Journal of Advanced Research in Applied Sciences and Engineering Technology 56, Issue 2 (2026) 196-205

SEMARAK ILMU PUBLISHING
202103268166(003316878-P)

Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

Exploring Prediction Models for Hospital Patient Discharge Turnaround Time: A Comparative Study

Ali Abd Almisreb[1], Hajar Jaafar[2,*]

[1] Faculty of Engineering and Natural Sciences, International University of Sarajevo, Hrasnička Cesta 15, Ilidža, Bosnia & Herzegovina
[2] School of Electrical Engineering, College of Engineering, Univerisiti Teknologi MARA (UiTM), 23000 Kuala Dungun, Terengganu, Malaysia

ABSTRACT

Patient Discharge Turnaround Time has a significant impact on healthcare quality. However, improving this process has been a challenge. This study focuses on applying predictive modelling techniques to analyse patient discharge patterns in hospital. The goal is to develop a model that can predict timely patient discharge based on various features. The dataset used is derived from the Hospital Operation Management and Information System (HOMIS) of Cotabato Sanitarium and General Hospital, situated at Maguindanao, Philippines encompassing comprehensive data from admission to discharge. The process involves retrieving data from the database, followed by data cleaning and preparation to ensure quality. Feature engineering is also performed to extract additional information. Several supervised and unsupervised predictive modelling algorithms are employed, and performance metrics are used to evaluate the models. Results indicate that XGBoost achieves the highest performance, with an AUC score of 0.8207 and an accuracy rate of 0.7521. The hour of the discharge order emerges as the most significant predictor for timely discharge. This study demonstrates the application of predictive modelling in healthcare, particularly in predicting patient discharge turnaround time, contributing to existing knowledge that aims to enhance health outcomes using machine learning techniques.

Keywords:

Hospital discharge; Turnaround time; Prediction model; XGBoost

1. Introduction

In the Philippines where the population is 105 million as of 2021, there is less than one hospital bed for every 1,000 people, falling significantly below the World Health Organization's standard of three beds per 1,000 people [1]. This poses a significant challenge to the country’s healthcare system as it highlights the limited accessibility and availability of healthcare services. As a result, some hospitals have reported congestion thus worsening the burden on the health care system [2]. The country’s Department of Health set key performance indicators for government hospitals to improve hospital processes and solve issues such as this, like reducing Emergency Room discharge time to 6 hours and discharge process to 4 hours [3].

* *Corresponding author.*
E-mail address: hajarj422@uitm.edu.my

https://doi.org/10.37934/araset.56.2.196205

Reducing delays in the discharge of patients who are deemed safe to leave the hospital is of utmost importance for minimizing complications, managing costs, and enhancing overall healthcare quality [4]. Delays in the discharge process can significantly impact hospital and emergency department (ED) throughput, leading to operational inefficiencies. The complexities associated with the discharge process present specific challenges that limit the generalizability of solutions [5].

This study focuses on addressing the challenges faced by Cotabato Sanitarium, a government hospital with a capacity of 140 beds, located in the municipality of Sultan Kudarat, Maguindanao, Philippines. As a Department of Health (DOH)-retained and PhilHealth Accredited hospital, Cotabato Sanitarium is obligated to adhere to the key performance indicator set by the DOH, which requires all government hospitals to discharge patients within a 4-hour timeframe following the issuance of a discharge order by the attending physician. To expedite the discharge process and meet the 4-hour target, Cotabato Sanitarium implemented an online clearance system to facilitate data gathering and reduce turnaround time associated with clearance procedures. However, the desired outcomes have not been fully achieved despite these efforts.

Therefore, the primary objective of this study is to develop a prediction model utilizing various predictive modelling techniques that aims to identify patients who are likely to exceed the 4-hour turnaround time, enabling the integration of the model into the hospital information system. By doing so, the study aims to provide valuable insights that can be used to implement necessary interventions during the discharge process, with the goal of improving patient outcomes.

By exploring different predictive modelling approaches, this research seeks to contribute to the existing body of knowledge on predicting patient discharge turnaround time. The findings of this study hold potential implications for improving hospital efficiency, patient flow, and resource allocation, ultimately leading to enhanced healthcare delivery.

