting a moderate level of regularization. In contrast, the Random Forest model performed optimally with 200 estimators, unrestricted tree depth, and a minimum of two samples required for splitting, suggesting that deeper and more complex trees enhanced predictive accuracy. Similarly, the Gradient Boosting model achieved its optimal configuration with 200 estimators, a maximum depth of five, and a relatively high learning rate of 0.2, underscoring the importance of balancing model complexity with learning speed.

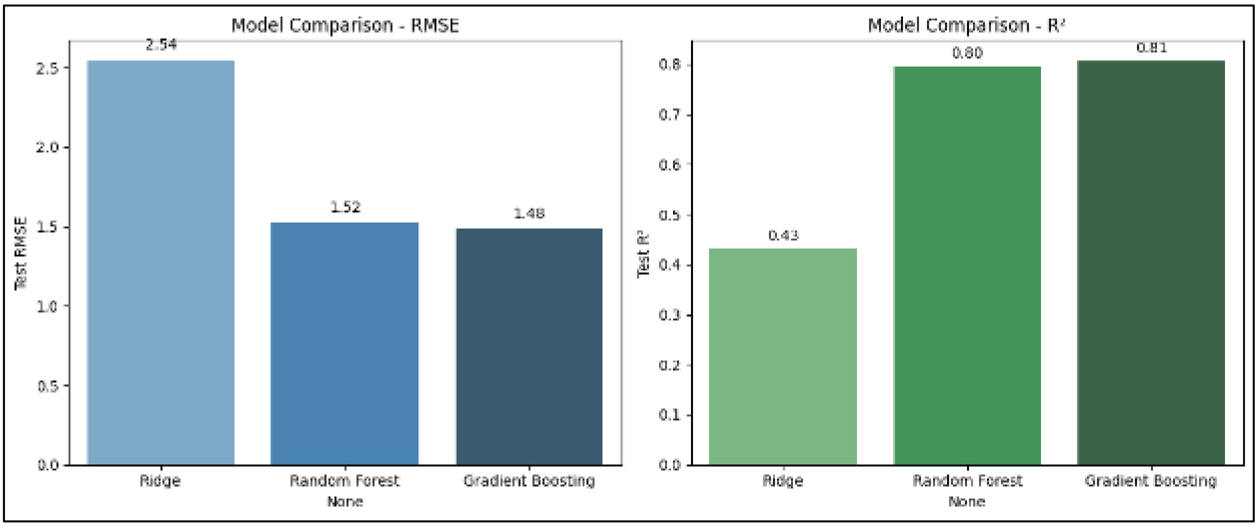

Figure 6 Model Performance Comparison

Figure 6 compares the predictive performance of Ridge, Random Forest, and Gradient Boosting models using test RMSE and R^2 values. The Ridge model performed weakest, with the highest RMSE (2.54) and the lowest R^2 (0.43), indicating limited explanatory power. In contrast, both ensemble methods substantially outperformed Ridge, with Random Forest achieving an RMSE of 1.52 and R^2 of 0.80, while Gradient Boosting slightly surpassed it with the lowest RMSE (1.48) and the highest R^2 (0.81). These results suggest that ensemble-based approaches captured the underlying data patterns more effectively than linear regularization, with Gradient Boosting providing the most accurate and reliable predictions overall.

