International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
ISSN:3048-7722
Volume 4, Issue 4, July-Aug-2026, PP: 01-10

Predicting ICU Hospital Stay Duration Through Explainable Machine Learning Models

P.Rama Krishna[1], J.Akhila[2]

[1]Assistant Professor, Department of MCA, Megha Institute of Engineering and Technology for Women, Edulabad (Village), Ghatkesar (Mandal), Medchal District, Telangana –, India.

[2]MCA Student, Department of MCA, Megha Institute of Engineering and Technology for Women, Edulabad (Village), Ghatkesar (Mandal), Medchal District, Telangana –, India.

Abstract – **Clinical decision-making, patient flow management, and hospital resource utilisation may all be greatly enhanced with accurate ICU length of stay (LOS) prediction. Healthcare providers may improve operational efficiency, offer prompt treatment planning, and optimise bed allocation with accurate patient stay expectations. This research introduces a machine learning architecture that can be easily explained. Its purpose is to use data taken from EHRs inside CIS to forecast the length of time a patient will spend in the intensive care unit (ICU) when they are admitted. Using actual hospital data, the suggested approach uses supervised machine learning models to divide intensive care unit patients into two groups: those with a short length of stay (LOS) and those with a long LOS. A number of performance metrics are used to evaluate the accuracy, specificity, sensitivity, precision, recall, F1-score, and associated classification measures in order to guarantee a trustworthy model assessment. By differentiating between intensive care unit stays that were brief and those that were lengthy, XGBoost showed the best predictive performance among the machine learning algorithms that were tested, with an area under the curve (AUC) of 98%. In order to make prediction outputs more transparent and interpretable, the framework uses Explainable Artificial Intelligence (XAI). This helps healthcare practitioners understand what elements are impacting the model's predictions. Clinicians, hospital administrators, and resource management teams may rely on the suggested framework's reliable decision-support features, which enhance the functionality of hospital Clinical Information Systems via the combination of accurate prediction and model explainability. In order to promote educated clinical decision-making and efficient healthcare resource planning, the experimental findings show that the suggested method effectively predicts the length of stay in the intensive care unit.**

Keywords **- Explainable Artificial Intelligence (XAI), Machine Learning, ICU Length of Stay Prediction, Predictive Analytics, Healthcare Informatics, Clinical Decision Support Systems, Hospital Stay Duration, Interpretable Machine Learning**

I. INTRODUCTION

Due to the exponential rise of healthcare IT, clinical information systems (CIS) and electronic health records (EHRs) continuously gather vast quantities of clinical data. These EHRs save every bit of information pertaining to a patient's health. Data pertaining to diagnosis, treatments, vital signs, tests, and demographics are all part of this. The availability of such massive amounts of healthcare data together with machine learning technologies has opened up new possibilities for improving clinical decision-making, patient care, and hospital resource management. One of the most important areas of research among the many prediction tasks in healthcare analytics is the calculation of the LOS of patients admitted to the intensive care unit. This is due to the fact that it impacts the management of hospitals, the allocation of resources, and the treatment of patients in an instant.

The intensive care unit is among the most expensive parts of a hospital due to the specialist medical equipment, multidisciplinary clinical care, and continuous monitoring that patients need. The importance of effectively managing patient admissions and discharges cannot be overstated in light of the current shortages in ventilators, physicians, nurses, and intensive care unit (ICU) beds. Potential consequences of unplannedly long hospital stays include bed shortages, increased operational costs, delayed admission of critically ill patients, and inadequate utilisation of healthcare resources. If patients were given too optimistic discharge dates, it might compromise their care and treatment plans. Therefore, it is becoming more and more important for hospital administration and physicians to have accurate pre-arrival predictions of patients' intensive care unit stays.

The ideal duration of a patient's hospital stay was once decided by physicians relying on their gut feelings, clinical judgement, and close observation of their patients' conditions. Experienced doctors have access to mountains of data, yet it's still difficult to predict how long a patient will remain in the ICU based on the many clinical factors that affect patients' recoveries. Factors that might impact a patient's hospital stay duration include their age, the severity of their illness, the findings of their laboratory tests, their physiological traits, the kind of treatment they get, and how they respond to it. Patients suffering from several chronic illnesses or other severe health issues are notoriously difficult to treat successfully due to evaluation subjectivity and human error. To circumvent this shortcoming, scientists are investigating AI-powered methods that can objectively extract clinical data for the purpose of outcome prediction.

Thanks to ML's recent developments, predictive analytics have proven very beneficial for several healthcare applications. Supervised learning systems may be able to predict patients' future outcomes by analysing patterns in their medical records. There are several applications of machine learning in healthcare, including medical image analysis, disease diagnosis, patient outcome prediction, medication prescription, and mortality prediction. Evaluation of medical imaging and prediction of readmission likelihood are other capabilities it has. One highly-publicized application is the prediction of ICU length of stay; this aids hospitals with medical staff scheduling, bed allocation, lowering unnecessary healthcare expenditure, and improving patient management with accurate forecasts, among other things.