2. Methodology

Data for this research study was obtained from the Hospital Operation Management and Information System (HOMIS) of Cotabato Sanitarium. As the online clearance system was implemented in August 2020, the data collected from 2021 and 2022 were utilized for the purpose of this study. HOMIS captures a wide range of information, including patient demographics, medical records, and details related to the patient's hospital stay.

To begin the data analysis process, data retrieval was performed by extracting relevant information from the HOMIS database. Subsequently, the extracted data from the Hospital Operation Management and Information System (HOMIS) of Cotabato Sanitarium and General Hospital underwent a thorough examination to determine its validity and suitability for processing in machine learning models, ensuring that only relevant and reliable data were used for predictive modelling. Data cleaning and preparation procedures were diligently performed to address any data errors, duplications, nomenclature issues, entry errors, documentation errors, and missing data values. The goal was to refine the dataset and ensure its readiness for analysis. Following data examination and cleaning, categorization techniques were applied to better comprehend the roles of different variables within the predictive modelling framework. The data were examined to identify equivalent categories that pertain to specific variables, considering the nature of the data and its potential impact on the prediction process. If any standard guidelines or established frameworks existed for categorization, they were thoughtfully applied to enhance consistency and comparability across the dataset.

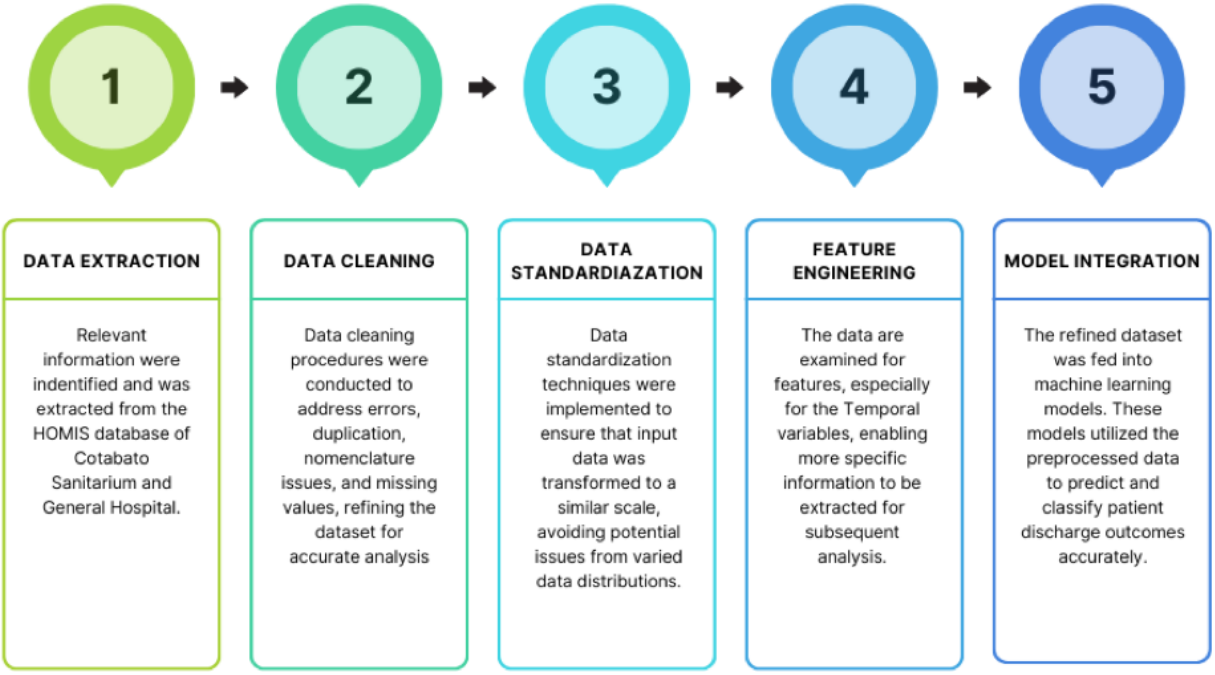

Fig. 1. Process flow of the research methodology

To address the lack of established predictive models specifically designed for predicting hospital turnaround time, various prediction modelling approaches were employed in this exploration. Both supervised and unsupervised models were utilized, drawing inspiration from churn modelling concepts used in the telecom sector. Churn modelling encompasses classification and clustering techniques to identify factors contributing to customer churn. In the context of this research study, the term "churning" refers to the timely discharge of patients, considering independent variables that may influence their discharge patterns. The collected data were subjected to analysis using various predictive models.