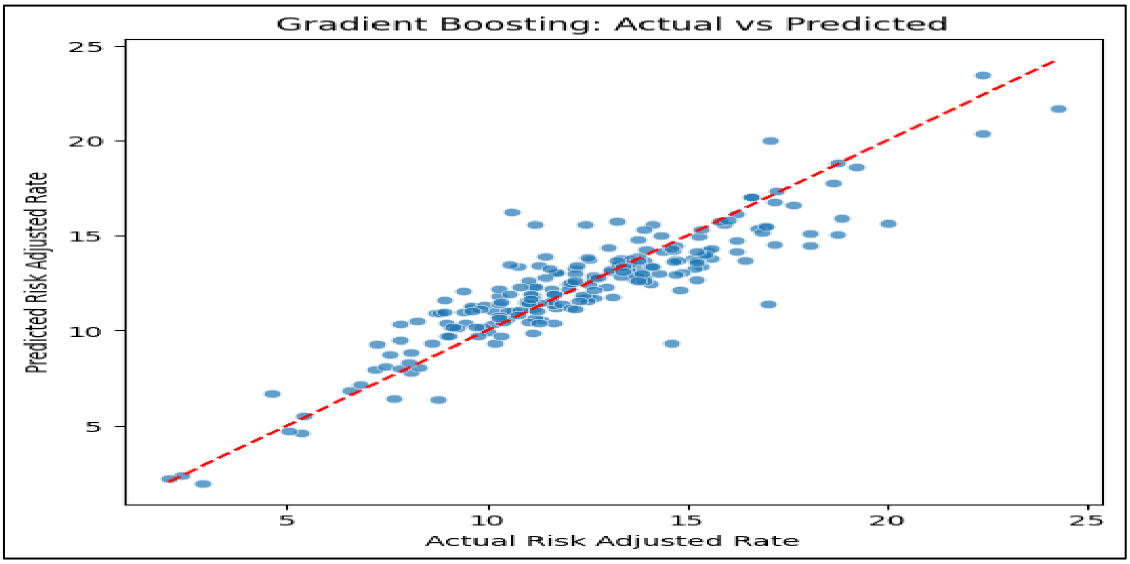

Figure 7 Hospital Readmission Risk Adjusted Rate: Actual Vs. Predicted

Figure 7 presents a scatter plot comparing actual and predicted hospital readmission risk-adjusted rates using the Gradient Boosting model. The diagonal reference line (y = x) represents perfect prediction, where the predicted values exactly match the observed outcomes. The majority of data points cluster closely around this line, indicating that the model achieved strong predictive accuracy, with relatively minor deviations between actual and predicted rates. However, some dispersion is evident, particularly at higher values, suggesting that while the model effectively captures overall patterns, prediction errors increase slightly at the extremes.

