Citation:
Elbatanouny, H and Tawfik, H and Khater, T and Gorbenko, A (2025) An interpretable machine learning model to predict hospitalizations. Clinical eHealth, 8. pp. 53-65. ISSN 2588-9141 DOI: https://doi.org/10.1016/j.ceh.2025.03.004

Link to Leeds Beckett Repository record:
https://eprints.leedsbeckett.ac.uk/id/eprint/12069/

Document Version:
Article (Published Version)

Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0

© 2025 The Authors

The aim of the Leeds Beckett Repository is to provide open access to our research, as required by funder policies and permitted by publishers and copyright law.

The Leeds Beckett repository holds a wide range of publications, each of which has been checked for copyright and the relevant embargo period has been applied by the Research Services team.

We operate on a standard take-down policy. If you are the author or publisher of an output and you would like it removed from the repository, please contact us and we will investigate on a case-by-case basis.

Each thesis in the repository has been cleared where necessary by the author for third party copyright. If you would like a thesis to be removed from the repository or believe there is an issue with copyright, please contact us on openaccess@leedsbeckett.ac.uk and we will investigate on a case-by-case basis.

Contents lists available at ScienceDirect

Clinical eHealth

journal homepage: ww.keaipublishing.com/CEH

An interpretable machine learning model to predict hospitalizations

Hagar Elbatanouny [a], Hissam Tawfik [a,b,*], Tarek Khater [c], Anatoliy Gorbenko [b]

[a] Department of Electrical Engineering, University of Sharjah, Sharjah, the United Arab Emirates
[b] School of Built Environment, Engineering and Computing, Leeds Beckett University, Leeds, UK
[c] Department of Biomedical Engineering, Khalifa University, Abu Dhabi, the United Arab Emirates

ARTICLE INFO

Article history:
Received 1 August 2024
Accepted 30 March 2025
Available online 4 April 2025

Keywords:
Hospital management
Pandemics
Machine learning
Explainable AI
COVID-19
Interpretable models

ABSTRACT

Hospital management plays a pivotal role in ensuring the efficient delivery of medical services, especially in the face of challenges posed by pandemics such as COVID-19. This paper explores the application of machine learning techniques in addressing the challenge of hospitalization during pandemics. Leveraging a comprehensive dataset sourced from the Mexican government, various supervised learning algorithms including Random Forest, Gradient Boosting, Support Vector Machine, K-Nearest Neighbors, and Multilayer Perceptron are trained and evaluated to discern factors contributing to hospitalizations. Feature importance analysis and dimensionality reduction techniques are employed to enhance models predictive performance. The best model was Gradient Boosting algorithm with an accuracy of 85.63% and AUC score of 0.8696. The interpretability plots showed that pneumonia had a positive impact on the hospitalization prediction of the model. Our analysis indicates that women aged over 45 with pneumonia and concurrent COVID-19 exhibit the highest likelihood of hospitalization. This study underscores the potential of interpretable machine learning in aiding hospital managers to optimize resource allocation, hospitalization cases, and make data-driven decisions during pandemics.
© 2025 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Hospitals are complicated establishments that require optimal performance in terms of service quality, service time, cost-effective rates, use of supplies, and overall service outcome.[1,2] Maintaining effective and organized operations, resolving problems, and maximizing resource utilization are all made possible by professional hospital management. It is critical to provide patients with practical and proactive services as well as to raising the standard of hospital care.[2] For better patient care quality and compliance, efficiency, minimizing mistakes, timeliness and outcome, and data privacy, management tools for hospitals are essential.[1,2]

The management of hospitals has faced many difficulties as a result of pandemics, including managing financial and resource allocation, employees and scheduling, and providing mental support to healthcare professionals.[3] To manage emergencies and decrease the effects of the pandemic on the community, hospital managers have to create elaborating plans for the epidemic.[3] In order to manage patients and stop the spread of diseases, pandemics have also brought attention to how crucial it is for hospitals to be adaptable and cooperative when collaborating with primary care.[4] Managers at hospitals have had to deal with the nursing staff shortage, create crisis management strategies, and give healthcare professionals organizational support.[5,6] Pandemics have further highlighted the need for easily accessible, precise, and effective diagnostic tools, treatment modalities, and preventive measures.[3]

Promising answers to the problems hospital managers encounter especially during pandemics can be found in machine learning (ML) techniques.[7] Hospitals that want to effectively manage a crisis can improve patient flow, predict the spread of disease, allocate resources efficiently, and make data-driven decisions through the use of Artificial Intelligence (AI) and ML.[8] In order to find trends and risk factors related to COVID-19 and other infectious conditions, machine learning algorithms are able to evaluate enormous amounts of patient data, including population demographics, medical histories, and clinical records.[9] These algorithms can help anticipate hospitalization rates, spot high-risk individuals who might benefit from prompt intervention, and maximize bed usage to handle the spike in patient numbers.[7,8] In order to ensure sufficient availability of crucial resources, ML models can also be used to develop forecasting models for medical supply chain management.[7,10] Additionally, by supporting staff scheduling and resource allocation decision-making processes, machine learning

* Corresponding author at: Department of Electrical Engineering, University of Sharjah, Sharjah, the United Arab Emirates.

E-mail addresses: U22102889@sharjah.ac.ae (H. Elbatanouny), htawfik@sharjah.ac.ae, h.tawfik@leedsbeckett.ac.uk (H. Tawfik), 100065162@ku.ac.ae (T. Khater), A.Gorbenko@leedsbeckett.ac.uk (A. Gorbenko).

https://doi.org/10.1016/j.ceh.2025.03.004
2588-9141/© 2025 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

algorithms can assist hospitals in making well-informed decisions based on real-time data.[8,9]

As an example, the COVID-19 pandemic has had a disastrous worldwide impact, having a severe influence on people's long-term health and livelihood. With a 7 million death toll, healthcare systems around the globe were compromised.[11] Never before had modern healthcare settings experienced such an influx for hospital beds, including ICU beds and intubation devices. In the face of this unprecedented crisis, it has been crucial to examine the factors that contribute to the severity of COVID-19 cases. The most severe cases typically end in death, but the moderately severe cases often result in hospitalization, which could last up to a month depending on the severity.[12]

This work endeavors to use AI technology to support the clinical decision-making process, aiding medical facilities in effectively preparing hospital beds and providing appropriate care for their patients. The research entails training various machine learning models, encompassing Random Forest (RF), Gradient Boosting (GB), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Multilayer Perceptron (MLP). Through an examination of patient data, including medical histories, the study aims to discern the extent to which specific pre-existing conditions or other attributes significantly elevate the risk of hospitalization using The COVID-19 Mexico Patient Health Dataset (Covid19MPD) as a case study.[13] This dataset was chosen because it includes patients medical history and several underlying conditions such as Pneumonia and Chronic Obstructive Pulmonary Disease (copd), contains over 97,000 instances, and has a hospitalization feature, making it suitable for the case study to predict hospitalization during pandemics. Additionally, the research employs several Explainable Artificial Intelligence (XAI) techniques to enhance the model's interpretability and explainability, facilitating a deeper understanding of the underlying factors influencing hospitalization outcomes.

The rest of the paper is structured as follows. Section 2 describes related work and a literature review. Section 3 describes the dataset utilized, examines the methodology used in this work, while the results and discussion are provided in Section 4. Section 5 describes the interpretability and add another layer of explainability to the models. Section 6 describes the current limitations while Section 7 concludes the paper highlighting the main findings and future work.

2. Literature review

Recently, hospital managers and healthcare providers have been increasingly using machine learning techniques to analyze real-time data for decision-making.[14] Salcedo et al.[15] evaluated the utilization of machine learning algorithms in the context of COVID-19. They examined 925 papers published between 2019 and 2022, culminating in the selection of 32 publications for further analysis. Their investigation revealed the diverse applications of machine learning algorithms in pandemic management efforts, including the prediction of in–hospital mortality among COVID-19 patients, the development of machine learning-driven models for predicting near-term in–hospital mortality in COVID-19 patients, and the assessment of the risk of community deaths associated with COVID-19 using various machine learning algorithms.[15]

To assist healthcare professionals in making decisions, De Holanda et al.[16] created two predictive machine learning models to identify COVID-19 patients at higher risk of hospitalization or death. Out of the fourteen machine learning algorithms that were tested, the gradient boosting model performed the best, predicting death with 83% accuracy and an Area Under the Curve (AUC) of 0.89, and hospitalization with 71% accuracy and an AUC of 0.75.

