Patient Flow Model for Hospital Admission Analysis

Prepared by

Ikmee U-daidee

Advisor

Dr. Nurul Izrin Binti Md Saleh

Department of Data Science and Analytic

Fatoni University

2024

Project Proposal

Topic: Patient Flow Model for Hospital Admission Analysis

Chapter 1: Introduction

1.1 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 the 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 community.

This project proposes k-Nearest Neighbor (KNN) model of Machine Learning to be employed and trained using hospital admission data encompassing attributes such as diagnosis, consultancy episodes, number of admission and demography. The model will identify patterns and trends to predict which diagnosis requires the patient to have longer hospital stays or readmissions to help stakeholders to prioritize resource allocation accordingly. Apart from that, this project also emphasizes on data visualization as it is essential for understanding and addressing the relationship between diagnosis and overcrowding in hospitals. It can help identify patterns, bottlenecks, and trends in the data, offering actionable insights for improving patient throughput and resource management. Data visualization using Python with libraries such as Matplotlib, Seaborn, and Plotly is a powerful approach to transforming raw data into meaningful insights through graphical representation.

1.2 Problem statement

a) Insufficient Understanding of Diagnosis-Specific Flow Patterns: The absence of data visualized of how specific diagnoses contribute to patient flow dynamics creates challenges in identifying which medical conditions are most closely associated with overcrowding at different times.

b) Difficulty in Integrating Historical Data for Predictions: Hospitals face challenges in integrating historical patient diagnosis data to create accurate prediction models, limiting their ability to anticipate and mitigate future overcrowding effectively.

c) Lack of Explainability in Prediction Models: Stakeholders struggle with interpreting and understanding the predictive models used for anticipating overcrowding. The absence of explainability makes it difficult for healthcare professionals to trust and act upon the predictions, which limits the effectiveness of these models in decision-making and patient flow management.

1.3 Objective

a) To conduct data visualization of diagnosis-specific flow patterns in order to identify which medical conditions are most closely associated with overcrowding during different times. This will aid in targeted resource allocation and improve patient flow management.

b) To develop robust methods for integrating historical patient diagnosis data into predictive models, enhancing the hospital's ability to accurately forecast diagnosed patient influx and manage future overcrowding proactively.

c) To improve the explainability of predictive models used for forecasting patient flow and overcrowding, ensuring that healthcare professionals can trust and comprehend the outputs, leading to better-informed decisions and improved management of hospital resources.

1.4 Scope of study

- Module to be developed:
  - Data use: Admitted Patient Care activity in England for the financial year 2022-23.2.
  - Tools: Implement data visualization, prediction models, and ExplainerDashboard using Jupyter Notebook.

- Target User:
  - The target user of this study are hospital executives who aim to manage resources effectively to reduce overcrowding in hospitals and improve the quality of healthcare services. They can use data and analysis to plan strategies for resource management and enhance operational efficiency. Additionally, it includes healthcare providers (doctors and nurses) who need in-depth information about patient flow, which will help them make better decisions in patient care.

1.5 Project Significance

1.6 Expected Output

1.7 Conclusion

Chapter 2: Literature review

2.1 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.

- **Matplotlib** is an established and popular library used for creating various types of graphs, such as line charts and histograms. It provides flexibility and detail for data visualization, making it a fundamental tool for healthcare data analysis and visualization [10].
- **Seaborn** builds on Matplotlib and is designed for statistical data visualization, facilitating the creation of complex plots like Heatmaps and Pair plots, which help in exploring data relationships and trends [10].
- **Plotly** supports interactive and 3D graph creation, enhancing the effectiveness of detailed dashboards and making it suitable for visualizing patient flow data [10].
- **Bokeh** focuses on creating interactive and web-based visualizations, ideal for detailed and specific data representation [10].

- **Altair** is known for its simple syntax for creating statistical graphs and interactive visualizations, making it suitable for in-depth data analysis and presentation [10].
- **ggplot** adapted from R, uses a grammar of graphics approach, allowing for straightforward and clear graph creation, which is beneficial for detailed data interpretation [10].

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]. Interactive dashboards also enable rapid data analysis, significantly improving response times in clinical settings and potentially saving lives [14].

2.2 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 improve 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.

