**Patient Flow Model for Hospital Admission Analysis**

**IKMEE U-DAIDEE**

**642437002**

**FATONI UNIVERSITY**

**1445/2024**

**Project Proposal**

**Topic:** Patient Flow Model for Hospital Admission Analysis

**Chapter 1:**

**1.1: Project Overview**

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**

1. 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.
2. 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.
3. 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 Objectives**

1. 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.
2. 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.
3. 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**

**ขอบเขตของการศึกษานี้มุ่งเน้นไปที่การวิเคราะห์ข้อมูลผู้ป่วยที่เข้ารับการรักษาในโรงพยาบาล โดยใช้ข้อมูลทางการแพทย์ เช่น การวินิจฉัยโรค การเข้ารับการรักษาซ้ำ และข้อมูลประชากร เพื่อสร้างโมเดล Machine Learning ประเภท K-Nearest Neighbor (KNN) สำหรับการทำนายโอกาสที่ผู้ป่วยจะเข้ารับการรักษาอีกครั้งหรือการรักษาที่มีความยาวนาน โดยขอบเขตครอบคลุมดังนี้:**

* **แหล่งข้อมูล: ข้อมูลจากโรงพยาบาล รวมถึงรายละเอียดการเข้ารับการรักษา การวินิจฉัยโรค จำนวนครั้งที่เข้ารับการรักษา ข้อมูลประชากร ฯลฯ**
* **โมเดล Machine Learning: จะใช้ KNN ในการสร้างโมเดลคาดการณ์ และจะทดสอบเปรียบเทียบกับโมเดลอื่น ๆ เพื่อเลือกวิธีที่เหมาะสมที่สุด**
* **Data Visualization: การแสดงผลข้อมูลในรูปแบบกราฟิก เพื่อทำความเข้าใจเกี่ยวกับรูปแบบการไหลของผู้ป่วย และปัจจัยที่มีผลต่อการเกิด overcrowding**
* **การวิเคราะห์ผล: การประเมินผลการทำนายของโมเดล และการเสนอแนวทางในการปรับปรุงการจัดการทรัพยากรในโรงพยาบาล**

**1.6 Expected Output**

**จากการพัฒนาและการใช้งานโมเดล KNN รวมถึงการวิเคราะห์ข้อมูลในโครงการนี้ ผลลัพธ์ที่คาดหวังได้คือ:**

* **โมเดลทำนาย: โมเดล Machine Learning ที่สามารถทำนายการไหลของผู้ป่วย การเข้ารับการรักษาซ้ำ และความแออัดในโรงพยาบาลได้อย่างมีประสิทธิภาพ**
* **การวิเคราะห์เชิงลึก: รายงานและการแสดงภาพข้อมูลที่แสดงถึงความสัมพันธ์ระหว่างการวินิจฉัยโรคกับปัญหาความแออัดในโรงพยาบาล**
* **การปรับปรุงการจัดการ: ข้อเสนอแนะสำหรับการจัดสรรทรัพยากรที่ดีขึ้นเพื่อลดความแออัด**
* **การเพิ่มประสิทธิภาพการจัดการผู้ป่วย: การใช้ข้อมูลและ AI ช่วยให้เจ้าหน้าที่สามารถจัดการผู้ป่วยในแบบเชิงรุก**

**Chapter 2:**

**Literature Review**

**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].

**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.

**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:**

**Methodology**

**3.1 Data Collection and Understanding**

**Collect Data**

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. The specific data file is publicly available

The dataset includes:

* Patient demographics: Age and gender.
* Admission details: Type of admission (emergency, waiting list and planned), length of stay, and Time waited.
* Diagnosis: Information on primary diagnoses

**Data Understanding**

To ensure a comprehensive understanding of the dataset, interviews were conducted with hospital staff, including doctors, nurses, and administrative personnel. The objectives of these interviews were:

* To fully understand the data: This involved clarifying the meanings of the various table columns, such as how diagnosis codes are structured, what each demographic attribute signifies, and any nuances in the data collection process.
* To understand the hospital process flow: Gaining insights into the patient journey from check-in to discharge helped identify critical touchpoints and stages that could impact patient flow and contribute to overcrowding. Understanding this flow is essential for recognizing how different departments interact and the potential bottlenecks that may arise during patient admissions and treatment.