Brnabic and Hess[17] carried out an extensive study into the methods and approaches used when applying machine learning to observational data to inform patient-provider decision making. In order to guarantee that decisions about patient care are supported by solid evidence, the study made clear how important it is to use a range of machine learning techniques, clearly define model selection strategies, and include both internal and external validation procedures. Random forest and decision tree (DT) approaches were the most often used techniques in these investigations. It is interesting to note that most of the studies that were analyzed relied on a single algorithm, and very few chose to use multiple machine learning algorithms in their analyses.[17]

Weissman et al.[18] attempt to predict when to expect an increase in clinical demand and present the best and worst-case scenarios for the local COVID-19-induced demand on hospital capacity. The developed modeling tool can help in the early stages of a pandemic, guide clinical operations and staffing demands, predict when hospital capacity may become overcrowded, and inform capacity strain preparations. Using a queuing model, the study offers a novel method for ventilator capacity planning during the early phases of the COVID-19 pandemic. This tactic was used to demonstrate how public health campaigns and social distancing could potentially prevent up to 50 deaths per day in British Columbia, Canada. The authors used the COVID-19 Hospital Impact Model for Epidemics (CHIME), which is based on the susceptible, infected, removed (SIR) model. The model's primary findings are the total counts of expected demand for hospital beds, ICU beds, and ventilators over time at three different levels of care. Their study predicts hospital bed capacity during the COVID-19 pandemic, providing vital information for effective planning during epidemics. CHIME estimated that it would take a maximum of 31 to 53 days for the number of patients with COVID-19 diagnoses to exceed the hospital's current capacity. The total number of beds that would be needed in three hospitals to accommodate a surge of COVID-19 patients was estimated to be between 3,131 and 12,650 in both optimistic and pessimistic scenarios. This included an estimated requirement for 118 to 599 ventilators and 338 to 1,608 ICU beds.[18]

Eight Ochsner Health hospitals used a SIR model created by Fort et al.[19] that was based on the University of Pennsylvania Model for Epidemics between March 16 and April 15, 2020. The model used ICU admissions of cases to estimate community case loads, especially in cases of delayed evaluation. Surprisingly, by April 6, 2020, the model accurately predicted the peak utilization of hospital beds ($n = 487$) and intensive care units ($n = 250$). Ochsner Health actively distributed 130 intensive care unit beds among its hospitals based on the patterns the model identified. This necessitated building more intensive care units (ICUs) and converting some hospital emergency rooms and surgical rooms into ICU beds. The model demonstrated the importance of hospital admission data and offered useful information about the prevalence of COVID-19 transmission within an area in situations where disease monitoring is limited or results are delayed. Different basic reproduction number (R0) plots in their model represented a range of scenarios, which aid in the planning of hospital managers and healthcare providers.[19]

Al Meslamani et al.[20] investigated the possible benefits of using machine learning in pandemic healthcare responses. These benefits included the ability to predict the spread of disease, identify individuals who are more susceptible to it, and optimize the allocation of resources. Machine learning can help policymakers make decisions about how best to use resources, which could ultimately lead to resource savings, reduced workloads for medical professionals, improved patient outcomes, and improved mental health. However, the authors draw attention to the paucity of studies on how machine learning can improve pandemic healthcare responses in practice.[20]

The Machine Learning-based Hospitalization Capacity Planning System for COVID-19 (CPAS) in the UK is a noteworthy initiative that aims to provide various stakeholders with actionable insights into the complex problem of ICU capacity planning as highlighted in.[21] In partnership with NHS Digital, CPAS has been effectively deployed at individual hospitals as well as regionally throughout the UK. Its main goal is to support the organization of ICU beds, supplies, and staff members in order to efficiently handle the extraordinary demands on ICU resources brought on by the COVID-19 pandemic. The system uses data to forecast hospital demands at the national, regional, hospital, and individual levels. Widespread deployment of it in the UK has aided in capacity planning initiatives designed especially for COVID-19 patients. Rather than relying on algorithms, Qian et al.[21] used the sophisticated automated machine learning tool AutoPrognosis, which is based on the ideas of Bayesian optimization (BO), to tune every pipeline step collaboratively. CPAS made use of three different patient-level data sources, each of which provides information on a different facet of the patient's health. To create aggregated trend forecasts, CPAS also integrates community movement trend data.

In order to deliver high-quality healthcare services and make the best possible utilization of scarce resources, Altintop et al.[22] claimed that knowledge developments or efficiency measurements on unprocessed data in the healthcare information system were required. This study uses a dataset of operational and financial healthcare records from over 600 hospitals in Turkey in 2013, which includes around 200 features divided into three categories: administrative features (such as the load factor of beds), earnings items, and cost items. Hospital locations are also categorized. To produce fuzzy linguistic overviews, a genetic algorithm is implemented. For the dataset that was provided, the parameters of the genetic algorithms in their system were determined through experimentation and error.

Garcia et al.[23] explored the utilization of simulation models to enhance hospital readiness in the face of epidemics, with a specific focus on the COVID-19 scenario. They introduced a discrete event simulation model designed to assist in the short-term planning of hospital resources, particularly ICU beds, to effectively manage outbreaks such as the COVID-19 pandemic. Their simulation model incorporates stochastic modeling of patient admission and patient flow processes. The Gompertz growth model is employed to represent the patient arrival process, accurately capturing the exponential growth, peak, and decline of COVID-19 cases. Through empirical analysis, their study concludes that the Gompertz model surpasses other sigmoid models in terms of fitting pandemic-related data and predicting capacity. Patient flow modeling encompasses various pathways and dynamic length-of-stay estimation based on patient-level data. Their simulation model was applied in two regions of Spain during the COVID-19 waves in 2020, providing daily predictions to guide healthcare planning teams. Their work[23] underscores the significance of simulation models in replicating the complexity and variability of healthcare systems, especially in uncertain and rapidly changing situations like pandemics. Their proposed simulation framework enables short-term forecasting of resource requirements for COVID-19 patients, facilitating proactive planning by health authorities. The paper also underscores the potential application of the simulation model in future outbreaks.

The potential of machine learning in addressing five major issues related to the COVID-19 pandemic response is examined by van der Schaar et al.[24] Risk assessment, treatment planning, allocating resources, forecasting diseases, and medication development are all included in these difficulties. The authors argue that these technologies have an opportunity to save lives when integrated into healthcare systems and offer specific strategies for their real-world application.

Arvind et al.[25] highlights the utilization of machine learning techniques to forecast the intubation of hospitalized COVID-19 patients based on contemporaneous laboratory measurements achieving area under the receiver operating characteristic curve (AUC) of 0.84. Research on predicting the outcomes of hospitalized patients is abundant, with several papers focused on predicting mortality or ICU admissions for hospitalized cases, including.[26–28]

Previous studies often failed to provide detailed information about the features that influence ML models, making model explanations challenging. The papers we reviewed achieved a maximum accuracy of 83%. This study seeks to close that gap by offering a framework for understanding ML models used to predict hospitalization. Our technique not only improves the ability to predict hospitalization cases—which is critical during pandemics—but it also illustrates how ML models arrive at their results, boosting resource management and decision-making during pandemics.

3. Methodology

The proposed framework for predicting hospitalization cases during pandemic is outlined in this section, along with details about data description, preprocessing, feature engineering and how the models are trained. Fig. 1 displays the process flow chart. Each Block willl be covered in details in this section.

3.1. Data description

The Covid19MPD Dataset serves as a comprehensive repository of patient data encompassing COVID-19 test results and various medical conditions.[13] Comprising 95,840 instances distributed across 20 features, the dataset transcends its initial focus on COVID-19 analysis to address a more pressing concern: forecasting the hospitalization needs of patients. This strategic shift aims to assist hospital managers in optimizing resource allocation, particularly during pandemic crises.

To facilitate efficient predictive modeling, the dataset underwent a restructuring process whereby hospitalization feature (patient type) was relocated as output variable. Table 1 serves as a guide, providing a detailed description of the dataset's columns.

An inconsistency was found in the dataset where patients marked as being in the ICU '1' were incorrectly classified as not hospitalized '0'. This was corrected by adjusting the hospitalization status to match the logical progression of patient care. After that, the features labeled "dead," "ICU," and "intubated" were eliminated since the primary objective is to forecast hospitalization before a patient's registration.

Fig. 2 offers valuable insights into the interplay among dataset features, showcasing each feature's correlation with others to help understanding the dataset's underlying dynamics. After the elimination of the 'ICU' and 'intubated' features, the correlation matrix makes it evident that no two attributes have a significant relationship with one another.