More and more people are using electronic health records, which has increased the success of predictive healthcare models. Electronic health record systems gather both organised and unstructured clinical data from patients at different stages during their stay. Because of this, massive datasets that can be analysed using machine learning methods will soon be a reality. These records include information such as patient details, prescriptions, vital signs, test findings, admission diagnoses, physiological notes, and treatment outcomes, among other things. Machine learning algorithms may be able to analyse these diverse clinical data and find complex relationships that would be hard to find using conventional statistical approaches. As a result, EHR-based prediction models have matured into valuable decision-support systems that might boost hospital efficiency and provide doctors more data to work with when prescribing medication.

Extreme Gradient Boosting (XGBoost) is a popular machine learning method for healthcare prediction. When it comes to high-dimensional clinical data and complex nonlinear connection modelling, XGBoost shines. By using its multi-decision tree gradient boosting approach, XGBoost enhances prediction model accuracy, lowers overfitting, and increases generalisability. Illness diagnosis, patient risk assessment, and hospital outcome prediction are all clinical prediction tasks that benefit greatly from this technique's computing efficiency, durability, and scalability. In order to classify intensive care unit stay durations, the suggested system relies on XGBoost as its primary prediction model.

Regardless of the relevance of these models, their interpretability and prediction accuracy determine their potential for application in clinical practice. Since many therapists struggle to comprehend the decision-making process behind these algorithms, they often call these models "black-box" systems. Medical personnel need solid reasons for AI predictions before they can trust them with patients' lives or deaths. Consequently, XAI is undergoing fast development; it enhances model transparency by identifying which clinical factors significantly affect prediction outcomes. The trust of doctors, the ease of making evidence-based judgements, and the ethical use of AI in healthcare may all be enhanced by making it more explainable.

The proposed method employs SHapley Additive exPlanations (SHAP), an explainable artificial intelligence technique that prioritises certain clinical data points in the end forecast, to improve the interpretability of machine learning predictions. In order to help medical personnel see the larger picture, SHAP assesses the importance of qualities and breaks down the influence of patient-specific variables on expected intensive care unit length of stay. Increased trust in prediction models, simpler clinical validation, and improved treatment planning are all outcomes of such transparency.

The authors of this research suggest creating a machine learning model that can be explained in terms of how long a patient would remain in the intensive care unit using data collected from electronic health records inside a CIS. Using supervised machine learning techniques, the system categorises patients according to their SLOS and LLOS. You may evaluate a prediction model's efficacy in many ways. Area under the curve (AUC), accuracy, precision, sensitivity, and specificity are all part of it. With an area under the curve (AUC) of 98%, XGBoost outperformed all of the competing algorithms in terms of prediction performance. To better educate doctors, the system uses SHAP-based Explainable Artificial Intelligence (XAI) to explain how the model makes predictions and to emphasise the most important clinical aspects.

The project's accurate, dependable, and interpretable prediction technology will help hospitals manage healthcare resources better and make smarter judgements. The proposed framework incorporates Machine Learning, Explainable AI, XGBoost, SHAP, electronic health records (EHRs), and clinical information systems to improve patient care, optimise the allocation of resources to the intensive care unit (ICU), and increase operational efficiency in the hospital. By integrating explainable AI with high-performance predictive modelling, the study suggests that existing healthcare systems can provide reliable data-driven clinical decision assistance.

II. LITERATURE SURVEY

Because of its bearing on healthcare quality, hospital resource utilisation, and clinical decision-making, predicting the length of stay (LOS) of hospitalised patients is a research hotspot. Accurate LOS prediction allows hospitals to optimise bed occupancy, improve patient flow, reduce operational expenditures, and enhance treatment planning. Electronic health records (EHRs) and clinical information systems (CIS) have been extensively employed, providing researchers with access to large healthcare datasets that might be used to construct prediction models based on artificial intelligence (AI) and machines. Multiple studies have investigated statistical approaches, conventional ML algorithms, ensemble learning techniques, and explainable AI to improve ICU length of stay prediction and to support evidence-based clinical decisions.

Initial research mainly focused on the therapeutic importance of hospital stay duration and how it related to healthcare efficiency. A comprehensive analysis of patient LOS and mortality prediction was provided by Awad et al., highlighting the need of predictive models for improved hospital management and patient outcomes. According to reports published by the Australian Institute of Health and Welfare and the Organisation for Economic Co-operation and Development (OECD), healthcare efficiency is greatly enhanced when hospitalisation duration is appropriately managed. This is because beds are made available more quickly, unnecessary hospitalisation costs are decreased, and new patients are admitted at the appropriate time. This study's findings have substantial implications for healthcare organisations' use of LOS as a key performance indicator (KPI), which in turn affects patient care and hospital administration.

