

Explainable Machine Learning for Emergency Admission Intensity: A Comparative Study of Predictive Models

^{*1}Nurul Izrin Md Saleh, ²Ikme U-Daidee, ¹Nurul A. Emran, ¹Nur Atikah Arbain, ²Sahidan Abdulmana

¹Department of Applied Data Engineering, Universiti Teknikal Malaysia Melaka, 76100 Durian Tunggal, Melaka, Malaysia

²Faculty of Science and Technology, Fatoni University, Pattani, Thailand

***Corresponding author**

DOI: <https://doi.org/10.51244/IJRSI.2026.1309000084>

Received: 07 September 2026; Accepted: 12 September 2026; Published: 08 October 2026

ABSTRACT

Hospital overcrowding requires capacity-based or flow-based measurements, such as occupied beds relative to staffed capacity, boarding time, or emergency-department census. The public NHS Hospital Admitted Patient Care Activity 2023-2024 data used in this study do not contain such an endpoint. We therefore reframe the analysis as a descriptive classification audit of an emergency admission intensity label, defined as emergency admissions divided by total admissions exceeding 85%. K-Nearest Neighbors (KNN), Random Forest (RF), and Gradient Boosting (GBM) were compared using accuracy, precision, recall, and F1-score. In the reported held-out row split, GBM obtained accuracy 0.967, precision 0.9255, recall 0.9062, and F1-score 0.9158. However, because the label is deterministically constructed from contemporaneous admission counts and emergency admissions are among the model inputs, these scores quantify reproduction of the constructed rule, not prospective forecasting of overcrowding. LIME examples are consequently interpreted only as local descriptions of model behavior. The study demonstrates both the apparent performance and the methodological limits of applying explainable classifiers to annual aggregate admission data, and specifies the hospital-by-time, independently measured outcomes required for a valid overcrowding forecasting study.

Keywords: Emergency Admission Intensity; Explainable Artificial Intelligence; LIME; Machine Learning; Methodological Audit; NHS Hospital Statistics

INTRODUCTION

The issue of hospital overcrowding has been a recurring issue, leading to long waiting hours and delayed admissions to intensive care wards. This has been identified as a major challenge facing hospitals globally [1]. Overcrowding occurs when healthcare is forced to operate beyond its capacity due to a shortage of medical staff and an excessive number of patients seeking medical treatment [2]. Hospital overcrowding is primarily caused by factors such as unnecessary patient visits, lack of inpatient beds, and prolonged waiting times for available beds in wards. Research indicates that unnecessary visits often stem from inadequate standard procedures, while a shortage of inpatient beds exacerbates delays in emergency departments (EDs) and contributes to increased mortality rates among vulnerable populations, such as chronic kidney disease patients [3]. To mitigate these effects, healthcare systems can implement several strategies. Enhancing bed management and fostering departments can streamline patient flow and reduce boarding times [4]. Additionally, optimizing staffing levels in outpatient departments and employing queuing models to manage patient arrivals can significantly decrease wait times and improve overall operational effectiveness [3]. These measures can help alleviate overcrowding and enhance patient care quality.

Patient flow plays a critical role in hospital overcrowding, as inefficient management of patient movement can lead to significant delays and negative outcomes. Research indicates that effective patient flow management, including the use of artificial intelligence (AI) tools, can enhance the forecasting and monitoring of patient

admissions, transfers, and discharges, thereby alleviating overcrowding in hospitals [5]. For instance, the implementation of discharge lounges has been shown to improve patient flow by increasing discharge rates and reducing turnaround times, which directly correlates with decreased overcrowding [6]. Additionally, systematic reviews highlight that managing patient flows across various hospital departments is essential, as disruptions in one area can impact the entire system. Factors such as prolonged waiting times and inadequate staffing in emergency departments exacerbate overcrowding, underscoring the need for targeted interventions to streamline patient flow [7]. Overall, optimizing patient flow is vital for improving hospital efficiency and patient care quality. The emerging technique of Artificial Intelligence (AI) has made it possible to manage overcrowding in emergency departments hence getting more attention in the community.

Hospital overcrowding remains an important operational problem, but the endpoint must be measured independently of the predictors used to forecast it. Bed occupancy is defined using occupied beds and available or staffed bed capacity; it is not equivalent to the proportion of admissions that are emergencies. The NHS annual aggregate dataset available for this study describes admission activity by diagnosis and demographic categories but does not establish a capacity-based overcrowding outcome. Accordingly, this revised study does not claim to predict overcrowding or provide advance warning. It evaluates how three classifiers reproduce a constructed emergency admission intensity label and uses XAI to inspect that reproduction.

The contribution is methodological rather than operational: (i) a transparent comparison of KNN, Random Forest, and Gradient Boosting on the constructed label; (ii) an examination of local model behavior using LIME; and (iii) an explicit audit of target validity, feature leakage, temporal interpretation, and causal overreach. This framing prevents classification performance on annual diagnostic aggregates from being mistaken for a deployable hospital forecasting system.

Machine learning performance is meaningful only when the target corresponds to the intended construct, predictors are available at the stated prediction time, and validation reflects the intended deployment setting. In the present data, the outcome is derived from contemporaneous admission counts, and the rows are annual diagnosis aggregates rather than hospital-by-time observations. These properties limit the analysis to descriptive rule reproduction. LIME can reveal which encoded features support a local model output, but it cannot identify causal drivers or justify operational interventions.

The objectives of this study are:

1. To compare KNN, Random Forest, and Gradient Boosting in classifying a constructed emergency admission intensity label and to benchmark their performance conceptually against the deterministic threshold rule.
2. To use LIME to describe selected local model outputs while assessing target validity, leakage, temporal validity, and the limits of operational interpretation.

RELATED WORK

This section reviews data visualization, machine learning for patient-flow applications, and explainable AI. Studies that predict future demand from hospital-by-time observations are distinguished from the present descriptive analysis of annual diagnosis aggregates. This distinction is essential because a model that reproduces a contemporaneous constructed label does not constitute an overcrowding forecast.

Data Analysis and Visualization in Patient Flow Management

Data analysis and visualization are crucial tools for managing and interpreting healthcare data, significantly enhancing the management of medical resources and improving patient service delivery [8]. Data visualization enables healthcare executives and staff to view data in an accessible format, making it easier to identify patterns, trends, and bottlenecks in patient admissions and movements [9]. Python libraries such as Matplotlib, Seaborn, Plotly, Bokeh, Altair, and ggplot are used to create detailed visualizations that support better decision-making [10]. Utilizing these visualization tools allows hospitals to optimize resource allocation, manage bed occupancy more effectively, and reduce patient waiting times clearly.

An example of data visualization application is a study in Southwest Ethiopia, where a health information system was developed to aggregate data from 21 healthcare facilities over 41 months. Using Python Sankey diagrams, the researchers visualized patient flow and employed machine learning algorithms to achieve high prediction accuracy for outpatient flows [11]. The study found that Sankey diagrams effectively visualized patient flow across healthcare facilities, enabling stakeholders to monitor and predict patient movements with high accuracy (up to 85%) [11].

Additionally, Exploratory Data Analysis (EDA) using Python libraries such as Pandas and Matplotlib plays a crucial role in cleaning and visualizing healthcare data. This aids in discovering trends and relationships that inform patient care strategies [12]. Data visualization enhances understanding of complex datasets, allowing healthcare professionals to identify patterns and relationships crucial for evidence-based decision-making [12], [13].

Machine Learning Techniques for Managing Overcrowding

Managing hospital overcrowding is a critical challenge that directly affects the quality of patient care and resource management within healthcare facilities. Machine learning (ML) techniques play an essential role in forecasting patient flow, optimizing resource allocation, and enhancing existing services to reduce congestion. Predicting patient flow and hospital admissions can be achieved through predictive modeling techniques such as Random Forest, K-Nearest Neighbors (KNN), and Support Vector Machines (SVM). These models are instrumental in analyzing patient data and forecasting future admission volumes. For instance, a study conducted in the southwestern region of Ethiopia utilized ML models to predict outpatient and inpatient flow, achieving an accuracy of up to 85% for outpatient admissions and 83% for predicting overall patient flow. Techniques like NearMiss, SMOTE, and SMOTE-Tomek were employed to address data imbalance issues commonly found in patient data, significantly enhancing model performance and reliability [11]. These models are invaluable for anticipating patient demand and effectively planning hospital resource allocation, helping reduce overcrowding and improving service delivery efficiency.