By leveraging a diverse range of prediction modelling techniques and carefully analysing the data, this study aims to contribute to the existing knowledge in the field and provide insights that can inform interventions to improve patient discharge processes.

3. Results

3.1 Data Preparation

The hospital database of Cotabato Sanitarium and General Hospital's information system is a vast, intricate, and interconnected relational database that consists of a substantial number of patient admissions, totalling 4,396 recorded during the years of 2021 and 2022. It collects comprehensive data encompassing not only patient admissions, but discharges, and all relevant information throughout the entire hospital visits too. Certain data entries are derived from existing data, providing additional insights into specific aspects of the patients' visit. These derived data entries are represented as row entries, expanding beyond simple data points. Examples of such derived data include the existence of patient procedures, the patient's need to apply for a birth certificate, the engagement in blood service-related activities, and the total bill amount with deductions.

Following the extraction of the data, a meticulous analysis was conducted to identify and rectify any potential data errors. Common types of errors observed included data duplication errors, nomenclature errors, data entry errors, documentation errors, and missing data values [6]. A comprehensive analysis of these data issues can be found in Table 1. Missing values were frequently encountered during data analysis, often indicating a negative response to the corresponding data entry. Columns such as "Existence of Birth Certificate," "Requires Blood Service-Related Activities," "Philhealth Insurance Beneficiary," "No Balance Billing Beneficiary," "Financial Capacity Index," and "With Enough Benefits" exhibited missing data values. These missing values were treated as indicators of absence of information and were processed through constant imputation.

The "Diagnosis Code" column presented challenges due to nomenclature errors, data entry errors, and documentation inconsistencies. To address this issue, strategies for handling inconsistent and invalid values were employed [6]. Additionally, since the data primarily consisted of Relative Value Scale or RVS codes rather than International Classification of Diseases (ICD) codes, known erroneous entries were converted to their corresponding ICD codes. This coding system is more reliable to understand the medical cases of the patients. Further data preparation methods were applied to certain variables to facilitate processing by the models. Categorical data, such as "Financial Capacity Index," "No Balance Billing Beneficiary," "Philhealth Insurance Beneficiary," and "With Enough Benefits," underwent data standardization [6]. This step was necessary because learning algorithms is sensitive to the scale of input data and may yield suboptimal performance if the features are not on a similar scale. Other variables that required similar data preparation included "Type of Service," "Diagnosis Category," "Sex," "Date/Time of Doctor's Order," "Date/Time of Admission," "Age," and "Total Bill."

All data were categorized based on the specific aspects they represented, facilitating a clear understanding of their relevance to the patient discharge process. Medical category contains patient data related to medical information gathered during their admission. Temporal category included variables such as the date and time of admission and the doctor's order for discharge. Financial category comprised data pertaining to the financial aspect of the patient's visit. Lastly, the social demographic category encompassed variables such as the age and sex of the patient.

Table 1
Summary of the dataset's cleaning process based on the errors observed by different data cleaning methods [6,7]

Description	Errors	Data Cleaning Method [6]
MEDICAL		
Type of Service	Duplicate Occurrences	Data Standardization
Procedure Code	High Frequency	Data Standardization
Diagnosis Code	Nomenclature Error, Data Entry, Documentation Error	Handling inconsistent data, Invalid Values
Diagnosis Category		Data Standardization
Has Birth Certificate	Missing Values	Constant Imputation
Has Blood Transfusion	Missing Values	Constant Imputation
*Length of Stay		Derived Value
TEMPORAL		
Date/Time of Patient Admission (Numeric)		Data Standardization
Date/Time of May Go Home Order		Data Standardization
*Time (Numeric)		Derived Value
*Hour		Derived Value
*Day of the Week		Derived Value
Admission Daily Volume		Derived Value
ER Daily Volume		Derived Value
OPD Daily Volume		Derived Value
*Office Hour		Derived Value
FINANCIAL		