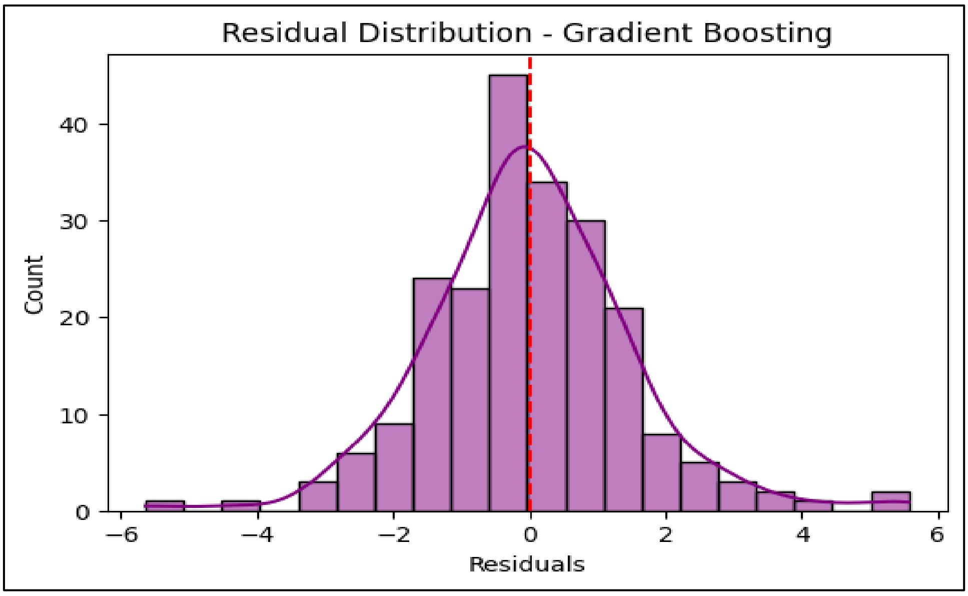

Figure 8 Residual Distribution

Figure 8 displays the residual distribution from the Gradient Boosting model, providing insight into the accuracy and bias of its predictions. The histogram shows that residuals are centered closely around zero, with a roughly symmetric, bell-shaped pattern, suggesting that the model's errors are generally small and evenly distributed. The vertical reference line at zero further highlights that most deviations are minor, with no substantial systematic over- or under-prediction. While a few residuals extend toward the tails, the overall distribution indicates that the Gradient Boosting model produces unbiased predictions with errors that approximate normality, supporting its reliability for modeling hospital readmission risk-adjusted rates.

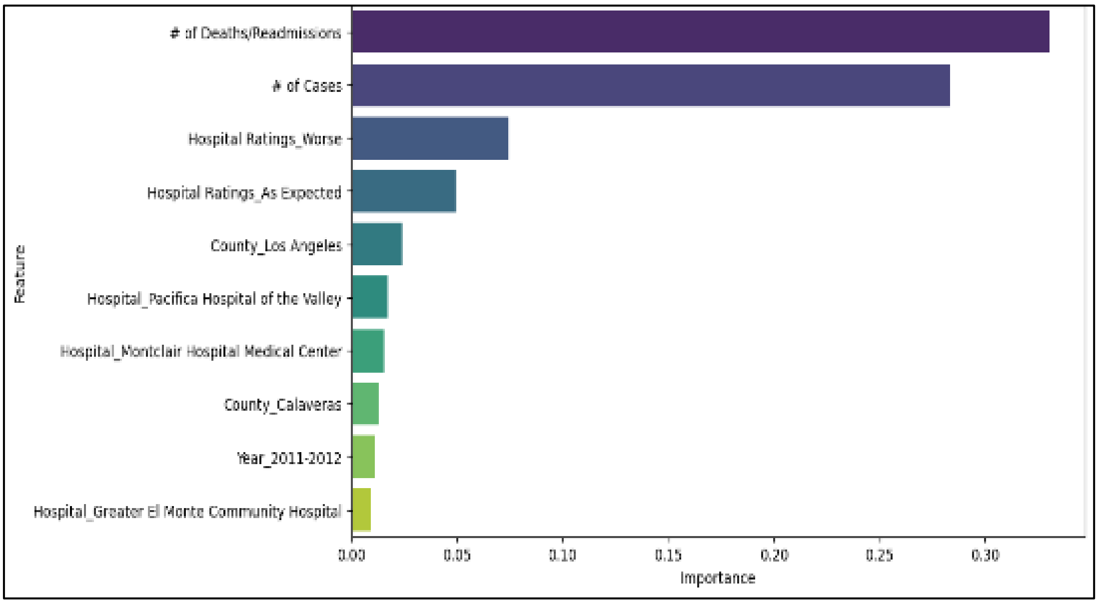

Figure 9 Top 10 Feature importances (Gradient Boosting)

Figure 9 presents the top ten feature importances derived from the Gradient Boosting model, highlighting the variables that most strongly influenced predictions of hospital readmission risk-adjusted rates. The number of deaths/readmissions emerged as the most influential predictor, followed closely by the number of cases, indicating that direct measures of patient outcomes and volume are central drivers of the model's performance. Hospital ratings also played a significant role, with "Worse" and "As Expected" categories contributing more to prediction accuracy than "Better" ratings, suggesting that lower-rated facilities carry stronger signals of risk. Geographic and institutional identifiers, such as county (e.g., Los Angeles, Calaveras) and specific hospitals, were among the top predictors, although with relatively lower importance, reflecting localized effects on outcomes. Temporal variation, represented by the 2011–2012 period, also contributed modestly to the overall trend.