In the context of managing patient flow in emergency departments, classification algorithms have been applied to predict and manage patient length of stay (LOS). A study in Nigeria explored various classification techniques, including SVM, Classification and Regression Trees (CART), and Random Forest, to forecast LOS in emergency rooms. The study found that the SVM algorithm performed the best, with an accuracy of 0.986984 and a Mean Squared Error (MSE) of 0.358594, demonstrating its effectiveness in predicting LOS and managing patient flow [15]. This high accuracy allows hospitals to better manage resources and patient treatment times, thereby reducing congestion and enhancing service efficiency.

Additionally, clustering techniques are pivotal in analyzing and managing inpatient bed demand by identifying patterns and trends within the data, which facilitates accurate predictions of bed requirements. The study "Machine Learning Based Forecast for the Prediction of Inpatient Bed Demand" employed K-means clustering combined with Support Vector Machine Regression (K-SVR) to predict inpatient bed demand. The study achieved a Mean Absolute Percentage Error (MAPE) ranging between 0.49% and 4.10%, highlighting the effectiveness of clustering and regression techniques in improving bed management and alleviating hospital congestion [16]. These techniques enable hospitals to better plan admissions, reduce waiting times, and optimize the allocation of limited bed resources. Overall, the application of machine learning techniques in hospital overcrowding management demonstrates significant potential in forecasting patient flow, optimizing resource allocation, and enhancing medical services. These approaches contribute to reducing congestion and improving the overall patient care experience.

The Role and Challenges of Explainable AI (XAI) in healthcare.

In recent years, artificial intelligence (AI) has become increasingly significant in healthcare, particularly in diagnostics and treatment recommendations. However, a crucial challenge is enabling users to understand and trust AI model outcomes. To address this issue, Explainable Artificial Intelligence (XAI) has emerged as a key concept, providing in-depth explanations of AI model operations. This helps medical professionals understand the rationale behind AI decisions, enhancing transparency and fostering trust between AI systems and

healthcare providers [17]. Additionally, XAI techniques contribute to improving decision-making processes, ensuring that AI systems operate effectively and are understandable [17].

Among the important XAI techniques are Local Interpretable Model-Agnostic Explanations (LIME) and Shapley Additive explanations (SHAP), which have been applied in various healthcare contexts. LIME is a tool that provides pixel-level explanations of model outcomes, which is particularly useful in medical imaging tasks such as breast cancer diagnosis. LIME allows physicians to visualize model operations and understand decision-making on a granular level [18]. SHAP, on the other hand, offers a robust framework for understanding the contributions of individual features to model predictions. SHAP provides clear attribution scores, enabling detailed analysis of model outcomes. However, its performance can be influenced by model choice and feature relationships [19].

Despite their significant benefits, XAI techniques like LIME and SHAP face several challenges. One key challenge is the need for systematic evaluation and improvement to ensure these methods are effective in diverse healthcare scenarios. Additionally, there is a need for developing mechanisms that can adapt to complex feature relationships, which remains a major limitation of current XAI applications [20]. Integrating XAI techniques with medical models requires systematic assessment to ensure reliability and practical applicability. Future research should focus on refining these methods to overcome existing limitations and enhance the capabilities of XAI in personalized medicine [21].

XAI supports model inspection, but an explanation is not evidence of causation. LIME fits a local surrogate around one observation and reports feature contributions to the model output under its perturbation scheme. Its weights should therefore be described as local associations with the predicted class, not as clinical effects, capacity mechanisms, or evidence that changing a feature will improve operations. Local fidelity and stability should be reported before explanations are used beyond exploratory model auditing.

METHODOLOGY

The analysis follows the CRISP-DM structure shown in Figure 1. In the revised framing, the task is a retrospective classification audit of a constructed emergency admission intensity label. No prediction time or future horizon is claimed because all variables summarize the same 2023-2024 financial year.

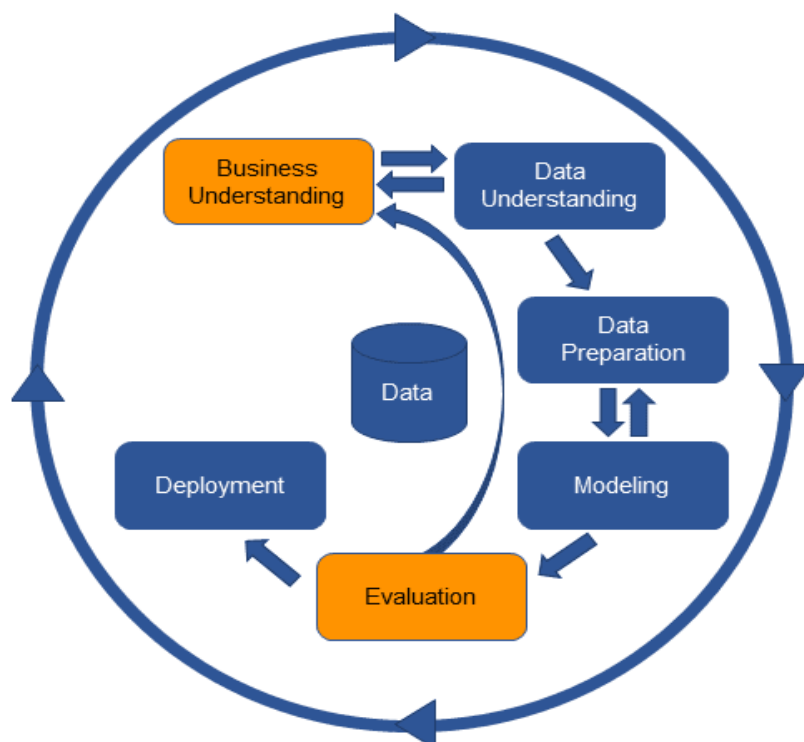


Figure 1: CRISP-DM

Business Understanding:

The source data are public NHS admitted-patient-care aggregates. Although an interview with staff at a Thailand hospital informed the broader motivation, it did not validate the constructed outcome, provide an independent capacity measure, or establish transportability to Thailand. The empirical conclusions therefore remain limited to the analyzed NHS aggregate records. The interview also identified the needs of key stakeholders—hospital administrators, clinicians, and resource managers, as they are the primary beneficiaries of such predictive insights. Despite the use of NHS data, predictions related to patient flow and resource optimization may prove valuable in settings similar to the Thailand hospital referenced in the interview. Attributes such as demographics, diagnoses, and admission histories in the NHS dataset were emphasized, as these factors often impact patient flow. From the interview, potential data limitations, especially concerning differences between the NHS data structure and the intended deployment environment are anticipated.

Finally, concrete data science tasks from these insights to address hospital management priorities are outlined. Below are findings gain from the interview:

- Models from NHS data to predict long hospital stays or frequent readmissions, offering hospital staff actionable insights into patient throughput and bed occupancy.
- Model would be deployed, whether for real-time resource planning or for periodic reporting is important for strategic oversight.
- By incorporating explainable AI (e.g., LIME), model's predictions are interpretable, allowing healthcare providers to act confidently on its insights.
- By aligning NHS data with practical applications in hospital management, resource optimization and improved patient care are able to establish the model as a valuable decision-support tool.

Data Understanding:

The analysis uses annual aggregate admission information. The observational unit is a diagnosis-level aggregate record rather than a patient encounter, hospital-day, or hospital-month. The workbook contains diagnostic totals and subcategories; where hierarchical levels overlap, a random row split may place related aggregates in both training and test sets. The archived analysis did not retain a complete reproducibility record of the exact workbook table, final row count, exclusions, class counts, random seed, or hospital identifiers. These omissions are reported as limitations and must be corrected in any confirmatory rerun.

Data Collection

The data used for this project was obtained from NHS Digital, specifically from the Hospital Episode Statistics (HES) dataset for admitted patient care during the 2023-24 financial year. The dataset was downloaded directly from the official NHS website [22], which provides comprehensive information about hospital admissions, diagnoses, treatments, and patient demographics.