**3.2 Data Preprocessing**

Handling Missing Values

Convert the data

Handling Invalid Values

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.

LIME (Local Interpretable Model — Agnostic Explanations) LIME can be adapted for time series data by generating explanations for specific predictions. For instance, it can help explain why the ARIMA model predicts a price spike at a particular time by approximating the model’s behavior locally around that prediction.

3.1 Data Preprocessing

ก่อนที่จะนำข้อมูลมาใช้ในโมเดล จะต้องมีการเตรียมข้อมูล (data preprocessing) เพื่อให้มั่นใจว่าข้อมูลมีความถูกต้องและพร้อมสำหรับการวิเคราะห์ การเตรียมข้อมูลในขั้นตอนนี้รวมถึง:

- การจัดการกับค่าที่หายไป (Missing Values): การตรวจสอบและจัดการค่าที่ขาดหายไปใน dataset เพื่อให้มั่นใจว่าข้อมูลมีความสมบูรณ์

- การจัดรูปแบบข้อมูล (Data Transformation): การจัดรูปแบบข้อมูล เช่น การแปลง categorical data ให้เป็น numeric values โดยใช้เทคนิค one-hot encoding หรือ label encoding

- การปรับขนาดข้อมูล (Scaling): การปรับขนาดของข้อมูล เช่น การปรับขนาดฟีเจอร์ให้มีช่วงค่าที่ใกล้เคียงกันเพื่อช่วยปรับปรุงประสิทธิภาพของโมเดล

3.2 Data Visualization

Python มี libraries สำหรับการแสดงผลข้อมูลที่หลากหลาย ซึ่งจะช่วยให้สามารถค้นหาความสัมพันธ์และสื่อสารรูปแบบของข้อมูลได้อย่างมีประสิทธิภาพ ในการศึกษาและวิเคราะห์ข้อมูลนี้ จะใช้ libraries ต่อไปนี้:

- Matplotlib: ใช้สำหรับการสร้างกราฟพื้นฐาน เช่น กราฟเส้น (line plot), กราฟแท่ง (bar plot), และฮิสโตแกรม (histogram)

- Seaborn: ใช้เพื่อการสร้างกราฟที่มีการออกแบบสวยงามและสะดวกในการวิเคราะห์เชิงสถิติ เช่น heatmap และ pairplot

- Plotly และ Bokeh: สำหรับการสร้างกราฟเชิงโต้ตอบ (interactive graphs) ซึ่งช่วยให้สามารถสำรวจข้อมูลได้ง่ายขึ้น

การแสดงผลข้อมูลเหล่านี้จะช่วยให้เราเข้าใจถึงความสัมพันธ์ระหว่างการวินิจฉัยโรค (diagnosis) และความแออัดในโรงพยาบาล (overcrowding) และช่วยชี้แนะทิศทางในการพัฒนาโมเดลทำนาย

3.3 Model Development

การพัฒนาโมเดลในโปรเจคนี้จะใช้ K-Nearest Neighbor (KNN) สำหรับการทำนายช่วงเวลาที่มีแนวโน้มจะเกิด overcrowding โดยพิจารณาจากข้อมูลการวินิจฉัยโรคและจำนวนผู้ป่วยที่เข้ารับการรักษา:

- K-Nearest Neighbor (KNN): KNN เป็นโมเดลที่อาศัยหลักการใกล้เคียง (proximity) ของข้อมูลในเชิงพหุมิติ โดยการทำนายจะพิจารณาข้อมูลใกล้เคียงที่เคยเกิดขึ้นในอดีต เช่น หากมีการเข้ารับการรักษาโรคที่คล้ายกันหรือรูปแบบการเข้ารับการรักษาซ้ำในอดีตที่เคยนำไปสู่ความแออัดในโรงพยาบาล โมเดลสามารถใช้ข้อมูลนั้นในการคาดการณ์การเกิด overcrowding ในอนาคต

3.4 Model Explanation and Interpretability

เพื่อเพิ่มความเข้าใจในโมเดลและทำให้ผลลัพธ์ที่ได้สามารถตีความได้ง่าย จะใช้เครื่องมือดังต่อไปนี้:

- LIME (Local Interpretable Model-Agnostic Explanations): LIME เป็นเทคนิคที่ช่วยอธิบายการทำงานของโมเดลที่ซับซ้อนโดยการสร้างคำอธิบายเชิงพื้นที่ (local explanations) สำหรับการทำนายแต่ละครั้ง ในโครงการนี้ LIME จะถูกนำมาใช้เพื่ออธิบายผลการทำนายของ KNN และช่วยให้เราเข้าใจว่าปัจจัยใดบ้างที่มีผลต่อการคาดการณ์การเกิด overcrowding ในช่วงเวลาต่าง ๆ

การใช้ LIME จะช่วยให้สามารถอธิบายได้ว่าทำไมโมเดลถึงทำนายว่าโรงพยาบาลจะมีความแออัดในช่วงเวลาหนึ่ง ซึ่งจะทำให้ผู้ที่เกี่ยวข้องสามารถนำข้อมูลที่ได้ไปปรับปรุงการจัดการและวางแผนทรัพยากรได้อย่างมีประสิทธิภาพ

3.5 Model Evaluation

โมเดลที่พัฒนาขึ้นจะถูกประเมินโดยใช้ metrics หลายตัว เช่น:

- Accuracy: วัดความถูกต้องของการทำนาย

- Precision, Recall และ F1-score: วัดประสิทธิภาพการทำนายในกรณีที่ข้อมูลมีความไม่สมดุล

- Confusion Matrix: ใช้เพื่อดูรายละเอียดการทำนายผิดพลาดในแต่ละคลาส

นอกจากนี้ การแบ่งข้อมูลเป็น train/test split หรือ cross-validation จะถูกนำมาใช้เพื่อป้องกันการเกิด overfitting และเพื่อประเมินความสามารถของโมเดลในการทำนายข้อมูลใหม่

1.4 **Data Collection**

1.4.1 Data Source

The data for this project is sourced from the NHS Digital website, specifically from the publication titled "Hospital Admitted Patient Care Activity 2022-23." The dataset used in this project is named "Hospital Episode Statistics Admitted Patient Care Diagnosis 2022-23". This dataset comes from the Hospital Episode Statistics (HES) system, which is the primary source of data on patient care in NHS hospitals in England. It includes detailed information on diagnoses, consultancy episodes, and demographic attributes of patients.

1.4.2 Data Description

The downloaded dataset is in Excel format and includes the following key components:

- \*\*Diagnosis Codes:\*\* This data consists of codes used to identify diagnoses received during patient admissions. It helps in analyzing the prevalence of various diseases in hospitals by categorizing and studying the frequency of different conditions.

- \*\*Consultancy Episodes:\*\* Records of consultations that occurred during patient admissions. It provides details on interactions between patients and medical specialists, including dates and nature of consultations. This information aids in understanding the management and follow-up of patients during their hospital stay.

- \*\*Demographic Details:\*\* Information on patient demographics such as age, gender, and patient type (e.g., inpatient or outpatient). This data supports the analysis of patient characteristics, which is important for studying trends and developing healthcare policies.

- \*\*Admission Details:\*\* Data on the type of admission (e.g., emergency or scheduled) and the duration of the stay. This information helps analyze the length of stay and bed management in hospitals, as well as evaluate the efficiency of patient care management systems.

1.4.2 Data Description

The dataset is in Excel format and includes the following key components:

Diagnosis Codes: Information on diagnoses made during patient admissions.

Consultancy Episodes: Records of consultation episodes associated with each admission.

Demographic Details: Attributes such as age, gender, and patient type.

Admission Details: Information on the type and length of admissions.

**Chapter 4:**

Python’s extensive set of libraries enables users to craft a wide variety of visual representations that can help uncover insights and communicate data patterns effectively. The libraries for data visualization in Python will be used including:

• Matplotlib: A foundational library that provides a wide range of functionalities for creating static, animated, and interactive visualizations in Python. It is especially useful for simple plots and basic visual representations.

• Seaborn: Built on top of Matplotlib, Seaborn offers a high-level interface for drawing attractive and informative statistical graphics. It simplifies the creation of complex visualizations and enhances the aesthetics of plots.