Financial Capacity Index	Missing Values	Constant Imputation
MSS Discount Amount	Missing Values	Constant Imputation
No Balance Billing	Missing Values	Encoding Categorical Values, Constant Imputation
Is PhilHealth Insurance Beneficiary	Missing Values	Encoding Categorical Values, Constant Imputation
Total Bill		
Has Enough Benefits	Encoding Categorical Values, Constant Imputation	Encoding Categorical Values, Constant Imputation
Payment Total		
Has Discount	Missing Values	Encoding Categorical Values, Constant Imputation
SOCIAL DEMOGRAPHIC		
Patient Age		Data Standardization
Patient Sex		

In this research study, the predictor variable of the dataset was derived by defining the target as patient hospital discharge time, which should be less than 4 hours. To obtain this variable, the date and time of the doctor's may-go-home order were extracted and the time difference was calculated. If the discharge time was less than four hours, it was labelled as "Timely Discharged", indicating that the patient's discharge occurred within the desired timeframe and met the criteria for timely discharge. Conversely, if the discharge time exceeded the four-hour threshold, it was labelled as "Untimely Discharged," signifying delays in the discharge process beyond the expected timeframe. These labels effectively distinguish between patients who were discharged in a timely manner and those who experienced delays in their discharge process.

After categorizing the discharge instances according to the defined criteria, it was observed that 1,845 or 37.38% of the cases were classified as 'timely discharged', indicating that these patients were successfully discharged within the desired timeframe. Conversely, 2,551 or 62.62% of the cases fell under the 'untimely discharged' category, suggesting that these patients experienced delays in their discharge processes. The distribution of instances across these categories is depicted in Figure 2, providing a visual representation of the cases' proportion in each category and the comparison of total cases between both categories was illustrated in Figure 3. In both figures, 'timely discharged' was labelled as [1] while 'untimely discharged' was labelled as [0].

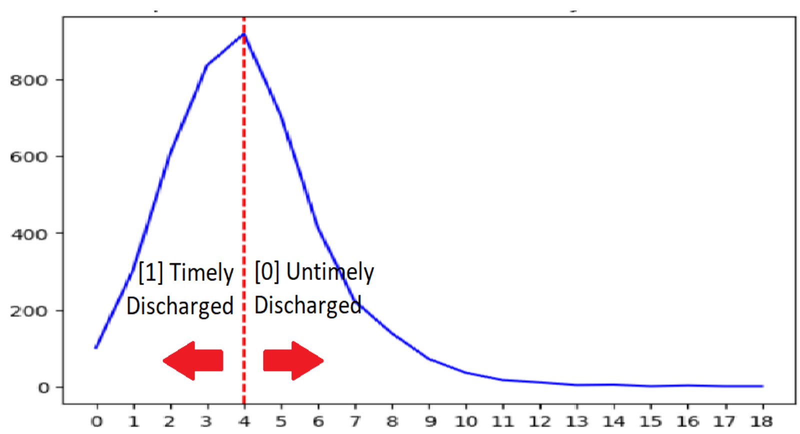

Fig. 2. This graph represents the distribution of patient discharge duration categorized for timely discharged and untimely discharged

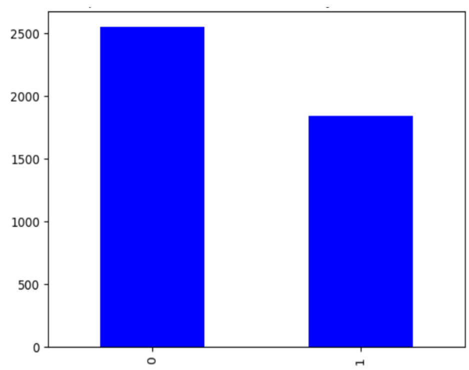

Fig. 3. Graph comparison of the Timely Discharged [1] and Untimely [0] Discharged