5. Discussion of Findings

The results revealed that ensemble methods, specifically Random Forest and Gradient Boosting, outperformed Ridge regression in predicting the 30-day hospital readmission rate adjusted for risk. Gradient Boosting performed best, with the lowest RMSE (1.48) and highest R^2 (0.81), followed closely by Random Forest (RMSE = 1.52; R^2 = 0.80). In contrast, Ridge regression exhibited very poor predictive power and explanatory capability (RMSE = 2.54; R^2 = 0.43). These results are consistent with earlier empirical studies that have shown ensemble-based machine learning models to outperform linear models in recognizing advanced non-linear patterns in healthcare data. Huang et al. (2022) proved that XGBoost has improved predictive potential compared to logistic regression when predicting 30-day readmission rates for enrolled patients with pneumonia. Similarly, Halac et al. (2025) found that Random Forest improved the accuracy of prediction compared to logistic regression in an ambience conducted in a tertiary center.

Residual analysis revealed that the errors of Gradient Boosting were centered around zero and nearly normally distributed, indicating that predictions were unbiased and did not contain systematic error. This development is partly in line with a more recent study by Almeida et al. (2025), who demonstrated improved calibration and reduced bias in readmission prediction using both structured and unstructured electronic health record data in models employing more sophisticated machine learning techniques. The sole restriction is the presence of some remaining outliers, particularly in higher-risk-adjusted rates, which is a common finding in the literature on prediction models (e.g., restrictions with readmission and adverse event prediction). It is not particularly noted that the models are less accurate when implemented in extreme or high-acuity conditions. Feature importance analysis indicated that the most predictive variables were the number of cases and the number of deaths/readmissions, followed by hospital ratings and other contextual variables, such as county and facility identifiers. This finding is consistent with earlier studies, which have shown that prior utilization, comorbidity burden, and quality measures of institutions are among the best predictors of readmission risk (Huang et al., 2022; Halac et al., 2025). The findings align with the previously discussed construct of readmission risk as multifactorial, in that patient-level outcomes have the most significant predictive power. At the same time, institutional and geographical factors contribute incremental explanatory variance to the models.

6. Conclusions

The results show that deep learning techniques, particularly ensemble techniques such as Gradient Boosting and Random Forest, are an exceedingly accurate means of predicting 30-day hospital readmissions, surpassing traditional linear models. The capacity of the models to identify non-linear patterns in the complex healthcare data has important implications for hospital performance monitoring and patient care planning. The models provide established predictors that are previously set, including, but not limited to, a history of deaths/readmissions, volume of cases, and hospital quality scores, which indicate possible interventions that clinicians can enact to activate patients with a greater risk of avoidable readmission. The models can enable hospitals to allocate resources more optimally, enhance the quality and/or intensity of discharge planning, and facilitate the use of follow-up care.

Recommendations

Given the insights derived from this study, the following are recommended

- Hospitals should adopt Gradient Boosting or Random Forest algorithms in electronic health records to identify high-risk patients at the time of discharge, enabling clinicians to prioritize follow-up and prevent avoidable 30-day readmissions through timely interventions.
- Predictive data must inform individualized discharge planning, including medication reconciliation, education, and follow-up visits. Individualised discharge processing, tailored to specific risk levels, can significantly impact readmission prevention and continuity of care.
- The ones with higher volumes of cases and lower scores must be addressed with increased personnel, quality improvement initiatives, and enhanced monitoring. Targeting these hospitals yields the greatest impact on the system weaknesses that contribute to readmissions.
- Machine learning-driven predictions may guide partnerships with home care agencies, community health workers, and outpatient clinics. Providing post-discharge monitoring and care for high-risk patients ensures continuity of care beyond the hospital walls, thereby reducing unnecessary readmissions.