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

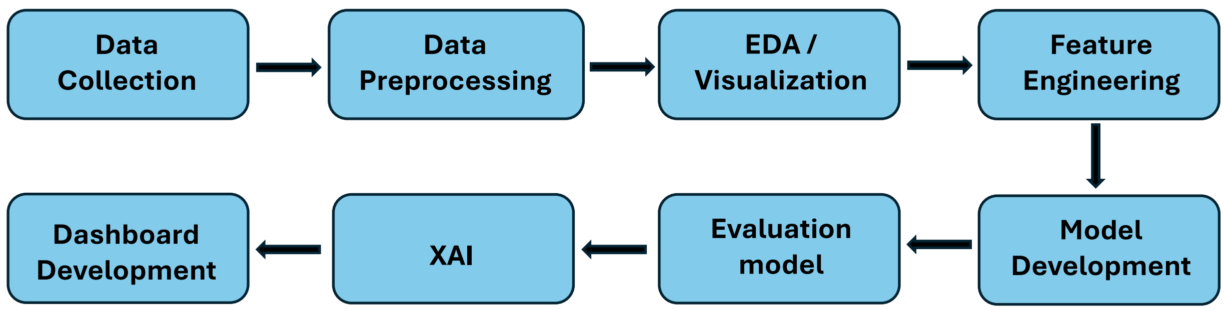
In recent years, artificial intelligence (AI) has become increasingly significant in healthcare, particularly in diagnostics and treatment recommendations. However, a crucial challenge is enabling users to understand and trust AI model outcomes. To address this issue, Explainable Artificial Intelligence (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 holds great potential for advancing and improving healthcare. By incorporating explanatory techniques such as LIME and SHAP, medical decision-making can become more transparent, fostering a better understanding of AI model outcomes and enhancing overall treatment efficacy. Ongoing research and development in this area will contribute to making AI systems more reliable and beneficial in the future.

Chapter 3: Project Methodology



3.1 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 2022-23 financial year. The dataset was downloaded from the NHS website, which provides comprehensive information about hospital admissions, diagnoses, treatments, and patient demographics.

3.2 Data Preprocessing

3.2.1 Handling Missing Values

In my context to encounter the missing:

- I use **fillna(0)** function**,** causing all missing values to be replaced with 0 so that no NaN is left in the data.

3.3.2 Convert the data

- Use **pd.to_numeric**: Correctly convert columns with numeric data

3.2.3 Handling Invalid Values

- Handle invalid values by converting them to NaN.

3.3 EDA and Visualization

I use key Python libraries for data analysis and visualization, such as Matplotlib, Seaborn, and Plotly.

This visualizing will help to understand the relationship between diagnoses and hospital overcrowding, and it will guide the direction for developing predictive models.

3.4 Feature Engineering

3.5 Model Development

KNN will be used to predict periods of overcrowding by looking at the diagnosis flow. If similar patterns in diagnoses (such as an influx of certain conditions) have previously led to overcrowding, the model could be used to flag potential future overcrowding when these patterns emerge again.

3.6 Evaluation of the Model

The model will be evaluated by splitting the data into training and testing sets (70% training, 30% testing). To accurately assess the model's performance, by using the following key metrics:

- Accuracy: Measures the correctness of the predictions.
- Precision, Recall, and F1-score: Measure the performance of predictions in cases where the data is imbalanced.
- Confusion Matrix: Used to observe details of mispredictions for each class.

I will employ k-fold cross-validation, which helps prevent overfitting and assess the model's ability to predict new data.

3.8 XAI

3.8 Dashboard Development

Chapter 4: Proposed Solution

4.1 Data Visualization

In this chapter, I have created visualizations using the libraries Matplotlib, Seaborn, and Plotly, which are popular tools for generating graphs and charts in Python.

This visualizing will help to understand the relationship between diagnoses and hospital overcrowding, and it will guide the direction for select the feature to developing predictive models.

4.2 KNN Modeling

4.3 Dashboard Development

Chapter 5: Experimental result

Reference:

[1] Aminuddin, W. M. W. M., & Ismail, W. R. (2021). Improving emergency department overcrowding in Malaysian government hospitals. Journal of Quality Measurement and Analysis JQMA, 17(1), 19-39.

[2] Savioli, G., Ceresa, I. F., Gri, N., Bavestrello Piccini, G., Longhitano, Y., Zanza, C., ... & Bressan, M. A. (202 2). Emergency department overcrowding: understanding the factors to find corresponding solutions. Journal of pers onalized medicine, 12(2), 279.