3.2 Prediction Model

In this study, we utilized various common supervised and unsupervised prediction modelling algorithms:

i. Naive Bayes is a classification algorithm known for its high bias and low variance, which allows it to build effective models even with limited data [8]. This makes it a valuable choice when dealing with small datasets, and it is particularly useful in text classification tasks and spam filtering.
ii. Logistic regression is a popular regression method used to examine the relationship between a binary or categorical outcome and multiple influencing factors [9]. It is commonly utilized in various fields, including medicine, economics, and social sciences, for its simplicity and interpretability.
iii. Generalized linear models (GLMs) extend traditional linear models and offer various advantages, including parallel computation, rapid model fitting, and scalability, particularly for models with a limited number of predictors with non-zero coefficients [8]. GLMs are widely used for various types of data, such as count data, binary data, and continuous data.
iv. Deep learning is a powerful approach based on multi-layer feed-forward artificial neural networks, trained using stochastic gradient descent and back-propagation [8]. This technique is highly effective in handling complex data, especially images, and speech, and has achieved remarkable success in various applications, including image recognition and natural language processing.
v. Random forests are ensemble learning methods that combine multiple tree predictors, with each tree depending on a random vector sampled independently and from the same distribution as other trees in the forest [10]. This approach is widely used in both classification and regression tasks, and it excels in handling high-dimensional data.
vi. The Fast Large Margin operator employs a rapid margin learner based on the linear support vector learning scheme [8]. This operator is usually used in large-scale classification problems, where it efficiently determines the margin between different classes.
vii. A gradient-boosted model is an ensemble of regression or classification tree models, where boosting, as a flexible nonlinear regression technique, enhances the accuracy of

individual trees [8]. Gradient boosting is widely applied in various domains due to its robustness and ability to handle complex data patterns.

viii. XGBoost is a scalable tree-boosting system widely used by data scientists to achieve state-of-the-art results in various machine learning challenges. It incorporates innovations such as a novel tree learning algorithm for handling sparse data and a weighted quantile sketch procedure for approximate tree learning with instance weights [11].

The dataset was randomly split into a training dataset (70%) and a test dataset (30%). The column "Timely Discharged" was used as the predictor variable in this model. The prediction models were applied to the dataset using the respective algorithms. Performance metrics such as accuracy rate, precision rate, recall, f1-score, and AUC-ROC were computed. The results of the prediction are summarized in Table 2, which provides an overview of the performance of the different prediction models in predicting the timely discharge of hospital patients based on the time of their discharge order.

Table 2
Summary of the prediction model results

Model	Accuracy	AUC	F1-Score	Precision	Recall	Specificity
Naive Bayes	0.673566	0.715254	0.719093	0.728366	0.710313	0.618966
Generalized Linear Model	0.730889	0.799769	0.764861	0.756859	0.775158	0.67253
Logistic Regression	0.414017	0.782804			0	1
Fast Large Margin	0.47453	0.584832	0.260321	0.642951	0.163712	0.880868
Deep Learning	0.731683	0.777342	0.791679	0.719684	0.880711	0.521851
Decision Tree	0.648097	0.69603	0.693149	0.698312	0.690776	0.585239
Random Forest	0.632182	0.733787	0.623834	0.754655	0.533342	0.763221
Gradient Boosted Trees	0.742816	0.790384	0.790804	0.743247	0.845305	0.600133
XGBoost	0.7521	0.8207	0.681	0.7914	0.5976	0.8748

Data in Table 2 were then used to create bar charts as a visual presentation for all 8 prediction model results as can be seen in Figure 4. Each bar chart represents each performance metrics thus depicting six different charts altogether.