Data Description

As described in Table 1, we gather hospital admission data that encompasses various details about patient diagnoses, demographic information, and admission types. The primary focus is on a comprehensive range of diagnoses, classified according to the ICD-10 system [23]. ICD-10 organizes diseases and health conditions into multiple groups [24], covering a broad spectrum of medical conditions:

Table 1: Hospital Admission Data by ICD-10 Diagnosis Classification

Code Range	Description
A00-B99	Certain infectious and parasitic diseases
C00-D49	Neoplasms

D50-D89	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
E00-E89	Endocrine, nutritional and metabolic diseases
F01-F99	Mental, Behavioral and Neurodevelopmental disorders
G00-G99	Diseases of the nervous system
H00-H59	Diseases of the eye and adnexa
H60-H95	Diseases of the ear and mastoid process
I00-I99	Diseases of the circulatory system
J00-J99	Diseases of the respiratory system
K00-K95	Diseases of the digestive system
L00-L99	Diseases of the skin and subcutaneous tissue
M00-M99	Diseases of the musculoskeletal system and connective tissue
N00-N99	Diseases of the genitourinary system
O00-O9A	Pregnancy, childbirth and the puerperium
P00-P96	Certain conditions originating in the perinatal period
Q00-Q99	Congenital malformations, deformations, and chromosomal abnormalities
R00-R99	Symptoms, signs, and abnormal clinical and laboratory findings, not elsewhere classified
S00-T88	Injury, poisoning, and certain other consequences of external causes
U04-U82	Codes for special purposes
Z00-Z99	Factors influencing health status and contact with health services

About Dataset

The records summarize completed admission activity by diagnosis within one financial year. They are not patient-level records and do not establish repeated observations for individual hospitals over time. Accordingly, the study cannot estimate future demand or validate transfer across hospitals.

Unit of analysis and scope

The dataset consists of hospital admission statistics for various diagnoses, detailing the outcomes of finished admission episodes (FAEs) across different categories. represented by ICD-10 codes. Table 2 lists all the variables used in the analysis.

Table 2: Independent variables in the study

Variable	Description
Code	ICD-10 code representing the diagnosis
Diagnosis	Description of the diagnosed condition
Diagnosis Category	Category under which the diagnosis falls, as classified by ICD-10
Finished Admission Episodes (FAEs)	Total number of completed admissions for each diagnosis
Admission Types	Classification of admissions as Emergency, Waiting List, Planned, and Other

Mean Time Waited (Days)	Average time waited for treatment for each diagnosis
Mean Length of Stay (Days)	Average duration of stay in the hospital for each diagnosis
Age Distribution	Number of patients in various age categories

Visualization

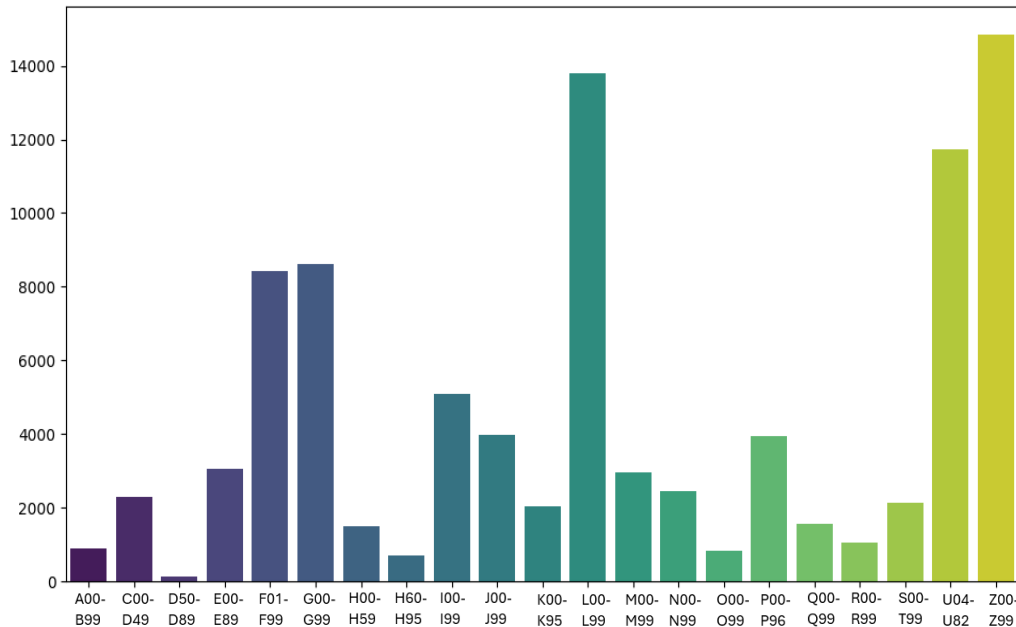


Figure 2: shows the distribution of emergency admissions across different diagnosis codes.

From the bar chart in Figure 2, the diagnosis codes Z00-Z99, K00-K95, and U04-U82 have the highest number of emergency admissions, respectively. This may indicate the severity and urgency of treatment required for diseases in these categories. On the other hand, the diagnosis codes D50-D89, H60-H95, O00-O9A, and R00-R99 show very low numbers of emergency admissions, which may reflect the ability to manage these conditions without the need for urgent care, such as planned treatments or effective symptom management. This analysis highlights the importance of prioritizing resource allocation, such as medical personnel and equipment, for diagnosis categories with high emergency admission numbers to ensure that treatment demands can be met efficiently and promptly.

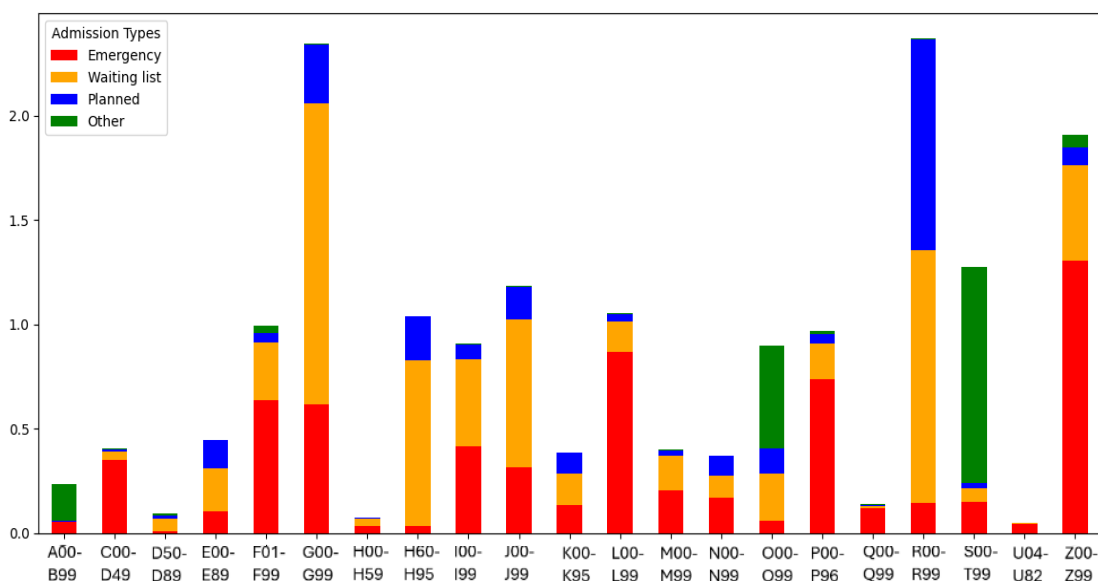


Figure 3: Distribution of Admission Types Across Different diagnosis codes

Figure 3 shows the distribution of admission types (Emergency, Waiting list, Planned, Other) across different diagnosis codes, providing a clear overview of how patient care is managed and distributed within each diagnosis group. Diagnosis codes G00-G99 shows a high number of patients requiring emergency care, with a notable proportion of emergency admissions and waiting list cases compared to planned or other types of care. This suggests that the conditions in this category are severe and require urgent treatment. In contrast, diagnosis codes R00-R99 and Z00-Z99 display a high number of emergency admissions, yet also show significant numbers of planned admissions, indicating that care for these conditions may involve both urgent and planned treatments. Additionally, some diagnosis codes, such as A00-B99, C00-D49, E00-E89, I00-I99, and J00-J99, demonstrate low emergency admissions but higher numbers in planned or waiting list categories, suggesting that these conditions are more manageable in advance without the need for emergency care. The information from this chart can be used for planning and allocating medical resources to respond effectively and promptly to patient needs in each category.