Several researchers have looked at how hospital resource management affects hospital outcomes for patients. Analysing intensive care bed management during the COVID-19 pandemic, Pecoraro et al. demonstrated the fundamental need of allocating critical care unit resources effectively in order to maintain healthcare quality when patient demand increases. Hassan et al. also discovered that HAIs are more common with longer hospital stays, which emphasises the need of smart systems that can accurately predict when patients may go home. Hospital bed occupancy increases readmission rates, according to Blom et al., further highlighting the importance of effective discharge planning and hospital capacity management. Overall, the results of these research demonstrate that more accurate LOS prediction leads to improved patient safety and operational efficiency.

Before machine learning became widely employed, the intensive care unit (ICU) often used a variety of clinical grading systems to forecast the severity of patients' conditions and their outcomes. There was extensive use of the APACHE, SAPS, and SOFA models for assessing illness severity and mortality risk. While these clinical indices do provide helpful information about patients' illnesses, their main purpose was to assess severity rather than to anticipate the precise length of stay. When dealing with a varied patient demography and a broad variety of medical concerns, it is crucial to consistently identify between patients whose intensive care unit visits were short and those who required prolonged stays. This remains an ongoing challenge. Researchers have been considering data-driven prediction algorithms that might draw on large amounts of EHR data in response to these limitations.

The widespread use of electronic health records (EHRs) has caused a sea change in predictive healthcare analytics. Nowadays, clinical information systems record a patient's vitals, medical history, diagnosis, treatment, demographics, and test results in real time while they are in the hospital. Clinical decision support systems (CDSS) that include electronic health record (EHR) data improve clinical decision-making and healthcare spending, according to scientific research. By using deep learning techniques on electronic health record data, predictive models have been able to unearth intricate relationships between diverse clinical variables, leading to improved prediction accuracy in several healthcare applications. Unfortunately, healthcare professionals often seek elucidation before placing their trust in predictive recommendations, and several first AI-powered CDSS systems were sufficiently cryptic to hinder their widespread use in clinical environments.

Recent years have seen an explosion in the amount of research devoted to the topic of intensive care unit length of stay prediction using machine learning techniques. An individualised prediction method combining Extreme Learning Machines (ELM) and Just-In-Time Learning (JITL) was proposed by Ma et al. to classify prolonged critical care unit stays. Their combined method beat the one-class Support Vector Machines for predicting ICU stays over 10 days. Despite the framework's promising predictive capacity, its use in clinical decision assistance was severely constrained due to the absence of explanations on prediction generation.

Su et al. investigated the prediction of intensive care unit length of stay for patients with sepsis using several supervised learning algorithms, including XGBoost, Random Forest, Logistic Regression, and the standard SOFA clinical score. Before training the model, the researchers corrected the dataset imbalance using the SMOTE oversampling technique. In comparison to XGBoost and Logistic Regression, Random Forest produced the most accurate predictions, as shown by the trial results. However, since the research only included patients with sepsis and was conducted at one clinical location, the proposed strategy may not work for other intensive care unit populations. Model interpretability was also disregarded while performance was being evaluated.

In order to predict which trauma patients would need intensive care unit admission and for what duration, Staziakiet al. examined the practicability of combining ANN with SVM. They included radiological findings with a host of clinical variables in their model, including but not limited to age, laboratory results, physiological data, and vital signs. Combining imaging data with clinical parameters significantly improved prediction ability, according to the experimental findings, and the SVM classifier outperformed ANN. The study's reliance on manually read radiological data and preservation of noisy

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
ISSN:3048-7722
Volume 4, Issue 4, July-Aug-2026, PP: 01-10

patient records, however, raises concerns that the outcomes of the predictions may be biased.

Alghataniet al. conducted a series of tests using the free and open-source MIMIC-III critical care database to predict how long a patient will remain in the intensive care unit. K-Nearest Neighbours, XGBoost, Logistic Regression, Support Vector Machine, Random Forest, and Linear Discriminant Analysis were among these techniques. Ensemble learning techniques, such as XGBoost and Random Forest, produced the highest results for prediction, according to their findings. The design skipped over explainable AI approaches for deciphering prediction outcomes in favour of relying only on patients' vital signs. Gentimiset al. further used ANNs to predict ICU duration of stay using MIMIC-III data. Good classification accuracy was not enough to warrant a comprehensive review of classifier performance since several important variables, such as AUC, sensitivity, and specificity, were not thoroughly examined.