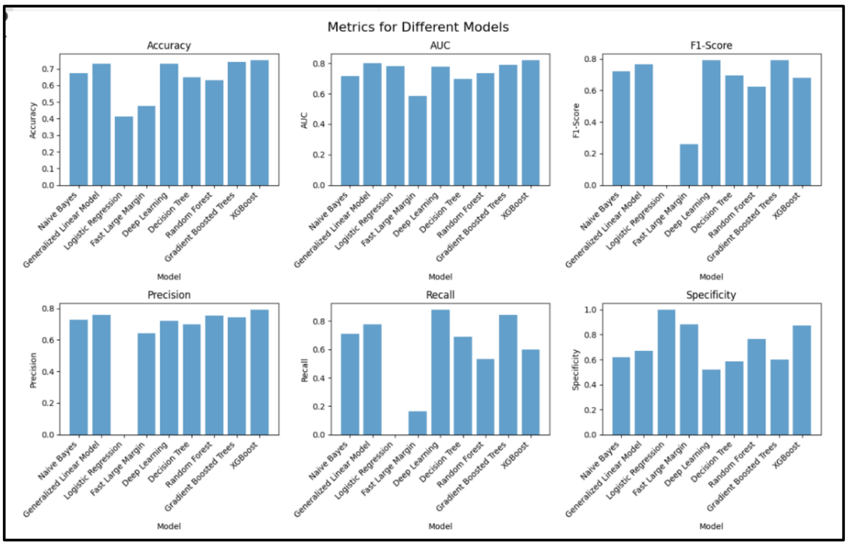

Fig. 4. Visual presentation of the output for different models at prediction of Discharge Time Turn-Around Time

4. Discussion

The dataset used in this prediction model is based on the data acquired from a level-1 140-bed capacity government hospital in the Philippines utilizing a wide range of complex financial, medical, temporal, and social demographic information of the patients to develop this prediction model. Given the complexity of the data, it is believed that an unsupervised model tree-based approach will perform well in this model. Obtained outputs and evaluation of the performance using appropriate performance indicators validated this idea.

The AUC (Area Under the Curve) score is a reliable indicator of a prediction model's performance in distinguishing between positive and negative occurrences [12]. Among the models tested, XGBoost achieved the highest AUC score of 0.8207, indicating its superior ability in classification. Additionally, XGBoost also exhibited the highest accuracy rate among all the models, with a value of 0.7521. The top three features with the highest weights in the XGBoost model were identified as follows: the hour of discharge order, with a weight of 272.0; the OPD Daily Volume, with a weight of 242; and the Length of Stay, with a weight of 196.0. These features played a significant role in the classification process within the XGBoost model.

Interestingly, the hour of discharge order emerged as a prominent attribute for classification across multiple models, including Naïve Bayes, Logistic Regression, Deep Learning, and Gradient Boosting. Conversely, the Length of Stay stood out as the primary identifier in the Random Forest and Decision Tree models, highlighting its importance in these models.

These three features with high importance in the prediction model provide valuable insights into the prediction process. The hour of discharge order represents the timing aspect, reflecting the efficiency of the discharge process. Additionally, the OPD Daily volume, which indicates the volume of patients in the out-patient department, serves as a factor that highlights potential process issues, as each hospital department handles all patients collectively for the OPD, ER, and admitted patients [15]. On the other hand, the Length of Stay signifies the duration of a patient's hospitalization, providing insights into the severity of illness or the complexity of treatment. The significance of these features in different models indicates their substantial impact on predicting patient discharge turnaround time, as captured by the models.

Predictive modelling techniques, like those employed in churn models [19], have the potential to predict factors that influence patient discharge time. By utilizing these techniques, hospitals can enhance their discharge processes [5], embracing a proactive approach to patient care, mitigating the risks of delayed discharges, and optimizing bed utilization to reduce unnecessary hospital stays. The integration of the predictive model into the hospital information system fosters a culture of data-driven decision-making, empowering hospital personnel to access timely and accurate predictions upon the discharge order. This Artificial Intelligent (AI)-driven approach enables the system to pre-emptively address potential discharge delays before they exceed the target timeframe, leading to more efficient and streamlined patient discharges. With strong leadership commitment, support, and front-line engagement [5,17], hospitals can leverage predictive modelling to identify patients at high risk of delayed discharge. This proactive approach enables hospitals to address potential issues and minimize discharge time, ultimately optimizing the discharge process, reducing unnecessary hospital stays, and improving overall healthcare quality [5,17,18].