- Hospitals need to improve data integration and aggregation at the departmental level. High-quality and comprehensive datasets enhance model performance, allowing for more accurate predictions and evidence-based decision-making, which in turn facilitates readmissions reduction and optimized resource allocation.
- Predictive models must be periodically retrained and reassessed with new patient data to remain current. Ongoing monitoring enables models to adapt to changing patient populations, medical practices, and healthcare policies, thereby maintaining long-term effectiveness in reducing readmissions.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Afrash, M. R., Kazemi-Arpanahi, H., Shanbehzadeh, M., Nopour, R., & Mirbagheri, E. (2022). Predicting hospital readmission risk in patients with COVID-19: A machine learning approach. *Informatics in Medicine Unlocked, 30*, 100908. doi:https://doi.org/10.1016/j.imu.2022.100908

[2] Alkhawaldeh, A., Abashtawy, M., Rayan, A., Abdalrahim, A., Musa, A., Eshah, N., . . . Albashtawy, S. (2023). Application and Use of Andersen's Behavioral Model as Theoretical Framework: A Systematic Literature Review from 2012–2021. *Iranian Journal of Public Health, 52*(7), 1346-1354. doi:https://doi.org/10.18502/ijph.v52i7.13236

[3] Almeida, T., Moreno, P., & Barata, C. (2025). Prediction of 30-day hospital readmission with clinical notes and EHR information.

[4] Alnazari, N., Alanazi, O. I., Alosaimi, M. O., Alanazi, Z. M., Alhajeri, Z. M., Alhussaini, K. M., . . . Azzam, A. Y. (2025). Development of explainable artificial intelligence based machine learning model for predicting 30-day hospital readmission after renal transplantation. *BMC Nephrology, 26*(1), 203. doi:https://doi.org/10.1186/s12882-025-04128-w

[5] Artetxe, A., Beristain, A., & Grana, M. (2018). Predictive models for hospital readmission risk: A systematic review of methods. *Computer Methods and Programs in Biomedicine, 164*, 49-64. doi:https://doi.org/10.1016/j.cmpb.2018.06.006

[6] Ashfaq, A., Sant'Anna, A., Lingman, M., & Nowaczyk, S. (2019). Readmission prediction using deep learning on electronic health records. *Journal of Biomedical Informatics, 97*, 103256. doi:https://doi.org/10.1016/j.jbi.2019.103256

[7] Davis, S., Zhang, J., Lee, I. R., Greiner, R., McAlister, A. F., & Padwai, R. (2022). Effective hospital readmission prediction models using machine-learned features. *BMC Health Services Research, 22*, 1415. doi:https://doi.org/10.1186/s12913-022-08748-y

[8] Gandra, A. (2024). Predicting hospital readmissions in diabetes patients: A comparative study of machine learning models. *International Journal of Health Sciences, 8*(3), 15189. doi:https://doi.org/10.53730/ijhs.v8n3.15189

[9] Gao, X., Alam, S., Shi, P., Dexter, F., & Kong, N. (2023). Interpretable machine learning models for hospital readmission prediction: a two-step extracted regression tree approach. *BMC Medical Informatics and Decision Making, 23*, 104. doi:https://doi.org/10.1186/s12911-023-02193-5

[10] Halac, D., Cocucci, C., & Camerlingo, S. (2025). Predictive machine learning model for 30-day hospital readmissions in a tertiary healthcare setting. *Bioinformatics Advances, 5*(1). doi:https://doi.org/10.1093/bioadv/vbaf121

[11] Henderson, M., Hirshon, J. M., Han, F., Donohue, M., & Stockwell, I. (2023). Predicting Hospital Readmissions in a Commercially Insured Population over Varying Time Horizons. *Journal of General Internal Medicine, 38*, 1417-1422. doi:https://doi.org/10.1007/s11606-022-07950-2

[12] Hogan, J., Arenson, M. D., Adhikary, S., Li, K., Zhang, X., Zhang, R., . . . Patzer, R. E. (2019). Assessing Predictors of Early and Late Hospital Readmission After Kidney Transplantation. *Transplantation Direct, 5*(8), E479. doi:https://doi.org/10.1097/TXD.0000000000000918

[13] Hosseinzadeh, A., Izadi, M., Verma, A., Precup, D., & Buckeridge, D. (2013). Assessing the Predictability of Hospital Readamission Using Machine Learning. *Innovative Applications of Artificial Intelligence Conference* , 1532-1538.