Figure 4: Mean Time Waited and Mean Length of Stay by Diagnosis Code

Figure 4 presents a graph the Mean Time Waited and Mean Length of Stay (LOS), categorized by different diagnosis codes. From the graph, it is evident that the Mean Time Waited (represented by the blue bars) varies significantly across different diagnosis codes, with some diagnosis codes showing a longer waiting time, such as those that require patients to wait over 90 days, while others have much shorter waiting times. Regarding the Mean Length of Stay (represented by the green bars), the length of stay for many diagnoses is relatively short, averaging no more than 10 days. However, there are some diagnosis codes that require longer stays, particularly those that involve special care needs. Despite the differences in the length of stay, waiting times are generally longer than the length of stay, highlighting challenges in bed management and resource allocation within hospitals. The data from this graph can be used to improve hospital bed management, particularly in predicting bed requirements and developing more efficient care plans.

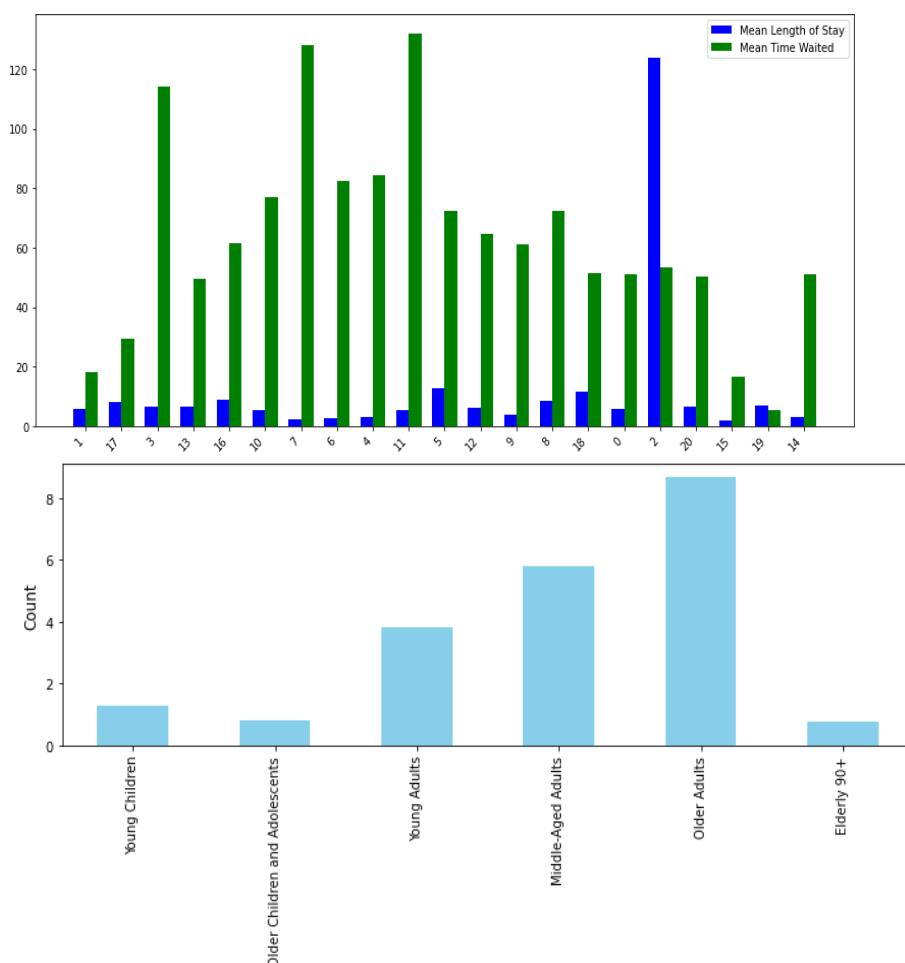


Figure 5: Distribution of Patients by Age Group

The bar chart in Figure 5 illustrates the distribution of patients across various age groups, showing a clear variation in the number of patients within each category. On the left side of the chart, we observe age groups with relatively few patients, which are likely to represent younger populations, such as children or young adults, where the incidence of medical conditions requiring hospitalization is lower. As we move towards the right side, there is a noticeable increase in the number of patients in older age groups, peaking at a specific age range. This suggests a higher prevalence of health conditions or hospitalizations in those age categories, potentially due to chronic diseases or more frequent medical interventions required at these ages. The age group with the tallest bar represents the category with the highest number of patients, which may indicate that this age group has the greatest healthcare needs, possibly related to common health issues among middle-aged adults or elderly individuals. Overall, the graph highlights how hospitalization trends differ by age, with older age groups showing a higher demand for healthcare services.

Data Preparation

In this project, the data was prepared and cleaned using a comprehensive set of steps (Figure 6) to ensure its readiness for analysis.



Figure 6: Data Preparation Process

Data Cleaning

Handling Missing Data:

The original workflow replaced all missing entries with zero. This step cannot distinguish genuinely absent activity from unavailable or suppressed values and may bias both distances and tree splits. Because the source missing-value codes were not retained in the archived analysis, the revised paper treats zero imputation as a limitation rather than a validated preprocessing choice. A reproducible rerun should document each source code, fit imputation using training data only, preserve suppression indicators where appropriate, and perform sensitivity analyses.

Data Transformation

Label Encoding:

The original workflow applied LabelEncoder to Diagnosis_categories. This imposes arbitrary ordinal distances between nominal diagnoses and is particularly unsuitable for Euclidean-distance KNN. The reported KNN result is retained only as an archival descriptive result. A rerun should use one-hot encoding or another justified nominal representation fitted on the training set, and should standardize continuous features within each training fold before KNN tuning.

Feature Engineering

Create new features:

‘Age Binning’

Age columns were consolidated for descriptive analysis. The source columns listed in the archived workflow do not support the stated 5-17 category because only ages 5-9 and 10-14 were included; ages 15-17 were not explicitly represented. The revised interpretation therefore labels that derived group as 5-14 unless the original workbook contains a separate 15-17 field. Every derived bin should be reconciled against source totals before model rerunning.

These were grouped into six major categories:

- Young Children (0-4 years): Combined data from 'Age 0' and 'Age 1-4'
- Older Children and Adolescents (5-14 years in the archived transformation): Combined data from 'Age 5-9' and 'Age 10-14'. Ages 15-17 require an explicit source column before a 5-17 label can be used.
- Young Adults (18-39 years): Spanned data from 'Age 18' to 'Age 35-39'
- Middle-Aged Adults (40-64 years): Included ages from 'Age 40-44' to 'Age 60-64'
- Older Adults (65-89 years): Grouped ages from 'Age 65-69' to 'Age 85-89'
- Elderly (90 years and above): Data from 'Age 90+'

Age aggregation can simplify presentation, but it does not by itself improve robustness or prevent overfitting. The validity of each bin depends on complete, non-overlapping source columns and verified reconciliation to the original totals. In this archived analysis, age binning should be regarded as a descriptive transformation. Its effect on model performance was not isolated through an ablation or sensitivity analysis.

‘Emergency Admission Ratio’

For each aggregate record, the constructed emergency-admission ratio was calculated as emergency admissions divided by total admissions.

$$\text{Emergency Admission Ratio (\%)} = \left(\frac{\text{Emergency}}{\text{Finished Admission Episodes}} \right) \times 100$$

A binary label was then defined as 1 when this ratio exceeded 0.85 and 0 otherwise. The label is named high emergency admission intensity, not overcrowding. The 0.85 value was originally borrowed from bed-occupancy discussions, but those measures have different denominators and meanings; therefore, 0.85 has no validated capacity interpretation in this dataset.

Variable Selection:

The target y is the constructed high emergency admission intensity label. Predictors X are contemporaneous annual aggregate features. Because emergency admissions and variables contributing to total admissions are retained among the predictors, the feature set contains direct ingredients of the target. Performance must therefore be interpreted as approximation of the labeling rule rather than independent prediction.

1. Feature matrix X: contemporaneous annual diagnosis-level aggregates, including admission-type and demographic features.
2. Target y: high emergency admission intensity, defined deterministically by emergency admissions / total admissions > 0.85.

Data Splitting

The archived workflow used a 70:30 random row split. This split can estimate reproduction of the constructed rule on held-out aggregate rows, but it cannot demonstrate transfer to a later period, another hospital, or future demand.

No temporal holdout was possible because the records summarize one financial year and do not establish hospital-by-time observations. Related diagnostic totals and subcategories may also cross the random split. A valid forecasting design would define a prediction time and horizon, use only information available at that time, and evaluate later periods and separate hospitals.