5. Conclusion

In conclusion, this study has demonstrated the practical application of predictive models in the context of hospital patient discharge. By utilizing predictive modelling techniques, patients who were more likely to experience untimely discharge after a discharge order has been carried out were able to identified. This paper findings indicate that XGBoost exhibited the highest performance among the evaluated models, with an AUC of 0.8207 and an accuracy rate of 0.7521. The time of the discharge order was identified as the most significant predictor for classifying patients as likely or unlikely to be discharged on time. Additionally, the OPD Daily volume and the patients' length of stay were found to be important factors attributing to timely discharge of the patients. To the best of the author's knowledge, this study represents a novel contribution to the existing body of literature as it is the first to apply predictive modelling techniques specifically aimed at predicting timely patient discharge turnaround time. By employing the ability of machine learning, the comprehension of the elements affecting discharge efficiency has advanced. However, the study has its limitations as the analysis was conducted using a dataset spanning only two years from a level-1 hospital with a capacity of 140 beds. Future research paths provide potential for validating and expanding on these results. Incorporating larger datasets from diverse hospital settings with varying bed capacities and demographics could enhance the generalizability and model performance. By encompassing a broader range of features, such as patient demographics, medical history, and hospital resources, we can achieve improved model stability and better future representation, ultimately leading to enhanced model generalization. Expanding the scope of the dataset to include a more diverse and extensive range of healthcare facilities across different levels of care and locations will provide a comprehensive understanding of patient discharge patterns in the Philippines. This broader perspective might shed light on common discharge practices and challenges, helping the Department of Health formulate policies and regulations that apply not only to level-1 hospitals but also to all healthcare facilities, thereby improving healthcare delivery and patient outcomes nationwide. By addressing the identified limitations and embracing these potential future research directions, we can continue to advance the field of machine learning in healthcare, ultimately leading to more efficient and effective health outcomes. These endeavours hold the promise of transforming patient care and hospital management, contributing to better healthcare delivery and improved patient experiences across the healthcare system.

Acknowledgement

The authors would like to thank the contributing researchers from Faculty of Engineering and Natural Sciences, International University of Sarajevo Sarajevo, Bosnia and Herzegovina and School of Electrical Engineering, College of Engineering, Univerisiti Teknologi MARA Cawangan Terengganu, Malaysia. This research was funded by a grant from Universiti Teknologi MARA, (600-RMC/KEPU 5/3 (004/2023).

References

[1] BusinessWorld Online. "Burden on hospitals can be eased by community-based care, say experts." (2021). https://www.bworldonline.com/health/2021/06/23/377332

[2] Toledo, R. C. "Overcrowding, long lines plague Manila City hospitals." *BusinessMirror*, (2017). https://businessmirror.com.ph/2017/09/14

[3] Department of Health. "Memorandum Order Number 2018-0081 [Office Performance and Commitment Review for Government Hospitals and Sanitaria]." (2018).

[4] Ragavan, Meera V., David Svec, and Lisa Shieh. "Barriers to timely discharge from the general medicine service at an academic teaching hospital." *Postgraduate medical journal* 93, no. 1103 (2017): 528-533. https://doi.org/10.1136/postgradmedj-2016-134529

[5] El-Eid, Ghada R., Roland Kaddoum, Hani Tamim, and Eveline A. Hitti. "Improving hospital discharge time: a successful implementation of six sigma methodology." *Medicine* 94, no. 12 (2015): e633. https://doi.org/10.1097/MD.0000000000000633

[6] Chapman, Arthur D. *Principles and methods of data cleaning*. GBIF, 2005.