3.2. Preprocessing and feature engineering

To preprocess the Covid19MPD Dataset, some steps were undertaken to ensure data integrity and suitability for modeling purposes. The following preprocessing procedures were applied to refine the dataset and rectify inconsistencies:

Initially, the 'age' attribute was binarized to simplify its representation while preserving its relevance to the model. Age values greater than 45 were assigned a value of 1, whereas those 45 or below were assigned a value of 0. The choice of 45 as the threshold

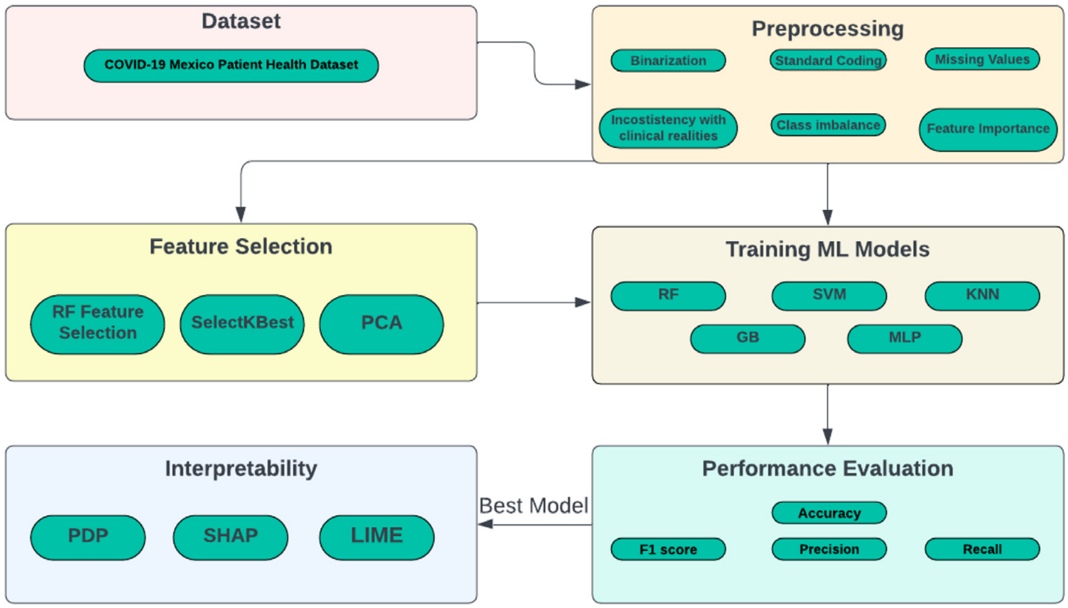


Fig. 1. The complete process for the interpretable ML model.

Table 1
Dataset description.

Feature	Description
Sex	Female/male.
Pneumonia	if the patient was diagnosed with pneumonia.
Age	Age of the tested group.
Pregnancy	if the patient is pregnant.
Diabetes	if the patient has a diagnosis of diabetes.
COPD	if the patient has a diagnosis of COPD.
Asthma	if the patient has a diagnosis of asthma.
Immunosuppression	if the patient has immunosuppression.
Hypertension	if the patient has a diagnosis of hypertension.
Other_disease	if the patient has a diagnosis of other diseases.
Cardiovascular	if the patient has a diagnosis of cardiovascular disease.
Obesity	if the patient is diagnosed with obesity.
Chronic_kidney_failure	if the patient has a diagnosis of chronic kidney failure.
Smoker	if the patient has a smoking habit.
Another_case	if the patient had contact with any other case diagnosed with covid.
Patient_type	identifies the type of care received by the patient in the unit, hospitalized(inpatient) or not (outpatient)
ICU	if the patient required to enter an Intensive Care Unit.
Intubated	if the patient required intubation.
death	if the patient passed away or survived the covid19.
Covid	the result of the analysis of the sample, if the patient got covid or not.

was based on its clinical significance, as individuals in this age group are more likely to develop several medical conditions.[29] Additionally, with the dataset's average age being approximately 42, selecting 45 provided a balanced division while maintaining predictive effectiveness. Although alternative thresholds such as 30, 40, and 50 were initially explored, experimental evaluations confirmed that 45 yielded the best model performance.

In the original dataset, the codes 97 and 98 were utilized to represent cases that were classified as "not applicable" or "not provided," respectively. These values were consistently reassigned to a standardized code '2' to maintain consistency across the dataset and minimize bias in subsequent analyses, to lessen their disproportionate influence on the model.

To assess the potential impact of dataset characteristics on model performance, we analyzed the skewness of each feature. Given that all attributes are categorical, skewness values reflect class imbalances, where a high positive skew indicates that the majority of instances belong to a single class, while a negative skew suggests the opposite. Several variables, including COPD (6.55), immunosuppression (6.74), chronic kidney failure (6.79), and cardiovascular disease (5.92), exhibited high positive skewness, suggesting that these conditions were rare in the dataset. Conversely, hospitalization (-1.25) displayed negative skewness, indicating that most cases fell into the positive class (hospitalized).

Feature importance analysis using RF is a valuable technique employed to discern the relative significance of different variables in predicting outcomes within a dataset.[30] In the context of the COVID-19 Mexico Patient Health Dataset, RF serves as a robust tool for assessing the importance of various features in predicting hospitalized cases and related outcomes.

Table 2 presents an overview of the feature importance rankings derived from the RF analysis and offers insights into the relative contributions of each attribute to the predictive power of the model. Features with higher rankings signify greater importance in influencing the model's predictions, indicating their significant impact on the outcome variables. Conversely, attributes with lower rankings are deemed to have relatively lesser influence on the model's predictive capabilities.

The dataset included 73,381 instances classified as hospitalized cases and 22,458 instances classified as outpatient cases. This emphasizes the dataset's underlying class imbalance, where inpatient cases are far more common than outpatient instances. In order to rectify the intrinsic class imbalance observed in the dataset, the oversampling and downsampling approaches were used. The details of these approaches' use and their effect on model performance and predicted accuracy are discussed in the Results and discussion section.

3.3. ML models

Machine learning provides powerful tools for identifying complicated patterns in data and building prediction models.[31] ML techniques, including RF, SVM, KNN, GB, MLP, DT, Logistic Regres-

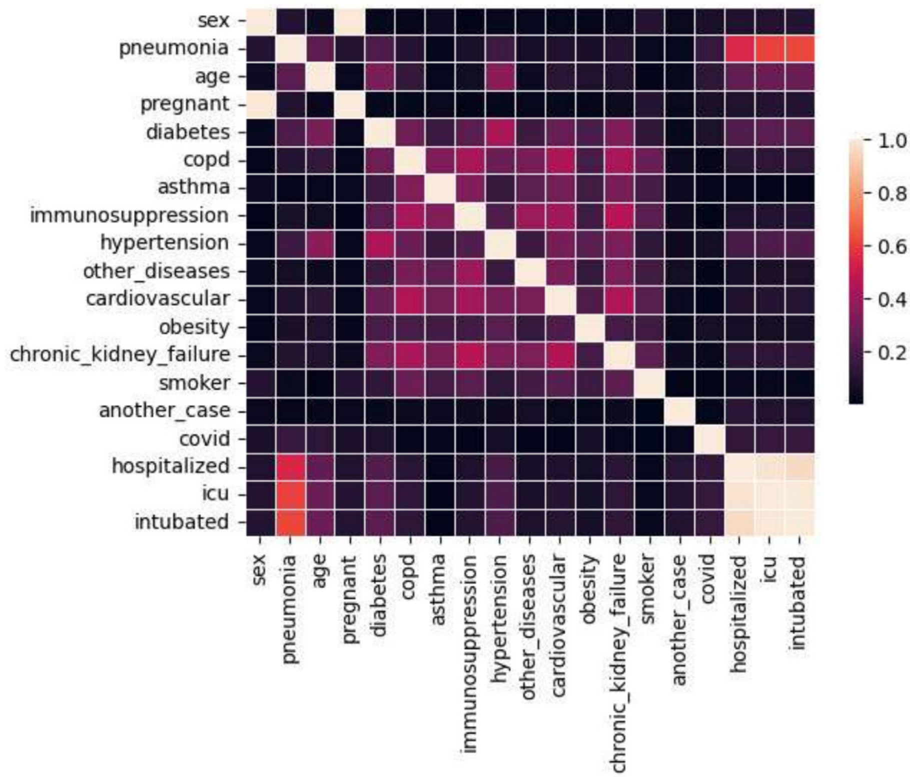


Fig. 2. The Features' Heat Map.

Table 2
Features Importance usung RF.

-	Feature	Feature Importance
1	Pneumonia	0.5134758
14	Another_case	0.10218845
2	Age	0.06774288
15	covid	0.04822036
4	Diabetes	0.04711961
8	Hypertension	0.03227444
11	Obesity	0.02457615
12	Chronic_kidney_failure	0.02334989
13	Smoker	0.02210968
5	COPD	0.02162293
9	Other_disease	0.02021054
7	Immunosuppression	0.01946763
10	Cardiovascular	0.01665956
3	Pregnancy	0.01569267
6	Asthma	0.0137337
0	Sex	0.01155569

sion (LR), and Naïve Bayes (NB) are widely used for classification and regression tasks.[32,33] To increase precision and decrease overfitting, RF builds numerous decision trees and aggregates their predictions.[34] SVM aims to discover the best hyperplane that splits data points into various classes in a high-dimensional space, making it effective for both linear and nonlinear classification.[35] KNN classifies data points based on the majority of their k-nearest neighbors, making it straightforward and flexible to varied datasets.[36] Gradient Boosting sequentially trains decision trees, with each consecutive tree rectifying the errors of its predecessor. This leads to enhanced accuracy and predictive power.[37] MLP is a popular form of artificial neural network (ANN) that comprises of an input layer, one or more hidden layers, and an output layer.[38,39] The MLP learning technique comprises forward propagation, error calculation, and backpropagation to update the model's weights over successive epochs to learn optimal weights.[40]

A variety of techniques are included in the classifier set used in this work, including RF, KNN, SVM, GB, MLP, DT, LR and NB. These models were chosen based on their effectiveness in similar predictive modeling applications, as discussed by Uddin et al.[33] Two different hidden layer configurations within the MLP architecture were investigated to maximize prediction accuracy and model performance. The first architecture comprised three hidden layers with node counts of 30, 20, and 4, respectively, while the second configuration featured two hidden layers with node counts of 50 and 20, respectively.