Modeling

Three classifiers—KNN, Random Forest, and Gradient Boosting—were compared as descriptive learners of the constructed label. Their scores are not interpreted as overcrowding forecasts, and they are not evidence of deployable clinical or operational utility.

The reasons for selecting these three models are as follows:

K-Nearest Neighbors (KNN):

KNN was chosen for its simplicity and suitability for less complex datasets. Although KNN does not require a complicated training process, it has limitations in terms of performance when working with large datasets or high-dimensional data [28]. Thus, KNN is ideal for smaller, less complex prediction tasks, helping us understand how a basic model works without requiring numerous parameters. It also serves as a benchmark for comparing more complex models.

Random Forest (RF):

Random Forest was selected for its ability to handle high-dimensional data and imbalanced classes effectively. It uses the bagging process, which helps reduce the risk of overfitting. This model tends to provide accurate results when dealing with complex data and uneven data distributions [29]. The use of Random Forest allows us to leverage ensemble learning, combining decisions from multiple trees to strengthen the model and reduce overall error.

Gradient Boosting (GB):

Gradient Boosting was chosen for its ability to capture non-linear relationships and improve model performance sequentially. This is particularly useful for complex, imbalanced datasets [28], [29]. Gradient Boosting often yields excellent results when predicting complex data with a lot of noise. By selecting Gradient Boosting, we aim to improve the accuracy of our predictions efficiently.

K-Nearest Neighbors (KNN)

KNN assigns a held-out aggregate record to the majority class among nearby training records under a chosen distance metric. In the archived workflow, diagnosis categories were label-encoded and feature scaling was not documented. These choices make the KNN distance geometry difficult to justify; the result is reported for completeness and should be rerun with nominal encoding and training-fitted scaling.

$$d(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

where:

- $d(p, q)$ is the distance between points p and q ,
- n is the number of features,
- p_i and q_i are the feature values for points p and q .

GridSearchCV was used to select k in the archived workflow. The candidate grid, cross-validation configuration, scoring rule, preprocessing pipeline, and random seed were not preserved in the manuscript and are required for reproducibility.

Random Forest

Random Forest combines bootstrapped decision trees and majority voting. In this study it classifies the constructed emergency admission intensity label; it does not predict an independently observed capacity state.

RandomizedSearchCV was used in the archived workflow, but the complete search space, selected hyperparameters, seed, and model-selection record were not retained in the manuscript. The reported score should therefore be treated as provisional until regenerated from a documented pipeline.

Gradient Boosting

Gradient Boosting is an ensemble learning technique that combines multiple weak models, typically decision trees, to create a stronger model through sequential training. In each step, a new model is trained to correct the errors (residual errors) made by the previous model. This makes the process additive, where each new model builds upon the errors of the previous one, leading to incremental improvements. The model minimizes the loss function by training a new model to fit the residual errors of the previous model. Parameters of the model are adjusted based on the gradient of the loss function using gradient descent. The gradient indicates the direction in which the model should adjust its parameters to reduce the loss value.

GridSearchCV was used for the archived Gradient Boosting analysis. The full parameter grid, selected values, fold assignments, scoring rule, and seed were not retained. These details, together with confusion matrices and uncertainty intervals, must be produced in a reproducible rerun before firm model ranking.

Evaluation

Accuracy, precision, recall, and F1-score describe agreement with the constructed label. They do not validate that label as overcrowding. Because the exact threshold rule has access to its own component variables, it provides a deterministic reference with perfect agreement by construction; a machine-learning model that uses the same information offers no independent forecasting evidence.

Accuracy: Measures the overall correctness of the model, calculated as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision: Indicates the proportion of true positive predictions among all positive predictions:

$$Precision = \frac{TP}{TP + FP}$$

Recall: Measures of the model's ability to identify true overcrowding instances:

$$Recall = \frac{TP}{TP + FN}$$

F1-Score: The harmonic means of Precision and Recall, providing a balance between both metrics:

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

TP: True Positive, TN: True Negative, FN: False Negative, FP: False Positive

The metrics reported below are the internally consistent values from the manuscript table. Confusion matrices, test-set counts, confidence intervals, calibration, and complete model-selection details were unavailable in the archived report and are not reconstructed here. They are mandatory outputs for the next reproducible run.

Explainable AI (XAI) using LIME

LIME is used only to describe local GBM behavior for selected test rows. The explained class and predicted probability are stated for each example. The archived analysis did not report the number of perturbations, discretization and sampling scheme, kernel width, local surrogate fidelity, or stability across random seeds; consequently, the explanations are illustrative rather than validated.

What is LIME?

LIME is a model-agnostic explainability technique that approximates the behavior of a complex model locally around a single prediction. It creates a simpler, interpretable model (e.g., a linear regression or decision tree) to explain the black-box model's decision for a specific instance. This is achieved by perturbing the input data and observing how the black-box model's predictions change [31].

Steps for Using LIME

For each selected row, LIME perturbed the encoded features and fitted a weighted local surrogate to approximate the GBM near that row. Reported weights indicate support for a model class within that local surrogate. They do not estimate causal effects. A reproducible analysis should report the perturbation distribution, number of samples, kernel settings, local fidelity (for example, weighted R-squared), and stability across repeated runs.

RESULT

New feature: Age Binning (Figure 7 & 8)

The archived transformation reduced 24 age-related columns to six reported groups. However, the group described as 5-17 was calculated only from 5-9 and 10-14 columns. It is therefore corrected to 5-14 for the current descriptive analysis; ages 15-17 must be added explicitly if available and all bin totals must be verified against the source workbook.

	Diagnosis	Age 0	Age 1-4	Age 5-9	...	Age 75-79	Age 80-84	Age 85-89	Age 90+
1	Cholera	0	0	0	...	0	1	0	0
2	Typhoid and paratyphoid fevers	1	46	93	...	7	1	1	0
3	Other salmonella infections	89	120	102	...	102	75	42	19
4	Shigellosis	2	31	45	...	18	21	12	8
5	Other bacterial intestinal infections	144	313	233	...	3946	3414	3079	1927

Figure 7: Example data before the binning process

	Diagnosis	Young Children	Older Children and Adolescents	Young Adults	Middle-Aged Adults	Older Adults	Elderly 90+
1	Cholera	0	0.0	3.0	12.0	2	0
2	Typhoid and paratyphoid fevers	47	214.0	470.0	206.0	39	0
3	Other salmonella infections	209	222.0	603.0	621.0	474	19
4	Shigellosis	33	63.0	202.0	146.0	99	8
5	Other bacterial intestinal infections	457	701.0	3916.0	7224.0	15526	1927

Figure 8: Example data after the binning process

The age groups improve readability, but no evidence is available that binning improved predictive performance. The transformed totals should be treated as provisional until reconciled with the source age columns.

Constructed high emergency admission intensity label (Figure 9)

The binary outcome equals 1 when emergency admissions / total admissions exceeds 0.85. It is a constructed emergency admission intensity indicator and is not a measure of bed occupancy, staffed capacity, boarding, queue length, or hospital crowding. The threshold is retained only to audit the original experiment; it is not presented as a clinically or operationally validated cutoff.

	Diagnosis	Diagnosis_categories	...	Emergency Admission Ratio (%)	Overcrowding_Status
1	Cholera	1	...	66.666667	0
2	Typhoid and paratyphoid fevers	1	...	72.515528	0
3	Other salmonella infections	1	...	82.642089	0
4	Shigellosis	1	...	88.919668	1
5	Other bacterial intestinal infections	1	...	81.266473	0
6	Other bacterial foodborne intoxications, not e...	1	...	96.963563	1
7	Amoebiasis	1	...	45.238095	0
8	Other protozoal intestinal diseases	1	...	85.751979	1
9	Viral and other specified intestinal infections	1	...	97.549924	1
10	Other gastroenteritis and colitis of infectiou...	1	...	75.737653	0

Figure 9: Sample data showing the constructed high emergency admission intensity label.

Descriptive agreement with the constructed label

Table 3 reports one internally consistent set of held-out row-split metrics from the archived analysis. The comparison evaluates agreement with the constructed label, not forecasting performance.