[7] Brownlee, Jason. *Data preparation for machine learning: data cleaning, feature selection, and data transforms in Python*. Machine Learning Mastery, 2020.
[8] RapidMiner. “Home - RapidMiner Documentation.” https://docs.rapidminer.com
[9] Wang, Q. Q., S. C. Yu, Xiao Qi, Y. H. Hu, W. J. Zheng, J. X. Shi, and H. Y. Yao. "Overview of logistic regression model analysis and application." *Zhonghua yu fang yi xue za zhi [Chinese journal of preventive medicine]* 53, no. 9 (2019): 955-960.
[10] Breiman, Leo. "Random forests." *Machine learning* 45 (2001): 5-32. https://doi.org/10.1023/A:1010933404324
[11] Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785-794. 2016. https://doi.org/10.1145/2939672.2939785
[12] Huang, Jin, and Charles X. Ling. "Using AUC and accuracy in evaluating learning algorithms." *IEEE Transactions on knowledge and Data Engineering* 17, no. 3 (2005): 299-310. https://doi.org/10.1109/TKDE.2005.50
[13] Ullah, Irfan, Basit Raza, Ahmad Kamran Malik, Muhammad Imran, Saif Ul Islam, and Sung Won Kim. "A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in telecom sector." *IEEE access* 7 (2019): 60134-60149. https://doi.org/10.1109/ACCESS.2019.2914999
[14] Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785-794. 2016. https://doi.org/10.1145/2939672.2939785
[15] Cotabato Sanitarium. “Quality Manual.” *Cotabato Sanitarium*, (2015).
[16] Pirooznia, Mehdi, Jack Y. Yang, Mary Qu Yang, and Youping Deng. "A comparative study of different machine learning methods on microarray gene expression data." *BMC genomics* 9 (2008): 1-13. https://doi.org/10.1186/1471-2164-9-S1-S13
[17] Bains, Amarpreet S., Erwin Wang, Deserie Duran, Lorna Lee-Riley, and Frank Volpicelli. "Maintaining throughput and reducing discharge delays after increasing capacity during the Covid-19 pandemic: A New York City hospital’s experience." *NEJM Catalyst Innovations in Care Delivery* 1, no. 6 (2020). https://doi.org/10.1056/CAT.19.1111
[18] Abuzied, Yacoub, Hassan Maymani, Basim AlMatouq, and Oweida AlDosary. "Reducing the length of stay by enhancing the patient discharge process: using quality improvement tools to optimize Hospital efficiency." *Global Journal on Quality and Safety in Healthcare* 4, no. 1 (2021): 44-49. https://doi.org/10.36401/JQSH-20-27
[19] Ajit, Pankaj. "Prediction of employee turnover in organizations using machine learning algorithms." *algorithms* 4, no. 5 (2016): C5. https://doi.org/10.14569/IJARAI.2016.050904
[20] Sharma, Tanu, Prachi Gupta, Veni Nigam, and Mohit Goel. "Customer churn prediction in telecommunications using gradient boosted trees." In *International Conference on Innovative Computing and Communications: Proceedings of ICICC 2019, Volume 2*, pp. 235-246. Springer Singapore, 2020. https://doi.org/10.1007/978-981-15-0324-5_20
[21] Ranapurwala, Shabbar I., Joseph E. Cavanaugh, Tracy Young, Hongqian Wu, Corinne Peek-Asa, and Marizen R. Ramirez. "Public health application of predictive modeling: an example from farm vehicle crashes." *Injury epidemiology* 6 (2019): 1-11. https://doi.org/10.1186/s40621-019-0208-9
[22] Yang, Christopher C. "Explainable artificial intelligence for predictive modeling in healthcare." *Journal of healthcare informatics research* 6, no. 2 (2022): 228-239. https://doi.org/10.1007/s41666-022-00114-1
[23] Bonnett, Laura J., Kym IE Snell, Gary S. Collins, and Richard D. Riley. "Guide to presenting clinical prediction models for use in clinical settings." *bmj* 365 (2019). https://doi.org/10.1136/bmj.l737
[24] Ibrahim, Walaa Hussein, Ahmed AbdelRhman Ahmed Osman, and Yusra Ibrahim Mohamed. "MRI brain image classification using neural networks." In *2013 international conference on computing, electrical and electronic engineering (ICCEEE)*, pp. 253-258. IEEE, 2013. https://doi.org/10.1109/ICCEEE.2013.6633943
[25] Esa, Muneera, Praveena Muniandy, and Nurul Ain Syahirah Binti Mohd Noor. "Awareness of Mental Health Issues in Malaysian Construction Industry." *Journal of Health and Quality of Life* 1, no. 1 (2024): 21-48. https://doi.org/10.37934/hqol.1.1.2148

Name of Author	Email
Ali Abd Almisreb	alimes96@yahoo.com
Hajar Jaafar	hajarj422@uitm.edu.my