The dataset was divided into three sets: training, validation, and testing. The training set, which accounts for 80% of the data, is used to train models and estimate parameters. To avoid overfitting, 10% of the data is allocated for validation, allowing for hyperparameter fine-tuning and performance evaluation on untrained data. The remaining 10% is reserved for the testing set, which evaluates the model's generalizability and accuracy on new data.

When analyzing the performance of a classification model, multiple metrics are often employed to comprehend various elements of the model's performance including accuracy, precision, recall, and F1 score.[41] Accuracy is the proportion of accurately predicted instances to all instances in the dataset. It is a simple metric that gives an overall impression of how frequently the model is true.[41] It can be calculated using Eq. (1):

$$Accuracy = \frac{TP + TN}{Total\_instances} \tag{1}$$

where TP is true positive cases and TN is true negative cases.[42]

Precision is defined as the ratio of accurately predicted positive observations to all expected positives. It determines how many cases anticipated as positive are actually positive.[41] It can be calculated using Eq. (2):

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

where TP is true positive cases and FP is false positive cases.[43]

Recall is the ratio of accurately predicted positive observations to total observations in the class.[41] It measures how many true positive instances are caught by the model and it can be calculated using Eq. (3):

$$Precision = \frac{TP}{TP + FN} \tag{3}$$

where TP is true positive cases and FN is false negative cases.[43] Recall is critical in cases where missing a positive occurrence (false negative) has serious effects.

The F1 Score represents the harmonic mean of precision and recall. It provides a single statistic that balances both concerns, making it a valuable measure for calculating false positives and false negatives.[41] It can be calculated using Eq. (4)[44]:

$$F1score = 2 * \frac{recall * precision}{recall + precision} \tag{4}$$

Another key performance metric for classification models is the AUC of the ROC curve. The AUC measures a model's ability to distinguish between classes and provides insight into the trade-off between sensitivity and specificity.[45] AUC can be calculated using Eq. (5):

$$AUC = \int_0^1 TPR(FPR)\, dFPR \tag{5}$$

where:

True Positive Rate (TPR) or recall is given by Eq. (3), and False Positive Rate (FPR) is given by Eq. (6):

$$FPR = \frac{FP}{FP + TN} \tag{6}$$

4. Results and discussion

Table 3 shows the comparison of classification performance of the ML algorithms used using 10-fold cross validation. It displays the performance of the trained model across multiple evaluation metrics, such as accuracy, precision, recall, F1 score and AUC.

Although the total accuracy percentages varied from 82% to 86%, more examination showed excellent recall, precision, and F1 scores for the "1" class, which corresponds to hospitalized cases. This demonstrates the model's ability to reliably identify situations that require hospitalization, as well as its potential application in clinical decision-making and resource allocation.

Despite the class imbalance, The use of AUC helped mitigate the effects of skewed distributions, as AUC evaluates model discrimination rather than absolute classification accuracy. Additionally, to minimize bias, the implemented stratified cross-validation ensured that training and validation sets maintained proportional class distributions.

Out of all the ML models that were assessed, the results show that GBM performed the best with AUC of 0.8696. This is due to its capacity to iteratively improve predictions by fixing cases that were incorrectly classified at every learning stage. In contrast to conventional ensemble techniques such as bagging, which train models separately, gradient boosting constructs trees in a sequential fashion, modifying each weak learner's contribution according to the residual errors of earlier iterations. This makes it possible for it to learn intricate, non-linear feature interactions, which is especially useful because our dataset is categorical.Gradient Boosting also dynamically modifies instance weights, improving its capacity to manage unequal class distributions. Models such as RF, on the

Table 3
Evaluation of model performance on the validation set.

ML Model		RF	GBM	KNN	SVM	MLP III	MLP II	DT	NB	LR
Not Hospitalized	Precision	0.6872 ± 0.0041	**0.7049 ± 0.0056**	0.6837 ± 0.0013	0.7136 ± 0.0053	0.6910 ± 0.0005	0.7109 ± 0.0032	0.6723 ± 0.0034	0.6141 ± 0.0055	0.7133 ± 0.0050
	Recall	0.5943 ± 0.0072	**0.6172 ± 0.0063**	0.5285 ± 0.0071	0.5822 ± 0.0023	0.5900 ± 0.0010	0.6076 ± 0.0042	0.5996 ± 0.0021	0.6453 ± 0.0049	0.5113 ± 0.0062
	F1 Score	0.6374 ± 0.0056	**0.6581 ± 0.0060**	0.5962 ± 0.0011	0.6412 ± 0.0030	0.6781 ± 0.0076	0.6623 ± 0.0044	0.6339 ± 0.0028	0.6293 ± 0.0057	0.5956 ± 0.0063
Hospitalized	Precision	0.8798 ± 0.0009	**0.8862 ± 0.0012**	0.8640 ± 0.0023	0.8780 ± 0.0018	0.8864 ± 0.0031	0.8852 ± 0.0020	0.8804 ± 0.0044	0.8888 ± 0.0025	0.8613 ± 0.0030
	Recall	0.9165 ± 0.0061	**0.9202 ± 0.0058**	0.9245 ± 0.0046	0.9279 ± 0.0062	0.9051 ± 0.0038	0.9119 ± 0.0047	0.9098 ± 0.0023	0.8748 ± 0.0042	0.9366 ± 0.0050
	F1 Score	0.8977 ± 0.0035	**0.9029 ± 0.0032**	0.8933 ± 0.0029	0.9023 ± 0.0035	0.8973 ± 0.0024	0.8995 ± 0.0036	0.8948 ± 0.0034	0.8818 ± 0.0045	0.8974 ± 0.0040
AUC		0.8525 ± 0.0101	**0.8696 ± 0.0095**	0.8217 ± 0.0123	0.8102 ± 0.0112	0.8264 ± 0.0097	0.8364 ± 0.0089	0.8042 ± 0.0109	0.8207 ± 0.0094	0.8363 ± 0.0103
Accuracy		84.17% ± 0.0043	**85.12% ± 0.0051**	83.11% ± 0.0047	84.64% ± 0.0056	82.64% ± 0.0049	83.64% ± 0.0045	83.67% ± 0.0048	82.07% ± 0.0050	83.63% ± 0.0046

other hand, give each instance the same weight, which can lessen sensitivity to patterns of minority classes.

To assess whether a more sophisticated ensemble method could enhance model performance, a stacking classifier using RF, GBM, and SVM as base models was used with LR serving as the meta-model. While stacking is often expected to improve classification performance by leveraging the strengths of multiple models,[46] in this case, it did not provide a significant advantage over individual models. The results showed that the stacking model performed comparably to the best individual classifiers, achieving accuracy of 85.04% and indicating that the base models were already capturing the key patterns in the data effectively. This suggests that, for this particular dataset and classification task, the added complexity of stacking did not translate into noticeable performance gains.

To optimize the GB model performance, we performed automated hyperparameter tuning using Optuna, a Bayesian optimization framework that efficiently searches for the best hyperparameters using a pruning-based approach.[47] Hyperparameters such as number of estimators, maximum depth, minimum samples per split, and subsample ratio were tuned. The optimal set of hyperparameters, which led to the best model performance, is reported in Table 4. The model's performance was only marginally improved through tuning. On the validation set, the model's accuracy increased to 85.31%, and on the testing set, it reached 85.61%.

The investigation of oversampling and downsampling strategies reveals the common issue of class imbalance in healthcare datasets, in which some classes are under-represented. As Table 5 demonstrates, attempts to rectify this imbalance were done by exploring the effects of both downsampling and oversampling, however neither strategy produced significant improvements in the GB model performance. The 'Over/Down sampling' results in Table 5 shows that resolving class imbalance remains a difficult issue that might require more advanced techniques or dataset-specific approaches.

RF feature selection with the top 8, 10, and 12 features, SelectKBest with k = 8, 10, and 12, and Principal Component Analysis (PCA) with 8 and 12 components were all utilized to enhance hospitalized case prediction using GB model. Despite these attempts, no significant enhancement in performance was noted. PCA, which is typically used for dimensionality reduction, and SelectKBest, which assesses features based on their statistical association with the target variable, both failed to improve performance. These results highlight the complexities of feature selection and its impact on model performance. Detailed performance results are provided in Table 6.

Table 7 shows that several feature selection strategies that we explored, such as RF and SelectKBest, perform similarly in a subset of 8 features. However, the difference in the 4th and 5th features chosen by RF (obesity) and SelectKBest (copd) implies that each technique evaluates feature value or relevance to the hospitalized case prediction task differently.

The preparation methods taken in this scenario, such as handling missing values, demonstrate the complexities of preparing healthcare datasets for predictive modeling and the importance of thorough data cleaning and feature engineering to ensure the dataset's integrity and dependability.