Table 3: Descriptive model agreement with the constructed high emergency admission intensity label

METRIC	K-NEAREST NEIGHBORS	RANDOM FOREST	GRADIENT BOOSTING
ACCURACY	0.8845	0.9629	0.967
PRECISION	0.7083	1	0.9255
RECALL	0.7083	0.8125	0.9062
F1-SCORE	0.7083	0.8966	0.9158

Using the values in Table 3, KNN achieved accuracy 0.8845, precision 0.7083, recall 0.7083, and F1-score 0.7083. Random Forest achieved accuracy 0.9629, precision 1.0000, recall 0.8125, and F1-score 0.8966. Gradient Boosting achieved accuracy 0.9670, precision 0.9255, recall 0.9062, and F1-score 0.9158. GBM has the highest reported accuracy, recall, and F1-score, whereas RF has the highest precision. These values replace the conflicting figures previously stated in the prose. Nevertheless, the deterministic threshold rule is the proper reference and would reproduce its own label exactly when its component variables are available. Thus, none of the model scores establishes advance warning, target validity, or superiority for hospital operations.

LIME for Explainable AI

LIME was applied to selected GBM outputs to inspect local model behavior. The labels "high intensity" and "low intensity" replace "overcrowding" and "normal." Reported contributions indicate how features supported a class in the local surrogate; they are neither causal effects nor intervention recommendations. Completed waiting-list admissions are counts of admissions from a waiting list, not the number of people currently waiting or a backlog measure.

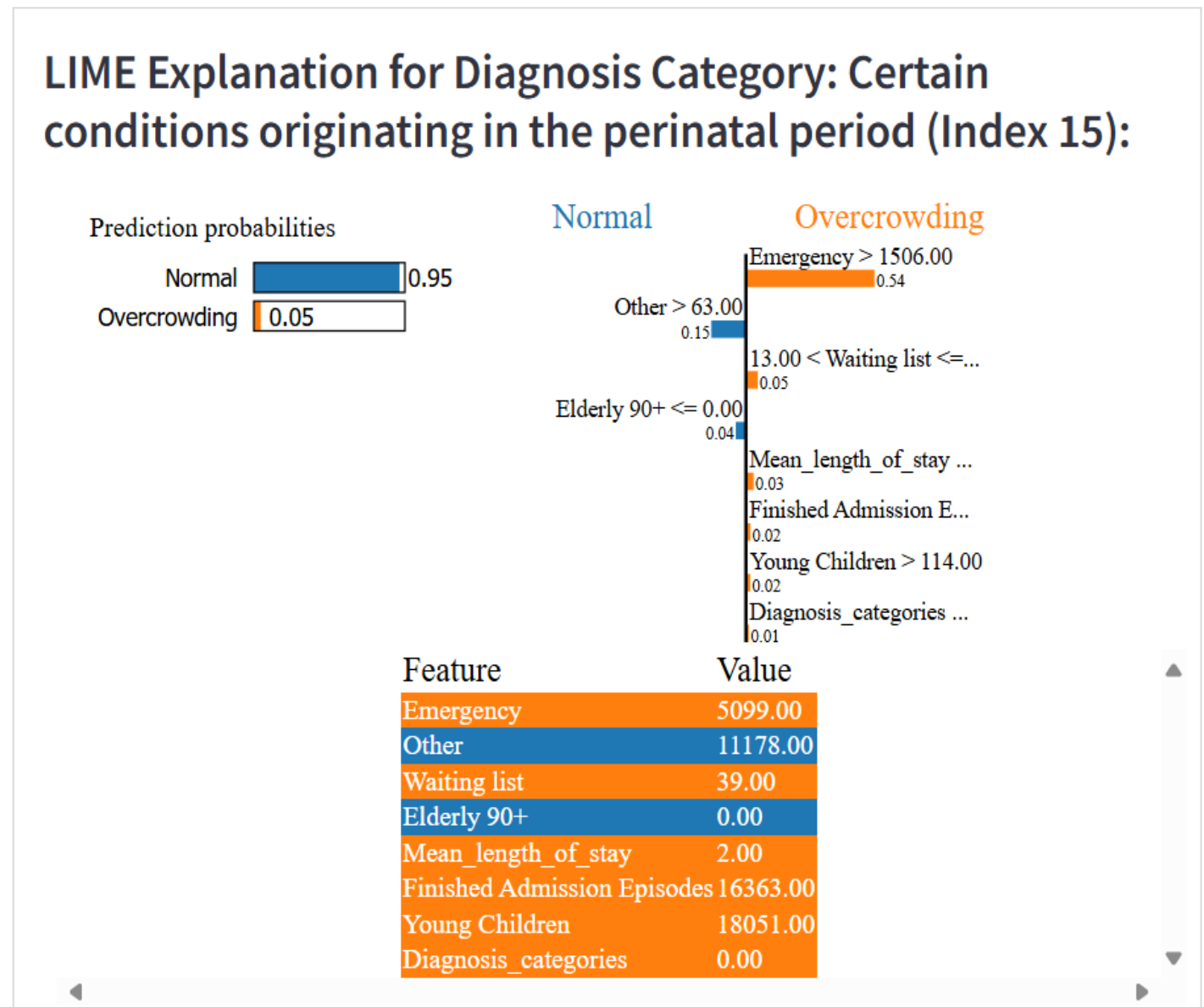


Figure 10: LIME explanation of a low emergency admission intensity output for perinatal conditions (Index 15)

Certain conditions originating in the perinatal period (Index 15)

- The GBM assigned 95% probability to low emergency admission intensity and 5% to high emergency admission intensity.
- Key contributors:
 - Emergency admissions (>1506) contributed 0.54 toward the high-intensity class in this local surrogate. This reflects model behavior and does not establish a capacity effect.
 - Other admissions (>63) and Elderly 90+ (=0) contributed toward the low-intensity class in this local surrogate.

- Other features such as Mean length of stay (≤ 2.0) had minor impacts on the overall prediction.
- Interpretation: the local surrogate relied strongly on emergency-admission volume, which is expected because emergency admissions directly contribute to the constructed target.

LIME Explanation for Diagnosis Category: Certain infectious and parasitic diseases (Index 0): ⇌

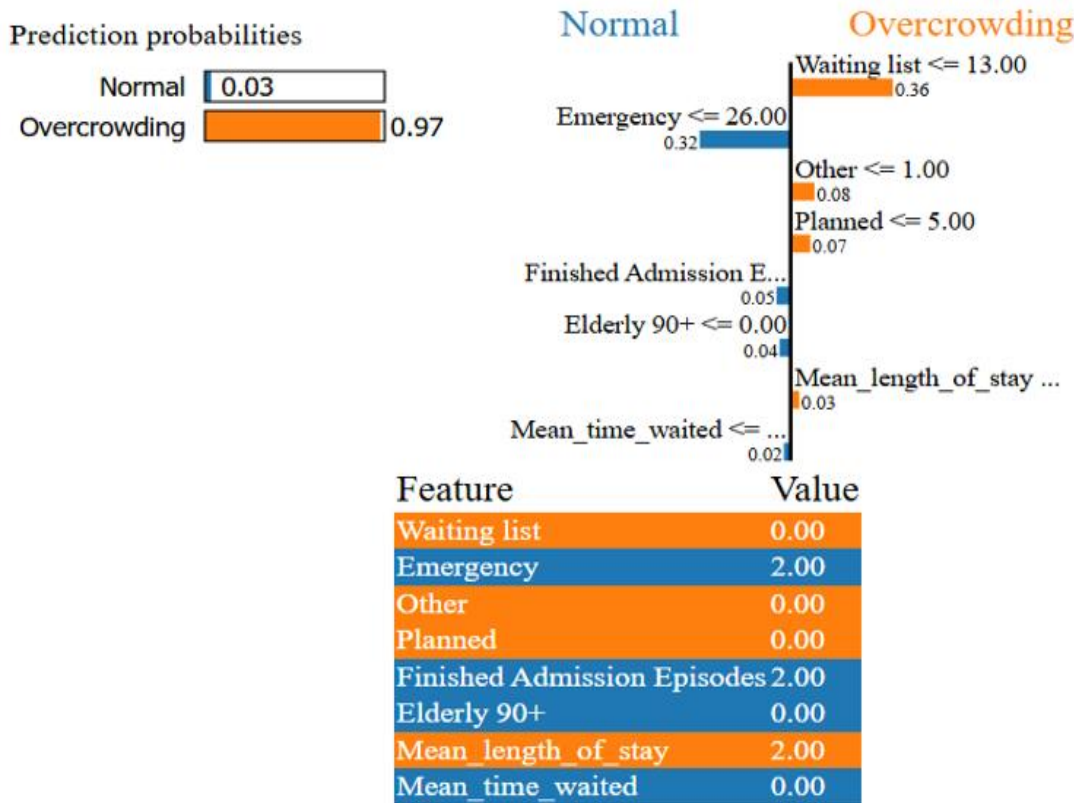


Figure 11: LIME explanation of a high emergency admission intensity output for infectious and parasitic diseases (Index 0)

Certain infectious and parasitic diseases (Index 0):

- The GBM assigned 97% probability to high emergency admission intensity and 3% to low emergency admission intensity.
- Key contributors:
 - Completed waiting-list admissions (≤ 13) contributed 0.36 toward the high-intensity class in this local surrogate. This variable is not the size of a waiting-list backlog, and no causal claim is made.
 - Emergency admissions (≤ 26) contributed toward the low-intensity class in this local surrogate.
 - Planned admissions (≤ 5) and Mean length of stay (≤ 2.0) had smaller impacts but still leaned toward overcrowding.
- Interpretation: these weights describe this model output only; they do not show that changing waiting-list activity would change crowding or capacity.