• Plotly Express: A library that enables the creation of interactive visualizations easily. Plotly Express is particularly beneficial for exploring data dynamically and allows users to hover over points to reveal additional information.

Using these libraries, I created various types of visualizations to analyze and present our findings, including bar charts to compare patient demographics, line charts to illustrate trends across different age groups, and stacked bar charts to analyze admission types. This approach not only aids in understanding the data better but also facilitates effective communication of our results.

**Chapter 5:**

**KNN**

**Chapter 6:**

**Explainable AI**

**Chapter 7:**

**Result and Conclusion**

**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].

This visualization illustrates the top 10 diseases with the highest number of patients based on hospital admissions data.

The top ten diagnoses with the highest number of cases include Pneumonia (J18) as the most prevalent, with 455,575 cases, followed by Liveborn infants according to place of birth (Z38) with 369,378 cases. Abdominal and pelvic pain (R10) ranks third with 326,270 cases. Senile cataract (H25) and other cataract (H26) occupy the fourth and fifth positions, respectively. Other notable diseases in this top 10 list are Pain in throat and chest (R07), Malignant neoplasm of breast (C50), Unknown and unspecified causes of morbidity (R69), Other disorders of urinary system (N39), and Maternal care for other known or suspected fetal problems (O36), each with over 230,000 cases.

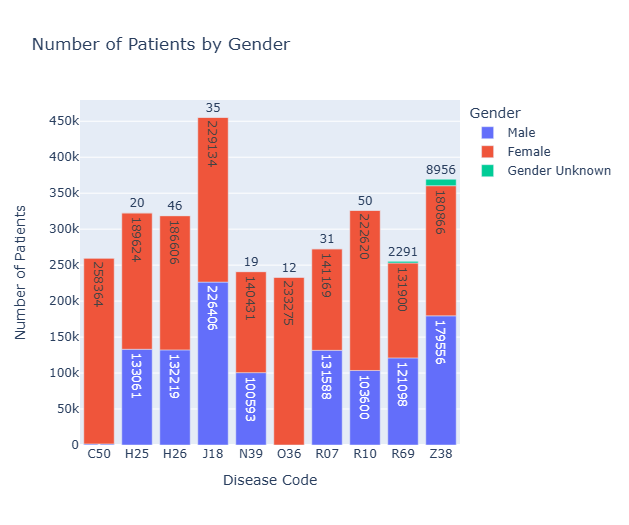


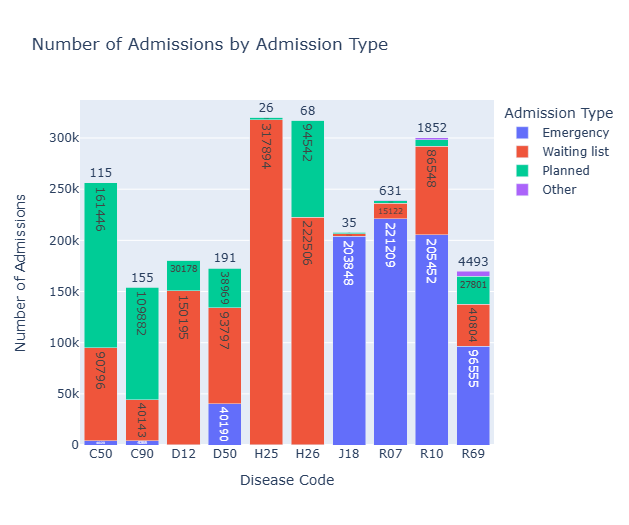
Fig. 1 Total number of patients by gender

In the (Fig. 1) shows the distribution of patients by gender according to disease codes. The Y-axis represents the total number of patients diagnosed with each disease, while the X-axis shows the disease codes. Each bar represents a specific disease code. and the color within each bar indicates the patient's gender. If male patient as blue, If it's female patient as red, If gender unknown as green.