To gain deeper insights into model behavior, error analysis was conducted by examining misclassification patterns across different models. One of the key findings was that false negatives (i.e., hospitalized cases misclassified as non–hospitalized) were generally more frequent than false positives. This trend was most prominent in models with lower AUC scores, such as NB, SVM, and DT, suggesting that these models struggled to differentiate hospitalized cases that shared similar feature distributions with the non–hospitalized group. On the other hand, models that achieved the highest AUC scores, such as GBM, and RF exhibited lower false negative rates, indicating their ability to better capture complex feature interactions. Conversely, simpler models like LR tended to produce higher false positive rates, meaning that some non–hospitalized cases were mistakenly classified as hospitalized. This suggests that these models may overestimate risk, potentially due to their reliance on linear decision boundaries.

In comparison to De Holanda et al.,[16] who developed two predictive machine learning models aimed at identifying COVID-19 patients at elevated risk of hospitalization or mortality, our study achieved higher predictive accuracy using the gradient boosting algorithm. Specifically, our model attained an accuracy of 85.63% and AUC of 0.8696 in predicting hospitalization during the COVID-19 pandemic, surpassing their reported accuracy of 71% with an AUC of 0.75 for hospitalization prediction.

Predicting hospitalization for pediatric COVID-19 patients in Malaysia was the main goal of Liew et al.[48] The Adaptive Boosting (AdaBoost) technique was determined to be the best-performing model with an AUROC of 0.95 after feature selection was done

Table 4
Best parameters obtained by Optuna.

Parameter	Value
learning_rate	0.05875538439237237
n_estimators	98
max_depth	5
min_samples_split	4
min_samples_leaf	5
subsample	0.9616982928990448

Table 5
Performance evaluation of the gradient boosting model with different sampling techniques.

		GB Model	Sampling Techniques	
		Original	Over-Sampling	Down-Sampling (1:2)
Not Hospitalized	Precision	0.71 ± 0.02	0.58 ± 0.03	0.65 ± 0.02
	Recall	0.64 ± 0.03	0.72 ± 0.04	0.63 ± 0.03
	F1-Score	0.66 ± 0.02	0.64 ± 0.03	0.64 ± 0.02
Hospitalized	Precision	0.89 ± 0.01	0.87 ± 0.02	0.86 ± 0.02
	Recall	0.92 ± 0.02	0.81 ± 0.03	0.84 ± 0.03
	F1-Score	0.90 ± 0.01	0.84 ± 0.02	0.85 ± 0.02
AUC		0.87 ± 0.01	0.79 ± 0.02	0.81 ± 0.02
Accuracy		85.35% ± 0.05	80.86% ± 0.075	82.71% ± 0.065

Table 6
Performance evaluation of the GB model with different feature selection techniques.

		RF Feature Selection			PCA Feature Selection		SelectKBest Feature Selection		
		Top 8	Top 10	Top 12	8 Components	12 Components	k = 8 Features	k = 10 Features	k = 12 Features
Not Hospitalized	Precision	0.70 ± 0.02	0.71 ± 0.02	0.71 ± 0.02	0.71 ± 0.03	0.71 ± 0.02	0.71 ± 0.02	0.71 ± 0.02	0.71 ± 0.02
	Recall	0.61 ± 0.03	0.61 ± 0.03	0.60 ± 0.02	0.61 ± 0.03	0.62 ± 0.03	0.62 ± 0.03	0.62 ± 0.03	0.62 ± 0.03
	F1-Score	0.65 ± 0.02	0.66 ± 0.02	0.65 ± 0.02	0.64 ± 0.03	0.66 ± 0.02	0.66 ± 0.02	0.66 ± 0.02	0.66 ± 0.02
Hospitalized	Precision	0.88 ± 0.02	0.88 ± 0.02	0.88 ± 0.02	0.88 ± 0.02	0.89 ± 0.02	0.89 ± 0.02	0.89 ± 0.02	0.89 ± 0.02
	Recall	0.92 ± 0.02	0.92 ± 0.02	0.92 ± 0.02	0.91 ± 0.02	0.92 ± 0.02	0.92 ± 0.02	0.92 ± 0.02	0.92 ± 0.02
	F1-Score	0.90 ± 0.01	0.90 ± 0.01	0.90 ± 0.01	0.90 ± 0.01	0.90 ± 0.01	0.90 ± 0.01	0.90 ± 0.01	0.90 ± 0.01
AUC		0.84 ± 0.023	0.84 ± 0.025	0.81 ± 0.051	0.82 ± 0.01	0.83 ± 0.098	0.83 ± 0.02	0.83 ± 0.005	0.84 ± 0.002
Accuracy		84.64% ± 0.02	84.64% ± 0.02	84.64% ± 0.01	83.88% ± 0.07	84.88% ± 0.02	84.88% ± 0.06	84.88% ± 0.09	84.88% ± 0.05

Table 7
The Top 8 Features selected by RF and SelectKBest.

Technique	Top 8 Features							
RF	pneumonia	age	diabetes	hypertension	obesity	chronic_kidney_failure	another_case	covid
SelectKBest	pneumonia	age	diabetes	copd	hypertension	chronic_kidney_failure	another_case	covid

using Recursive Feature Elimination (RFE). Direct performance comparison is challenging due to the disparities in target demographics, even though their work sheds light on pediatric admissions. But in contrast to their study, which mostly used characteristics based on demographics and symptoms, we included a wider variety of clinical and comorbidity-related characteristics.

Performance evaluations for several machine learning models are included in the internal validation results from Liew et al.[48] Although the RF classifier in their investigation had a PPV of 1.00 and a specificity of 1.00, it only had a sensitivity of 0.23 and an AUROC of 0.94. With an AUROC of 0.93 and a sensitivity of 0.32, the XGBoost model demonstrated a similar pattern of high specificity and poor sensitivity. This implies that their models were more cautious, giving specificity precedence over sensitivity, which would have resulted in a higher percentage of false negatives when it came to hospitalization predictions.

In contrast, considering the real-world applications of hospitalization prediction, our model aimed to strike a balance between sensitivity and specificity. With an AUROC of 0.8696, our GB model demonstrated a more equitable trade-off between specificity and sensitivity. Additionally, there were notable differences in the F1 scores between classifiers according to Liew et al.[48] For example, LR received an F1 score of 0.55, whereas RF received 0.37. Our models demonstrated a superior balance between precision and recall by achieving higher and more consistent F1 scores across several classifiers.

Buenrostro-Mariscal et al.[49] used the Mexican Health and Aging Study (MHAS) to predict hospitalizations in older persons. They created an RF model and used permutation importance and impurity reduction approaches to evaluate the relevance of the variables. They obtained a specificity of 0.4935 and a sensitivity of 0.7215. The model mostly used socioeconomic and functional health characteristics, and the best predictors were age, history of cerebrovascular accidents, and functional limits. Instead of concentrating solely on aging populations, our study is based on hospitalization prediction for COVID-era patients with different clinical problems. To guarantee resilience, the model in[49] employed a variety of partitioning techniques including cross-validation. Although our study's validation methodology was similar, our feature set was very different, emphasizing health condition predictions above functional limitations.

The use of XAI strategies to improve model interpretability is a significant difference in our research. Although Liew et al.[48] employed RFE to choose features, they did not go into great detail about model interpretability outside of feature ranking. Similarly, Buenrostro-Mariscal et al.[49] did not use explainability techniques like SHAP in their variable importance study. Our method incorporates explainability to help with clinical interpretability by offering insights into the model's decision-making process.

5. Interpretability results

To ensure transparency and a better understanding of the variables and features, interpretable ML is used to provide explanations for the model's behavior, addressing the gap in the literature regarding explainability and completing the cycle of creating an explainable ML model. The application of interpretation tools after model training is referred to as "post hoc interpretability." These strategies can be applied to models that are intrinsically interpretable, such as GB, RF, and decision trees.[50,51] Permutation feature importance[51] provides a simple way for determining the significance of features in a ML model by calculating the increase in prediction error after permuting the feature's values, thereby disturbing its association with the actual outcome.[51] Another useful technique for interpreting ML models is the partial dependence plot (PDP), which reveals the marginal effect of one or two parameters on the predicted outcome.[50] PDPs are commonly used for looking into the relationship between features and the desired variable in ML models.[50,51]

To better understand the contribution of individual features, ablation study was conducted by systematically removing each feature and evaluating its impact on AUC and accuracy. As shown in Fig. 3, removing critical clinical attributes, including pneumonia status resulted in a notable decline in model performance. This finding emphasizes the significance of pre-existing medical conditions in predicting hospitalization outcomes.