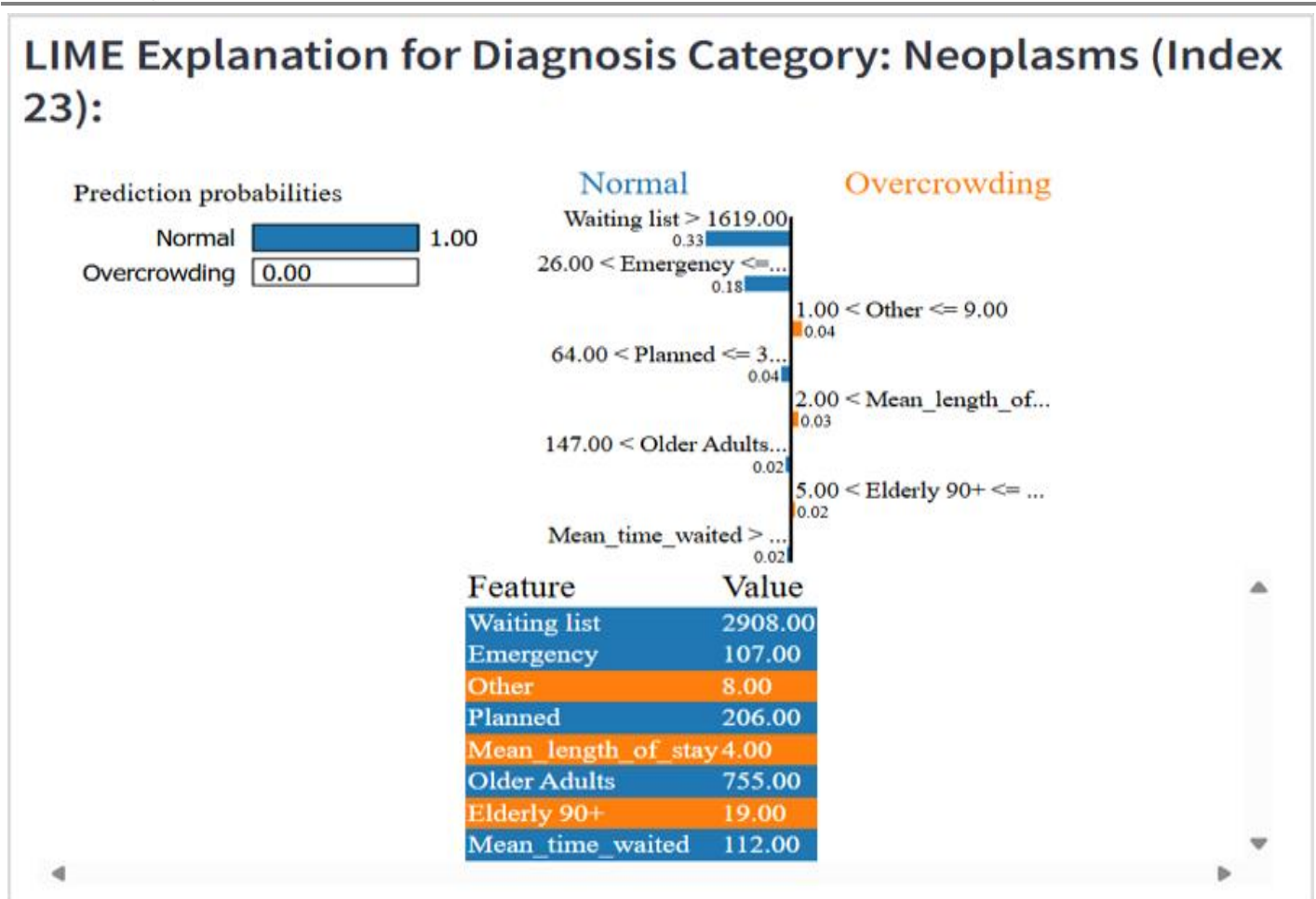


Figure 12: LIME explanation of a low emergency admission intensity output for neoplasms (Index 23)

Neoplasms (Index 23):

- The GBM assigned 100% probability to low emergency admission intensity and 0% to high emergency admission intensity.
- Key contributors:
 - Completed waiting-list admissions (>1619) contributed 0.33 toward the low-intensity class in this local surrogate. This is not evidence that a large waiting list maintains capacity.
 - Emergency admissions (≤ 26) also contributed toward the low-intensity class in this local surrogate.
 - Features like Planned admissions (≤ 64) and Older Adults admissions (≤ 147) had smaller positive impacts on the "Normal" prediction.
 - Other admissions (≤ 9) contributed minimally toward the high-intensity class.
- Interpretation: the explanation is a local account of GBM behavior and has no causal or operational meaning without an intervention study and an independently measured capacity outcome.

Across the three examples, LIME makes the GBM dependence on contemporaneous admission variables visible. Because those variables define or closely track the constructed target, this dependence primarily reveals circularity. The explanations should not guide resource allocation.

CONCLUSION

This revised study evaluates how three classifiers reproduce a constructed high emergency admission intensity label in annual NHS diagnosis aggregates. GBM has the highest reported accuracy (0.9670), recall (0.9062), and F1-score (0.9158), while RF has the highest precision (1.0000). These metrics are descriptive agreement measures, not evidence that either model forecasts hospital overcrowding.

LIME clarifies which encoded features support selected GBM outputs, but it does not identify causal drivers. The examples also expose the central limitation: emergency-admission variables contribute directly to a label derived from emergency-admission share. The exercise is therefore best understood as a target-validity and leakage audit.

Significance and implications

- The main contribution is a caution for explainable healthcare machine learning: high held-out performance can be misleading when the target is manufactured from contemporaneous predictors or does not measure the intended clinical or operational construct.
- The revised analysis provides a transparent baseline for redesigning the study, but it does not support deployment, prospective alerts, capacity planning, or claims of improved patient care.

Limitations

The data do not contain occupied beds, staffed capacity, boarding time, queue length, or another independently validated crowding endpoint. Rows are annual diagnosis aggregates rather than hospital-by-time observations. The archived workflow lacks exact row and class counts, complete workbook provenance, exclusions, seeds, confusion matrices, uncertainty intervals, calibration, selected hyperparameters, and full LIME settings. A random row split may separate overlapping hierarchical aggregates. Zero imputation, incomplete age binning, label encoding of nominal diagnoses, and undocumented KNN scaling further limit reproducibility and validity.

Future work

- A forecasting study should obtain hospital-by-time observations with an independently defined crowding endpoint, specify prediction time and horizon, retain only features available at that time, and evaluate later periods and separate hospitals. Preprocessing and model selection must be fitted within training folds, with documented seeds, hyperparameters, class counts, confusion matrices, confidence intervals, discrimination, and calibration.
- LIME analyses should report the explained class, perturbation distribution, sample count, kernel settings, local fidelity, and stability across repeated seeds. SHAP or other methods may supplement, but cannot repair an invalid target or establish causality.

In conclusion, the present results show that conventional classifiers can approximate a constructed emergency admission intensity rule, with GBM giving the strongest reported agreement among the tested learners. They do not demonstrate overcrowding forecasting. Establishing that claim requires new hospital-by-time data, an independent crowding outcome, temporal validation, and a fully reproducible analysis pipeline.