[3] Jefferson, Guerrero., A., Alqarni., Rock, Parreno, Cordero., Imad, AlJarrah., Mohsen, Almahaid. (2024). Perceived Causes and Effects of Overcrowding Among Nurses in the Emergency Departments of Tertiary Hospitals: A Multicenter Study. Risk Management and Healthcare Policy, doi: 10.2147/rmhp. s454925

[4] Mathieu, Oberlin., Emmanuel, Andrès., Martin, Behr., Sabrina, Kepka., P., Le, Borgne., Pascal, Bilbault. (2020). La saturation de la structure des urgences et le rôle de l'organisation hospitalière: réflexions sur les causes et les solutions. Revue de Médecine Interne, doi: 10.1016/J.REVMED.2020.05.023

[5] CADTH. (2024). Artificial Intelligence for Patient Flow. Canadian journal of health technologies, doi: 10.51731/cjht.2024.877

[6] Rebecca, Woods., Rebecca, Sandoval., Gregory, Vermillion., Brenda, Bates-Jackson., Audrey, Nwankwo., Catherine, P, Canamar., Laura, Sarff. (2020). The

Discharge Lounge: A Patient Flow Process Solution.. Journal of Nursing Care Quality, doi: 10.1097/NCQ.0000000000000469

[7] Maji, Ibrahim, Abubakar. (2022). 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, doi: 10.46792/fuoyejet.v7i4.922

[8] V. Domova and S. Sander-Tavallaey, "Visualization for Quality Healthcare: Patient Flow Exploration," in Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 2019, pp. 1072-1079, doi: 10.1109/BigData47090.2019.9006351.

[9] M. Islam and S. Jin, "An Overview of Data Visualization," 2019 International Conference on Information Science and Communications Technologies (ICISCT), Tashkent, Uzbekistan, 2019, pp. 1-7, doi: 10.1109/ICISCT47635.2019.9012031.

[10] Addepalli, L., Gaurav, L., Sindhuja, S., Hussain, S., Mookerjee, J., Uppalapati, V., Ali, W., & Sagar, V. (2023). Assessing the Performance of Python Data Visualization Libraries: A Review. International Journal of Computer Engineering in Research Trends, 10(1), 28-39. https://doi.org/10.22362/ijcert/2023/v10/i01/v10i0104

[11] Kassie, B. A., Senay, A. B., & Tegenaw. (2024). Developing a patient flow visualization and prediction model using aggregated data for a healthcare network cluster in Southwest Ethiopia. PLOS Digital Health, 3(2), e0000455. https://doi.org/10.1371/journal.pdig.0000455

[12] Gupta, A., Singh, K. B., Singh, K., Kushwaha, P. K., Lohani, B. P., & Kumar, S. (2024). Unveiling Insights: Exploring Healthcare Data through Data Analysis.

[13] G., Ratthna S. J., Jothikumar, K., & Priyadharshini, P. (2024). Advancing Healthcare Through Data Science Techniques for Comprehensive Analysis and Visualization of Healthcare Data. Advances in Healthcare Information Systems, IGI Global, pp. 1-15.

[14] Battineni, G., Mittal, M., & Jain, S. (2021). Data Visualization in the Transformation of Healthcare Industries. Springer, Singapore, pp. 1-23.

[15] Asuquo, D., Umoren, I., Osang, F., & Attai, K. (2023). A Machine Learning Framework for Length of Stay Minimization in Healthcare Emergency Department. Studies in Engineering and Technology, 10(1), Article 6372. https://doi.org/10.11114/set.v10i1.6372

[16] Tello, M., Reich, E., Puckey, J., Maff, R., García Arce, A., Bhattacharya, B., & Feijoo, F. N. (2022). Machine learning-based forecast for the prediction of inpatient bed demand. BMC Medical Informatics and Decision Making, 22(1).

[17] Singhal, S., Sharma, A. K., Singh, A. K., Pandey, A., & Sharma, A. K. (2024). Industry-specific AI and ML applications in healthcare focus on improving diagnostics, patient care, and medical research. Advances in Systems Analysis, Software Engineering, and High-Performance Computing. IGI Global, pp 110-124.

[18] Ponce, A. G., López-Bautista, J., & Fernandez-Beltran, R. (2024). Interpretando Modelos de IA en Cáncer de Mama con SHAP y LIME. Ideas en Ciencias de la Ingeniería, Universidad Autónoma del Estado de México, Vol. 2, Iss: 2, pp. 15-15.

[19] Salih, A., et al. (2024). A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME. Advanced Intelligent Systems.

[20] Huang, G., et al. (2024). 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.

[21] Kovvuri, V. R. R. (2024). Explainable Artificial Intelligence across Domains: Refinement of SHAP and Practical Applications. [No Journal Name Provided].