Conversely, demographic attributes such as smoking status, and prior contact with another COVID-19 case had minimal impact on

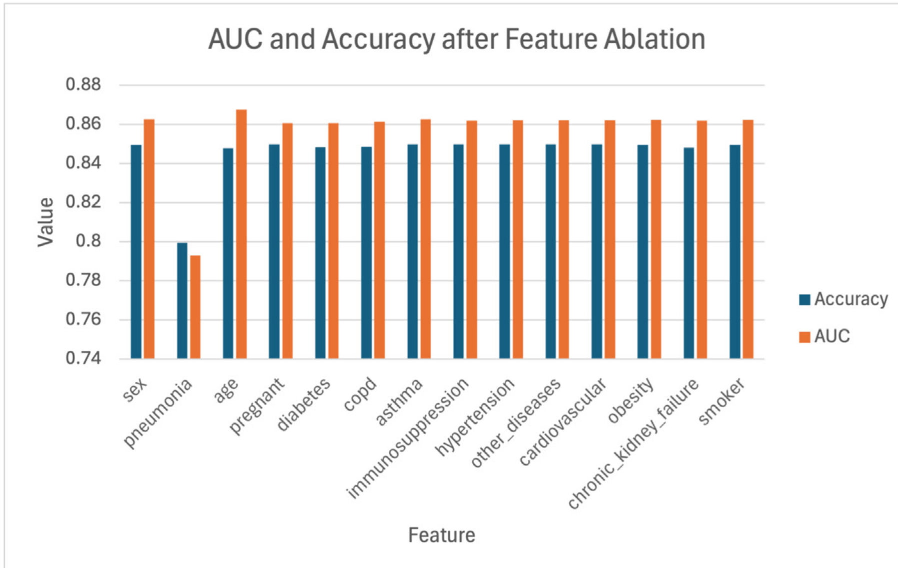


Fig. 3. AUC and Accuracy after Feature Ablation.

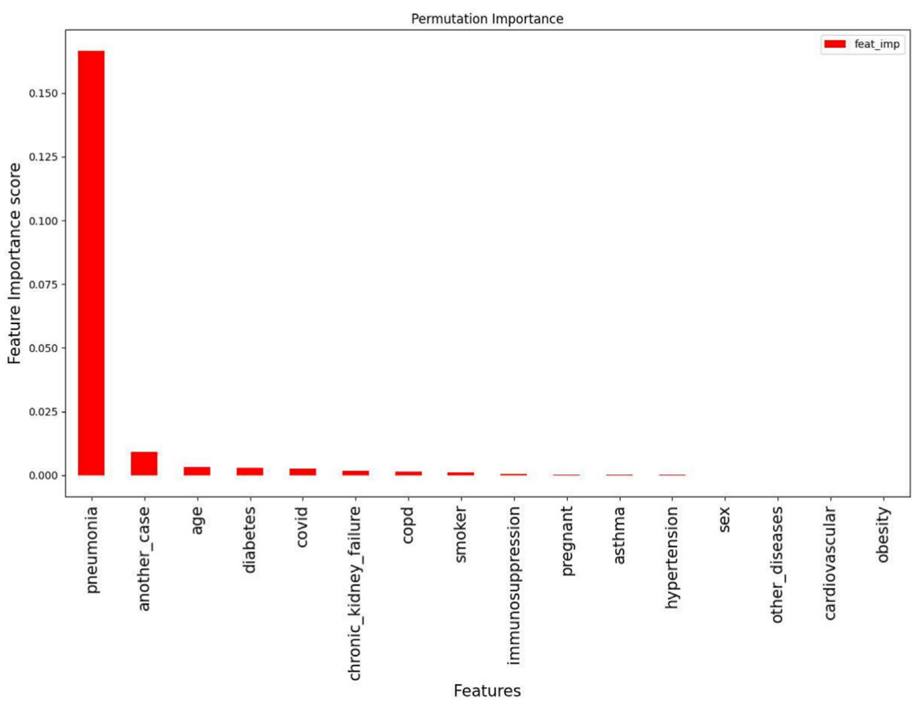


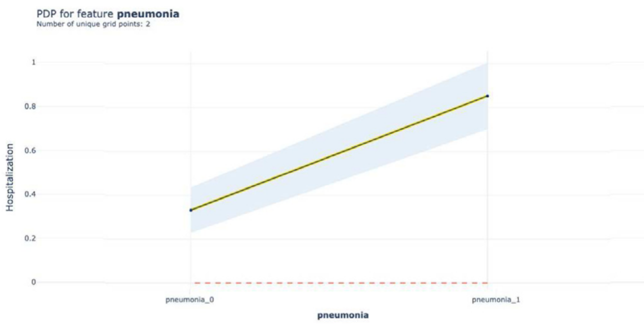
Fig. 4. Feature Importance Plot.

model performance, suggesting that they contribute less to hospitalization prediction.

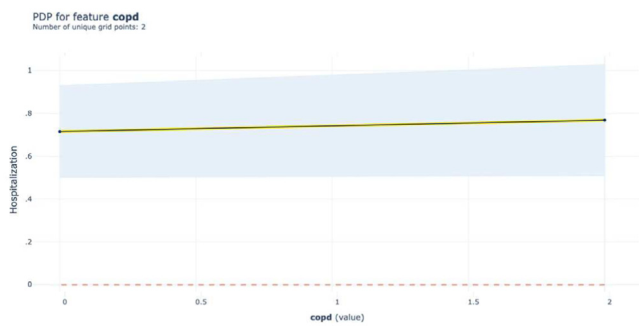
SHapley Additive exPlanations (SHAP) is another way for explaining individual predictions provided by ML algorithms. Its key objective is to explain the prediction of a given instance by calculating the contribution of each feature to that prediction.[50,51]

In this paper, post hoc interpretability methods are utilized to interpret the GB model's result including permutation importance, PDP, and SHAP. The permutation importance plot in Fig. 4 indicates that "Pneumonia" is the most crucial feature for the model's judgment regarding the patients' likelihood of being hospitalized or not, which is in line with our findings regarding the significance of RF features in Table 2.

Figs. 5a, 5b, and 5c, represent the PDP plots for the features 'Pneumonia', 'COPD', and 'COVID', respectively. The impact of Pneumonia on hospitalization cases is evident from the PDP plot in



(a) PDP plot for Pneumonia Feature.



(b) PDP plot for Copd Feature.

(c) PDP plot for Covid Feature

Fig. 5. Partial Dependence Plots.

Fig. 5a. When the 'pneumonia' feature is zero, the hospitalization probability is low; however, it greatly increases when the 'pneumonia' is one. This result is consistent with the previous analyses and emphasizes the significance of pneumonia as a predictor. Patients who have pneumonia have a significantly higher chance of being admitted to the hospital.

The PDP plots for COVID-19 and COPD features in Figs. 5b and 5c reveal a small increase in the likelihood of hospitalization when these symptoms exist. However, this rise is not as significant as the one seen with pneumonia.

The interaction plot in Fig. 6 shows that when a patient does not have Covid or pneumonia, the model generally classifies them as non–hospitalized. In contrast, the model frequently classifies patients with pneumonia as hospitalized. When a patient has pneumonia as well as COVID-19, the likelihood of hospitalization increases significantly.

For our SHAP analysis, we selected a singular instance from the dataset, focusing on non–hospitalized as the target variable, as depicted in Fig. 7a. The discernible trend indicates that in instances where pneumonia is absent (denoted by 0), the model tends to predict a non–hospitalized outcome.

In Fig. 7b, the SHAP plot depicts that age and 'another_case' feature have the most impact on the decision to hospitalize the patient. It also shows that women aged 45 and older who present with pneumonia alongside COVID-19 are notably predisposed to hospitalization.

SHAP produces global explanations for the ML results using kernel explainer. The SHAP plot summary in Fig. 8 shows that Pneumonia affects positively the model prediction of the hospitalization.

6. Limitations and Challenges

This study draws attention to a number of limitations in the dataset and methodology. The data imbalance, where one class greatly outweighs the other, is one prominent limitation that could cause skewed model predictions and decreased accuracy. Moreover, the dataset was gathered during the COVID-19 pandemic, making the results practical and suitable for COVID-19 pandemic situations. However, more research is needed to generalize these findings to the broader context of other pandemics or healthcare scenarios.

Furthermore, the dataset's features might be constrained because the ML models' predictive power could be enhanced by integrating more comprehensive patient data, such as demographics, vital signs, and blood test results.

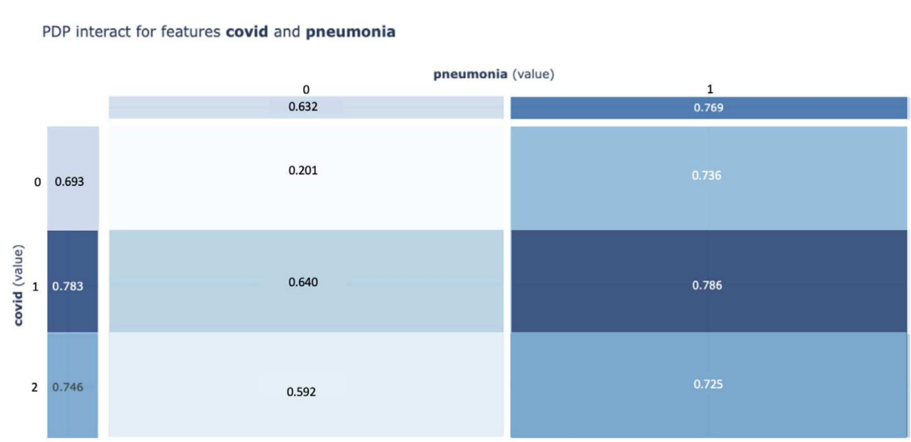
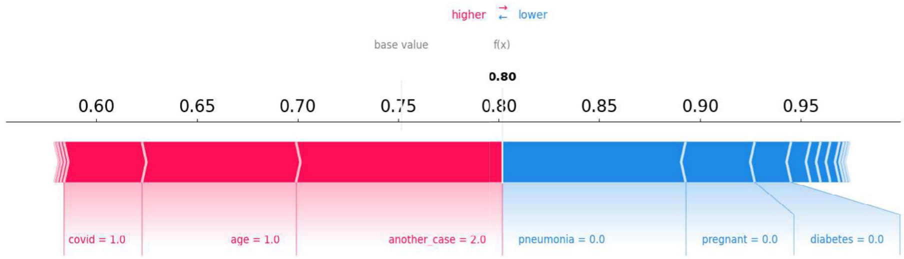


Fig. 6. Interaction PDP between Covid and Pneumonia.