ACKNOWLEDGMENT

This work was supported by the Malaysia Ministry of Higher Education under the Fundamental Research Grant Scheme (Early Career Researcher) (FRGS/1/2024/ICT02/UTEM/02/14) and iDEAL Research Group of Universiti Teknikal Malaysia Melaka, UTeM. The authors also appreciate the valuable comments and suggestions provided by the anonymous reviewers and the editor, which helped improve the quality of this manuscript.

REFERENCES

1. W. M. W. M. A. and W. R. I., "Improving emergency department overcrowding in Malaysian government hospital," *Journal of Quality Measurement and Analysis* , vol. 17, no. 1, pp. 19–39, 2021.
2. G. Savioli, I. F. Ceresa, N. Gri, G. B. Piccini, Y. Longhitano, C. Zanza, A. Piccioni, C. Esposito, G. Ricevuti, and M. A. Bressan, "Emergency Department Overcrowding: Understanding the Factors to Find Corresponding Solutions," *Journal of Personalized Medicine* , vol. 12, no. 2, p. 279, Feb. 14, 2022.
3. J. G. Guerrero, A. S. Alqarni, R. P. Cordero, I. Aljarrah, and M. A. Almahaid, "Perceived Causes and Effects of Overcrowding Among Nurses in the Emergency Departments of Tertiary Hospitals: A Multicenter Study," *Risk Management and Healthcare Policy* , vol. 17, pp. 973–982, Apr. 20, 2024.
4. M. Oberlin, E. Andr s, M. Behr, S. Kepka, P. Le Borgne, and P. Bilbault, "Emergency overcrowding and hospital organization: Causes and solutions," *La Revue de M decine Interne* , vol. 41, no. 10, Oct. 2020.
5. CADTH, "Artificial Intelligence for Patient Flow," *Canadian Journal of Health Technologies* , vol. 4, no. 5, 2024.
6. R. Woods, R. Sandoval, G. Vermillion, B. Bates-Jackson, A. Nwankwo, C. P. Canamar, and L. Sarff, "The Discharge Lounge: A Patient Flow Process Solution," *Journal of Nursing Care Quality* , vol. 35, no. 3, pp. 240–244, Jul./Sep. 2020.
7. M. I. Abubakar, "Stochastic Simulation Modelling and Analysis of Out-Patient Flow: A Case Study of General Hospital Hadejia, North-western Nigeria," *FUOYE Journal of Engineering and Technology* , vol. 7, no. 4, Nov. 2022.
8. V. Domova and S. Sander-Tavallaey, "Visualization for Quality Healthcare: Patient Flow Exploration," in *2019 IEEE International Conference on Big Data (Big Data)* , Los Angeles, CA, USA, 2019.
9. M. Islam and S. Jin, "An Overview of Data Visualization," in *2019 International Conference on Information Science and Communications Technologies (ICISCT)* , Tashkent, Uzbekistan, 2019.
10. L. Addepalli, L. G. S. S. S. Hussain, M. J. V. Uppalapati, W. Ali, and V. S., "Assessing the Performance of Python Data Visualization Libraries: A Review," *International Journal of Computer Engineering in Research Trends* , vol. 10, no. 1, pp. 28–39, Jan. 2023.
11. B. A. Kassie, A. B. Senay, and G. S. Tegenaw, "Developing a patient flow visualization and prediction model using aggregated data for a healthcare network cluster in Southwest Ethiopia," *PLOS Digital Health* , vol. 3, no. 2, Feb. 7, 2024.
12. A. Gupta, K. B. Singh, K. Singh, P. K. Kushwaha, B. P. Lohani, and S. Kumar, "Unveiling Insights: Exploring Healthcare Data through Data Analysis."
13. G. R. S. J., K. Jothikumar, and P. Priyadharshini, "Advancing Healthcare Through Data Science Techniques for Comprehensive Analysis and Visualization of Healthcare Data," in *Cybersecurity and Data Management Innovations for Revolutionizing Healthcare* , IGI Global, 2024, pp. 1–15.
14. D. Asuquo, I. Umoren, F. Osang, and K. Attai, "A Machine Learning Framework for Length of Stay Minimization in Healthcare Emergency Department," *Studies in Engineering and Technology* , vol. 10, no. 1, 2023.
15. M. Tello, E. Reich, J. Puckey, R. Maff, A. Garc a Arce, B. Bhattacharya, and F. Feijoo, "Machine learning-based forecast for the prediction of inpatient bed demand," *BMC Medical Informatics and Decision Making* , vol. 22, no. 1, 2022.
16. S. Singhal, A. Sharma, A. Singh, and A. Pandey, "Industry-specific AI and ML applications in healthcare focus on improving diagnostics, patient care, and medical research," in *Advances in Systems Analysis, Software Engineering, and High-Performance Computing* , IGI Global, 2024, pp. 110–124.
17. A. Ponce, J. L pez-Bautista, and R. Fernandez-Beltran, "Interpretando Modelos de IA en C ncer de Mama con SHAP y LIME," *Ideas en Ciencias de la Ingenier a* , vol. 2, no. 2, p. 15, 2024.
18. A. Salih et al ., "A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME," *Advanced Intelligent Systems* , 2024.
19. G. Huang et al ., "From Explainable to Interpretable Deep Learning for Natural Language Processing in Healthcare: How Far from Reality?," *Computational and Structural Biotechnology Journal* , vol. 24, pp. 362–373, 2024.

20. V. R. R. Kovvuri, “Explainable Artificial Intelligence across Domains: Refinement of SHAP and Practical Applications,” 2024.
21. NHS Digital, “Hospital admitted patient care activity, 2023–24,” 2023. [Online]. Available: [NHS Digital – Hospital Admitted Patient Care Activity 2023–24](https://digital.nhs.uk/data-and-information/publications/statistical/hospital-admitted-patient-care-activity/2023-24?utm_source=chatgpt.com). [Accessed: Nov. 11, 2024].
22. J. Hirsch, G. Nicola, G. McGinty, R. Liu, R. Barr, M. Chittle, and L. Manchikanti, “ICD-10: History and context,” *American Journal of Neuroradiology*, vol. 34, no. 4, pp. 596–599, 2016.
23. “International Classification of Diseases 10th Revision (ICD-10),” 2019. [Online]. Available: [WHO ICD-10](https://icd.who.int/browse10/2019/en?utm_source=chatgpt.com).
24. J. Hippisley-Cox and C. Coupland, “Predicting risk of emergency admission to hospital using primary care data: Derivation and validation of QAdmissions score,” *BMJ Open*, vol. 3, no. 8, 2013.
25. N. Kraaijvanger, D. Rijpsma, L. Roovers, H. van Leeuwen, K. A. H. Kaasjager, L. van den Brand, L. Horstink, and M. J. R. Edwards, “Development and validation of an admission prediction tool for emergency departments in the Netherlands,” *Emergency Medicine Journal*, vol. 35, no. 8, pp. 464–470, 2018.
26. A. A. Tokuç, “Splitting a Dataset into Train and Test Sets,” *Baeldung*, May 12, 2023. [Online]. Available: [Baeldung – Splitting a Dataset into Train and Test Sets](https://www.baeldung.com/cs/train-test-datasets-ratio?utm_source=chatgpt.com). [Accessed: Nov. 11, 2023].
27. D. N. Cosenza, L. Korhonen, M. Maltamo, P. Packalen, J. L. Strunk, E. Næsset, T. Gobakken, P. Soares, and M. Tomé, “Comparison of linear regression, k-nearest neighbour and random forest methods in airborne laser-scanning-based prediction of growing stock,” *Forestry: An International Journal of Forest Research*, vol. 94, no. 2, pp. 311–323, Apr. 2021.
28. L. Dube and T. Verster, “Enhancing classification performance in imbalanced datasets: A comparative analysis of machine learning models,” *AIMS Data Science*, Oct. 13, 2023.
29. R. Duda, P. Hart, and D. Stork, *Pattern Classification*. John Wiley & Sons, 2001.
30. J. Dieber and S. Kirrane, “Why model why? Assessing the strengths and limitations of LIME,” 2020.
31. C. Barchielli, M. Vainieri, C. Seghieri, E. Salutini, and P. Zoppi, “The Function of Bed Management in Pandemic Times—A Case Study of Reaction Time and Bed Reconversion,” *International Journal of Environmental Research and Public Health*, vol. 20, no. 12, p. 6179, 2023.