[14] Huang, Y., Talwar, A., Lin, Y., & Aparasu, R. R. (2022). Machine learning methods to predict 30-day hospital readmission outcome among US adults with pneumonia: analysis of the national readmission database. *BMC Medical Informatics and Decision Making, 22*(1), 288. doi:https://doi.org/10.1186/s12911-022-01995-3

[15] Krzyz, E. Z., Martinez, O. F., & Lin, H. R. (2023). Uses of Andersen health services utilization framework to determine healthcare utilization for mental health among migrants—a scoping review. *Frontiers in Public Health, 11*, 1284784. doi:https://doi.org/10.3389/fpubh.2023.1284784

[16] Lederle, M., Tempes, J., & Bitzer, E. M. (2021). Application of Andersen's behavioural model of health services use: a scoping review with a focus on qualitative health services research. *Bio-Medical Journal Open, 11*, e045018. doi::10.1136/bmjopen-2020-045018

[17] Linzey, J. R., Foshee, R. L., Srinivason, S., Fiestan, G. O., Mossner, J. M., Gemmete, J. J., . . . Pandey, A. S. (2020). The Predictive Value of the HOSPITAL Score and LACE Index for an Adult Neurosurgical Population: A Prospective Analysis. *World Neurosurgery, 137*, e166-e175. doi:https://doi.org/10.1016/j.wneu.2020.01.117

[18] Liu, V. B., Sue, L. Y., & Wu, Y. (2024). Comparison of machine learning models for predicting 30-day readmission rates for patients with diabetes. *Journal of Medical Artificial Intelligence, 7*, 23. doi:https://doi.org/10.21037/jmai-24-70

[19] Liu, W., Stanbury, C., Singh, K., Ryan, A. M., Sukul, D., Mahmoudi, E., . . . Nallamothu, B. K. (2020). Predicting 30-day hospital readmissions using artificial neural networks with medical code embedding. *Plos ONE, 15*(4), e0221606. doi:https://doi.org/10.1371/journal.pone.0221606

[20] Lundberg, S., & Lee, S. I. (2017). A Unified Approach to Interpreting Model Predictions. doi:http://dx.doi.org/10.48550/arXiv.1705.07874

[21] Lv, J., Zhang, M., Fu, Y., Chen, M., Chen, B., Xu, Z., . . . Zhao, N. (2023). An interpretable machine learning approach for predicting 30-day readmission after stroke. *International Journal of Medical Informatics, 174*, 105050. doi:htpps://doi.org/10.1016/j.ijmedinf.2023.105050

[22] Michailidis, P., Dimitriadou, A., Papadimitriou, T., & Gogas, P. (2022). Forecasting Hospital Readmissions with Machine Learning. *Jounral of Healthcare, 10*(6), 981. doi:https://doi.org/10.3390/healthcare10060981

[23] Miotto, R., Li, L., Kidd, B. A., & Dudley, J. T. (2016). Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. *Scientific Reports, 6*, 26094.

[24] Morel, D., Yu, K. C., Liu-Ferrara, A., Caceres-Suriel, A. J., Kurtz, S. G., & Tabak, Y. P. (2020). Predicting hospital readmission in patients with mental or substance use disorders: A machine learning approach. *International Journal of Medical Informatics, 139*, 104136. doi:https://doi.org/10.1016/j.ijmedinf.2020.104136

[25] Pons-Suner, P., Arnal, L., Signol, F., Mateos, M. J., Martinez, B. V., & Perez-Cortes, J. C. (2023). Prediction of 30-day unplanned hospital readmission through survival analysis. *Heliyon, 9*(10), e20942. doi:https://doi.org/10.1016/j.heliyon.2023.e20942

[26] Radhamony, R., Cross, W. M., & Banik, B. (2024). Culturally and Linguistically Diverse Community Access and Utilisation of the Mental Health Service: An Explanation Using Andersen's Behavioural Model. *Issues in Mental Health Nursing, 45*(7), 758-765. doi:https://doi.org/10.1080/01612840.2024.2359602