(a) Local SHAP explanations for non-hospitalized patients.

(b) Local SHAP explanations for hospitalized patients.

Fig. 7. Local SHAP plots.

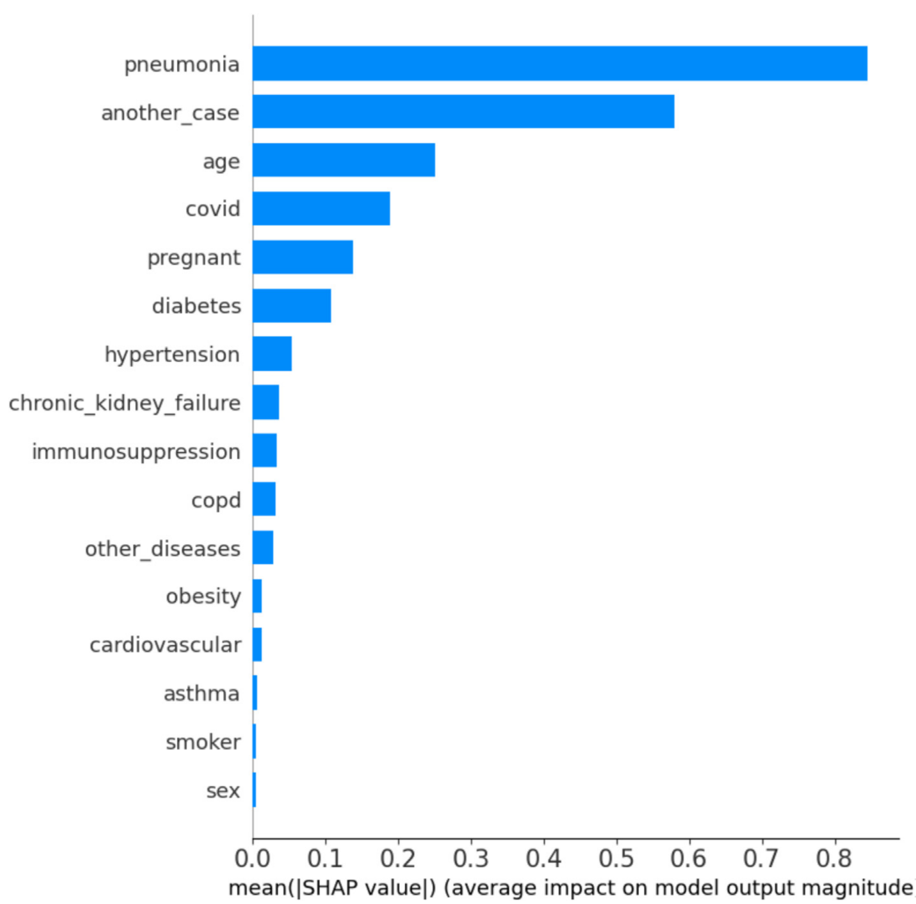


Fig. 8. Global SHAP explanations for hospitalized patients.

7. Conclusion and future work

Hospital management presents significant challenges that call for effective solutions, particularly during pandemics. This paper uses ML on a dataset obtained during the COVID-19 pandemic as a case study to forecast hospitalizations, shedding light on the significant role machine learning approaches may play in tackling the challenges of hospital management during pandemics. Several supervised learning algorithms were used and assessed through the examination of the dataset, yielding good prediction accuracy and insights into the variables impacting hospitalization outcomes. With the Gradient Boosting model, hospitalization cases may be predicted with 85.63% accuracy and AUC score of 0.8696. Notwithstanding obstacles including unequal class distribution and data scarcity, the findings show that machine learning has the ability to enhance hospital management in the event of a pandemic.

Moreover, the utilization of interpretability methods facilitates the understanding of machine learning models, developing confidence and trust in their predictions. For Example, according to our analysis, the most important factor in determining whether or not the patients would be hospitalized is "pneumonia". The SHAP analysis underscores that elderly women diagnosed with both pneumonia and COVID-19 are at significantly elevated risk of hospitalization. Explainable machine learning helps stakeholders and hospital managers make well-informed decisions about patient care and resource allocation by offering insights on feature relevance and model behavior.