The graph shows that females have the highest number of cases for Malignant neoplasm of breast (C50), with a total of 258,364 patients. This is followed by diseases related to pregnancy, such as Maternal care for other known or suspected fetal problems (O36) and Abdominal and pelvic pain (R10), both of which have over 200,000 female patients. Meanwhile, males have the highest number of cases for Pneumonia (J18), with 226,406 patients, followed by Senile cataract (H25) and Other cataract (H26), both of which have over 130,000 male patients.

From these results, it can be concluded that females are more likely to suffer from breast cancer and pregnancy-related conditions, while males are more prone to pneumonia and cataract-related diseases.

Fig. 2 Total Number of Admission by Admission Types



In the (Fig. 2) shows the distribution of patients by gender according to disease codes. The Y-axis represents the total number of patients diagnosed with each disease, while the X-axis shows the disease codes. Each bar represents a specific disease code. and the color within each bar indicates the patient's gender. If male patient as blue, If it's female patient as red, If gender unknown as green.

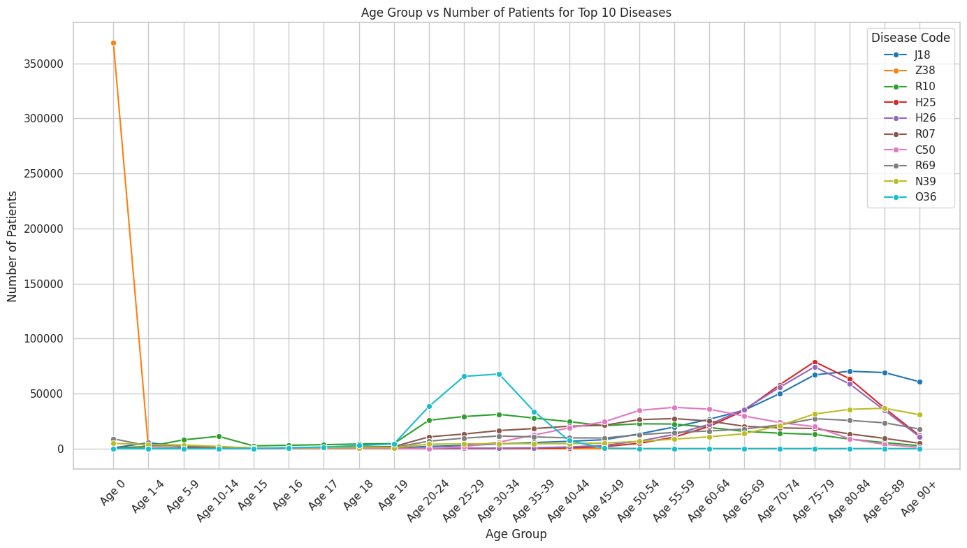


Fig. 3 Total Number of Patients by Age Group

(Fig. 3) show the differences in the distribution of various diseases across age groups. The X-axis represents the age groups of patients, ranging from newborns (Age 0) to those older than 90 years (Age 90+), while the Y-axis shows the number of patients in each age group.

From the graph, it is observed that Z38 (Liveborn infants according to place of birth) has the highest number of patients in the newborn age group (Age 0), with over 350,000 cases, and almost none in other age groups. J18 (Pneumonia, organism unspecified) sees a significant rise in the number of patients in the 75-79 and 85-89 age groups, with a slight decrease in patients aged 90+. H25 (Senile cataract) and H26 (Other cataract) have the highest number of patients in the 75-79 age group, with a decline in the 80-84 and 85-89 age groups. R10 (Abdominal and pelvic pain), R07 (Pain in throat and chest), R69 (Unknown and unspecified causes of morbidity), N39 (Other disorders of urinary system), and C50 (Malignant neoplasm of breast) are most prevalent in the 65-74 age group, particularly between 70-74 for R10, R07, R69, and N39. In age groups older than 74, the number of patients begins to decrease steadily.

This graph clearly shows that different diseases are distributed across age groups in various ways, but most notably, the number of patients tends to increase in the older adult population, especially in those aged 65 and above.

รูปภาพประกอบด้วย ภาพหน้าจอ, ข้อความ, พล็อต, ไลน์

คำอธิบายที่สร้างโดยอัตโนมัติรูปภาพประกอบด้วย ข้อความ, ภาพหน้าจอ, ไลน์, พล็อต

คำอธิบายที่สร้างโดยอัตโนมัติ