[27] Rajaguru, V., Kim, T. H., Han, W., Shin, J., & Lee, S. G. (2022). LACE Index to Predict the High Risk of 30-Day Readmission in Patients With Acute Myocardial Infarction at a University Affiliated Hospital. *Frontier in Cardiovascular Med, 9*(1). doi:https://doi.org/10.3389/fcvm.2022.925965

[28] Rumshisky, A., Ghassemi, M., Naumann, T., Szolovits, P., Castro, V. M., McCoy, T. H., & Perlis, R. H. (2016). Predicting early psychiatric readmission with natural language processing of narrative discharge summaries. *Transl Psychiatry, 6*(10), e921. doi:https://doi.org/10.1038/tp.2015.182

[29] Ryu, B., Yoo, S., Kim, S., & Choi, J. (2021). Thirty-day hospital readmission prediction model based on common data model with weather and air quality data. *Scientific Reports, 11*, 23313. doi:https://doi.org/10.1038/s41598-021-02395-9

[30] 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*, 57. doi:https://doi.org/10.1186/s12911-021-01423-y

[31] Sherman, E., Alejo, D., Wood-Doughty, Z., Sussman, M., Schena, S., Ong, C. S., . . . Whitman, G. (2022). Leveraging Machine Learning to Predict 30-Day Hospital Readmission After Cardiac Surgery. *The Annals of Thoracic Surgery, 114*(6), 2173-2179. Retrieved from https://www.annalsthoracicsurgery.org/issue/S0003-4975(21)X0013-6

[32] Subasi, A. (2024). Hospital readmission forecasting using artificial intelligence. *Applications of Artificial Intelligence in Healthcare and Biomedicine*, 455-520. doi:https://doi.org/10.1016/B978-0-443-22308-2.00006-8

[33] Tabak, Y. P., Sun, X., Nunez, C. M., Gupta, V., & Johannes, R. S. (2016). Predicting Readmission at Early Hospitalization Using Electronic Clinical Data. *Journal of Medical Care, 55*(3), 267-275. doi:https://doi.org/10.1097/MLR.0000000000000654

[34] Tavakolian, A., Rezaee, A., Hajati, F., & Uddin, S. (2023). Hospital Readmission and Length-of-Stay Prediction Using an Optimized Hybrid Deep Model. *Journal of Future Internet, 15*(9), 304. doi:https://doi.org/10.3390/fi15090304

[35] Teo, K., Lai, K. W., Yong, C. W., Pingguan-Murphy, B., Chuah, J. H., & Tee, C. A. (2020). Prediction of Hospital Readmission Combining Rule-based and Machine Learning Model. *International Computer Symposium (ICS)*, 352-355. doi:https://doi.org/10.1109/ICS51289.2020.00076

[36] Wang, S., & Zhu, X. (2022). Predictive Modeling of Hospital Readmission: Challenges and Solutions. *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 19*(5), 2975-2995. doi:https://doi.org/10.1109/TCBB.2021.3089682

[37] Xiong, L., Sun, T., & Green, R. (2022). Predictive analytics for 30-day hospital readmissions. *Mathematical Foundations of Computing, 5*(2), 93-111. doi:https://doi.org/10.3934/mfc.2021035

[38] Yang, F., Li, J., Yang, Z., Wu, L., Wang, H., & Sun, C. (2025). Construction and validation of a machine learning-based model predicting early readmission in patients with decompensated cirrhosis: a prospective two-center cohort study. *Bio Data Mining, 18*, 63. doi:https://doi.org/10.1186/s13040-025-00479-0

[39] Zhang, Y., Xiang, T., Wang, Y., Shu, T., Yin, C., Li, H., . . . Liu, X. (2024). Explainable machine learning for predicting 30-day readmission in acute heart failure patients. *iScience, 27*(7), 110281. doi:https://doi.org/10.1016/j.isci.2024.110281