Acquiring a broader range of datasets, using sophisticated preprocessing approaches, and directly collaborating with medical experts and hospital managers can improve the models' accuracy and expand their applicability in predicting hospitalization outcomes and guiding clinical judgment during pandemics. Further insights into patient demographics, regional variances, and socioeconomic aspects can be obtained by integrating novel and valuable datasets from other sources. Such richer set of data could help to strengthen and generalize ML models, increasing their usefulness in hospital management conditions.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Jani K, Chaudhary B. Importance of Hospital Management, Springer Nature Singapore. *Singapore*. 2023;25–41. https://doi.org/10.1007/978-981-19-6692-7_2.
2. Bastani P, Mohammadpour M, Bahmaei J, Ravangard R, Mehralian G. Hospital management by health services management graduates: the change paradigm in Iran. *Heliyon*. 2021;7(11):e08414.
3. Mohammadinia L, Saadatmand V, Khaledi Sardashti H, Darabi S, Esfandiary Bayat F, Rejeh N, Vaismoradi M. Hospital response challenges and strategies during covid-19 pandemic: a qualitative study. *Front Public Health*. 2023;11. https://doi.org/10.3389/fpubh.2023.1167411. URL: https://www.frontiersin.org/articles/10.3389/fpubh.2023.1167411.
4. Ravaghi H, Khalil M, Al-Badri J, Naidoo AV, Ardalan A, Khankeh H. Role of hospitals in recovery from COVID-19: Reflections from hospital managers and frontliners in the eastern mediterranean region on strengthening hospital resilience. *Front Public Health*. 2023;10:1073809.
5. Ebrahimi Rigi Z, Namjoo Z, Jabarpour M, Ahmadinejad M, Ahmadipour M, Mangolian Shahrbabaki P, Dehghan M. Hospital challenges and managerial approaches to combat covid-19 outbreak: a qualitative study in southeastern iran. *BMC Health Serv Res*. 2023;23(1):690. https://doi.org/10.1186/s12913-023-09631-0.
6. Hossny EK, Morsy SM, Ahmed AM, Saleh MSM, Alenezi A, Sorour MS. Management of the COVID-19 pandemic: challenges, practices, and organizational support. *BMC Nurs*. 2022;21(1):196.
7. Sarmiento Varón L, González-Puelma J, Medina-Ortiz D, Aldridge J, Alvarez-Saravia D, Uribe-Paredes R, Navarrete MA. The role of machine learning in health policies during the covid-19 pandemic and in long covid management. *Front Public Health*. 2023;11. https://doi.org/10.3389/fpubh.2023.1140353. URL: https://www.frontiersin.org/articles/10.3389/fpubh.2023.1140353.
8. A.Z. Al Meslamani, A.S. Jarab, M.A. Ghattas, The role of machine learning in healthcare responses to pandemics: maximizing benefits and filling gaps, Journal of Medical Economics 26 (1) (2023) 777–780. doi:10.1080/13696998.2023.2224018. URL: doi: 10.1080/13696998.2023.2224018.
9. Cossio M, Gilardino RE. Would the use of artificial intelligence in covid-19 patient management add value to the healthcare system? *Front Med*. 2021;8. https://doi.org/10.3389/fmed.2021.619202. URL: https://www.frontiersin.org/articles/10.3389/fmed.2021.619202.
10. Elbatanouny H, Kleanthous N, Dahrouj H, Alusi S, Almajali E, Mahmoud S, Hussain A. Insights into parkinson's disease-related freezing of gait detection and prediction approaches: A meta analysis. *Sensors*. 2024;24(12):3959.
11. Covid - coronavirus statistics - worldometer," covid - coronavirus statistics - worldometer (2001). URL: https://www.worldometers.info/coronavirus/.
12. Alimohamadi Y, Yekta EM, Sepandi M, Sharafoddin M, Arshadi M, Hesari E. Hospital length of stay for covid-19 patients: a systematic review and meta-analysis. *Multidiscip Respirat Med*. 2022;17(1).
13. Covid-19 mexico patient health dataset, https://www.kaggle.com/datasets/riteshahlawat/covid19-mexico-patient-health-dataset/ (2019).
14. Maleki Varnosfaderani S, Forouzanfar M. The role of ai in hospitals and clinics: Transforming healthcare in the 21st century. *Bioengineering*. 2024; 11(4):337.
15. Salcedo D, Guerrero C, Saeed K, Mardini J, Calderon-Benavides L, Henriquez C, Mendoza A. Machine learning algorithms application in covid-19 disease: a systematic literature review and future directions. *Electronics*. 2022;11(23). https://doi.org/10.3390/electronics11234015. URL: https://www.mdpi.com/2079-9292/11/23/4015.
16. de Holanda WD, e Silva LC, Sobrinho ÁAdCC. Machine learning models for predicting hospitalization and mortality risks of covid-19 patients. *Exp Syst Appl*. 2024;240:122670.
17. Brnabic A, Hess LM. Systematic literature review of machine learning methods used in the analysis of real-world data for patient-provider decision making. *BMC Med Inform Decis Mak*. 2021;21(1):54. https://doi.org/10.1186/s12911-021-01403-2.
18. Weissman GE, Crane-Droesch A, Chivers C, Luong T, Hanish A, Levy MZ, Lubken M, Becker M, Draugelis ME, Anesi GL, Brennan PJ, Christie JD, Hanson CW, Mikkelsen ME, Halpern SD. Locally informed simulation to predict hospital capacity needs during the covid-19 pandemic. *Ann Intern Med*. 2020;173 (1):21–28. https://doi.org/10.7326/M20-1260.
19. Fort D, Seoane L, Unis GD, Price-Haywood EG. Locally informed modeling to predict hospital and intensive care unit capacity during the covid-19 epidemic. *Ochsner J*. 2020;20(3):285. URL: http://www.ochsnerjournal.org/content/20/3/285.abstract.
20. A.S.J. Ahmad Z. Al Meslamani, M.A. Ghattas, The role of machine learning in healthcare responses to pandemics: maximizing benefits and filling gaps, Journal of Medical Economics 26 (1) (2023) 777–780, pMID: 37293941. arXiv: https://doi.org/10.1080/13696998.2023.2224018, doi:10.1080/13696998.2023.2224018. URL: doi: 10.1080/13696998.2023.2224018.
21. Qian Z, Alaa AM, van der Schaar M. Cpas: the uk's national machine learning-based hospital capacity planning system for covid-19. *Mach Learn*. 2021;110 (1):15–35. https://doi.org/10.1007/s10994-020-05921-4.
22. Altintop T, Yager RR, Akay D, Boran FE, Ünal M. Fuzzy linguistic summarization with genetic algorithm: an application with operational and financial healthcare data. *Int J Uncertain, Fuzziness Knowledge-Based Syst*. 2017;25 (04):599–620. https://doi.org/10.1142/S021848851750026X.
23. Garcia-Vicuña D, Esparza L, Mallor F. Hospital preparedness during epidemics using simulation: the case of covid-19. *Central Eur J Oper Res*. 2022;30 (1):213–249. https://doi.org/10.1007/s10100-021-00779-w. URL: https://doi.org/10.1007/s10100-021-00779-w.
24. van der Schaar M, Alaa AM, Floto A, Gimson A, Scholtes S, Wood A, McKinney E, Jarrett D, Lio P, Ercole A. How artificial intelligence and machine learning can help healthcare systems respond to covid-19. *Mach Learn*. 2021;110(1):1–14. https://doi.org/10.1007/s10994-020-05928-x.
25. Arvind V, Kim JS, Cho BH, Geng E, Cho SK. Development of a machine learning algorithm to predict intubation among hospitalized patients with covid-19. *J Crit Care*. 2021;62:25–30.
26. Galanter W, Rodríguez-Fernández JM, Chow K, Harford S, Kochendorfer KM, Pishgar M, Theis J, Zulueta J, Darabi H. Predicting clinical outcomes among hospitalized covid-19 patients using both local and published models. *BMC Med Inform Decis Mak*. 2021;21(1):224.
27. Cheng F-Y, Joshi H, Tandon P, Freeman R, Reich DL, Mazumdar M, Kohli-Seth R, Levin MA, Timsina P, Kia A. Using machine learning to predict icu transfer in hospitalized covid-19 patients. *J Clin Med*. 2020;9(6):1668.
28. Moulaei K, Shanbehzadeh M, Mohammadi-Taghiabad Z, Kazemi-Arpanahi H. Comparing machine learning algorithms for predicting covid-19 mortality. *BMC Med Inform Decis Mak*. 2022;22(1):1–12.
29. Choudhury NA, Mukherjee S, Singer T, Venkatesh A, Perez Giraldo GS, Jimenez J, Miller J, Lopez M, Hanson BA, Bawa AP, et al. Neurologic manifestations of long covid disproportionately affect young and middle-age adults. *Annals Neurol*. 2025;97(2):369–383.

30. Musolf AM, Holzinger ER, Malley JD, Bailey-Wilson JE. What makes a good prediction? feature importance and beginning to open the black box of machine learning in genetics. *Hum Genet*. 2022;141(9):1515–1528.
31. Zhou L, Pan S, Wang J, Vasilakos AV. Machine learning on big data: Opportunities and challenges. *Neurocomputing*. 2017;237:350–361.
32. Nasteski V. An overview of the supervised machine learning methods. *Horizons b*. 2017;4:51–62.
33. Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. *BMC Med Inform Decision Mak*. 2019;19(1):1–16.
34. Wikipedia contributors, Random forest — Wikipedia, the free encyclopedia, [Online; accessed 15-February-2024] (2024). URL: https://en.wikipedia.org/w/index.php?title=Random_forest&oldid=1203774559.
35. Saravanan R, Sujatha P. A state of art techniques on machine learning algorithms: a perspective of supervised learning approaches in data classification. *2018 Second international conference on intelligent computing and control systems (ICICCS)*. IEEE; 2018. 945–949.
36. Xing W, Bei Y. Medical health big data classification based on knn classification algorithm. *IEEE Access*. 2019;8:28808–28819.
37. Wikipedia contributors, Gradient boosting — Wikipedia, the free encyclopedia, [Online; accessed 15-February-2024] (2023). URL: https://en.wikipedia.org/w/index.php?title=Gradient_boosting&oldid=1190910306.
38. Gardner MW, Dorling S. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmos Environ*. 1998;32(14–15):2627–2636.
39. Durairaj M. Revathi, Prediction of heart disease using back propagation mlp algorithm. *Int J Scient Technol Res*. 2015;4:235–239. URL:https://api.semanticscholar.org/CorpusID:26834630.
40. Wikipedia contributors, Multilayer perceptron — Wikipedia, the free encyclopedia, [Online; accessed 15-February-2024] (2024). URL: https://en.wikipedia.org/w/index.php?title=Multilayer_perceptron&oldid=1195568237.
41. S.A. Alvarez, An exact analytical relation among recall, precision, and classification accuracy in information retrieval, Boston College, Boston, Technical Report BCCS-02-01 (2002) 1–22.
42. Eusebi P. Diagnostic accuracy measures. *Cerebrovasc Diseases*. 2013;36(4):267–272.
43. S. Raschka, An overview of general performance metrics of binary classifier systems, arXiv preprint arXiv:1410.5330 (2014).
44. L. Derczynski, Complementarity, F-score, and NLP evaluation, in: N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis (Eds.), Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), European Language Resources Association (ELRA), Portorož, Slovenia, 2016, pp. 261–266. URL: https://aclanthology.org/L16-1040.
45. Huang J, Ling CX. Using auc and accuracy in evaluating learning algorithms. *IEEE Trans Knowledge Data Eng*. 2005;17(3):299–310.
46. Mahajan P, Uddin S, Hajati F, Moni MA, Gide E. A comparative evaluation of machine learning ensemble approaches for disease prediction using multiple datasets. *Health Technol*. 2024;14(3):597–613.
47. Shekhar S, Bansode A, Salim A. A comparative study of hyper-parameter optimization tools*2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*. IEEE; 2021:1–6.
48. Liew C-H, Ong S-Q, Ng DC-E. Utilizing machine learning to predict hospital admissions for pediatric covid-19 patients (prepcovid-machine). *Scient Rep*. 2025;15(1):3131. https://doi.org/10.1038/s41598-024-80538-4.
49. Buenrostro-Mariscal R, Montesinos-López OA, Gonzalez-Gonzalez C. Predicting hospitalization in older adults using machine learning. *Geriatrics*. 2025;10(1). https://doi.org/10.3390/geriatrics10010006. URL: https://www.mdpi.com/2308-3417/10/1/6.
50. Gianfagna L, Cecco A. *Explainable AI with Python*. 2021. https://doi.org/10.1007/978-3-030-68640-6.
51. Molnar C. *Interpretable Machine Learning*. 2nd Edition,, 2022. URL: https://christophm.github.io/interpretable-ml-book.