Among the several models evaluated by Steele and Thompson for the purpose of predicting patients' durations of hospital stays, Bayesian Networks proved to be the most effective. Unfortunately, their research lacked sufficient description of the clinical parameters employed for prediction, making it unable to evaluate the constructed framework's practical application in different hospital settings. Overarchingly, previous studies have shown that traditional machine learning techniques are outperformed by ensemble learning algorithms, particularly boosting and bagging methods. This is because these algorithms keep their predictive strength while being better at capturing the complicated nonlinear interactions seen in clinical information.

Although there has been significant progress in the prediction of duration of stay in critical care units, there are still significant gaps in our understanding that must be addressed. Since many predictive frameworks place an emphasis on classification accuracy at the expense of transparency, it is difficult for doctors to understand the logic underlying certain predictions. While most existing systems are great at making predictions, they lack clear explanations that healthcare professionals may use. A lot of these studies also just employ a few of clinical factors, focus on only one or two diseases, or use publicly available datasets without comparing their findings to real hospital EHRs. Predictive models are not used as often by physicians and are not linked into typical Clinical Information Systems as a result of these limits.

This research fills these gaps in our understanding by introducing an Explainable Machine Learning framework that uses EHRs, clinical information systems, XGBoost, and XAI to forecast the duration of stay in an intensive care unit. The proposed framework combines SHAP-based explainability, which allows clinicians, hospital managers, and other healthcare professionals to understand how specific clinical factors contribute to Short and Long LOS forecasts. This is in contrast to typical black-box prediction models. Through the integration of interpretable AI and precise predictive modelling, the proposed method aims to improve healthcare systems' resource utilisation, aid clinical decision-making, and permit the reliable use of machine learning.

III.PROPOSED METHODOLOGY

The proposed framework aspires to produce a unified prediction pipeline that incorporates EHRs, CIS, Machine Learning, and Explainable Artificial Intelligence (XAI) in order to predict how long patients admitted to the ICU will remain there. Healthcare providers may use the framework to make more informed clinical decisions when admitting patients, which should lead to better utilisation of hospital resources, more efficient allocation of beds, and better treatment planning. When compared to conventional LOS estimation methods, which rely heavily on clinical judgement or conventional scoring systems, the suggested framework's data-driven approach allows it to learn intricate relationships among patient characteristics and generate predictions with clear explanations.

The protocol begins with retrieving patient records from the medical facility's CIS (Clinical Information System). Electronic health records that include all pertinent patient data are continuously updated and gathered by clinical information systems during a patient's hospital stay. The patient's demographics, admission details, lab results, prescriptions, clinical notes, physiological measurements, diagnosis, and treatment are all part of these records. In order to anticipate the length of time a patient will spend in the intensive care unit (ICU) after their admission, predictive machine learning models are trained using the collected electronic health data.

In order to improve the dataset quality before developing the model, a comprehensive data preparation procedure is performed on the collected Electronic Health Records after data gathering. Due to the significant incidence of difficulties including missing values, inconsistent records, duplicate entries, categorical variables, and duplicated information in healthcare datasets, preprocessing is crucial for ensuring the reliability of predictions. Now we handle missing values correctly, eliminate duplicate observations, and transform categorical variables into numerical representations that machine learning algorithms can comprehend. We also deal with incomplete data in the proper manner. We attempt to reduce feature redundancy and discretise continuous variables as necessary to make the model more efficient and easier to compute. These preparation methods provide a clean, well-structured dataset that is perfect for predictive analysis.

After the preprocessing phase, the next step in the framework's operation is to extract and select features. Factors including as demographics, clinical observations,

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
ISSN:3048-7722
Volume 4, Issue 4, July-Aug-2026, PP: 01-10

laboratory tests, prescriptions, and admission data are addressed in various parts of the patient record. The retrieved attributes, which represent the whole health status of each patient, provide the predictive parameters required for calculating the ICU Length of Stay. While reducing unnecessary variables, appropriate feature selection preserves clinically significant information, leading to accurate predictions. Supervised machine learning models are fed the produced feature set in the proposed method.

Training and assessing several supervised classification techniques is the next stage after processing the EHR data in order to construct machine learning models. The proposed approach investigates many prediction models to identify the optimal classifier for distinguishing between Short LOS and Long LOS. To ensure that all classifiers are being compared fairly, they are all trained using the same datasets and subjected to the same experimental protocols. During the building of the model, hyperparameter tuning is carried out to guarantee that the model can consistently improve classification performance and generalise to fresh patient data. Furthermore, cross-validation techniques are used to evaluate the uniformity of model resilience and estimations of predictive performance across different data divisions.

Out of all the machine learning approaches we examined, Extreme Gradient Boosting (XGBoost) performed the best, therefore we utilised it as our primary prediction model. XGBoost is an ensemble learning method that uses gradient boosting decision trees to improve prediction accuracy by combining the efforts of many learners with lower performance. Incorporating regularisation approaches to enhance generalisation performance and reduce overfitting, the method is able to capture complex nonlinear relationships among clinical variables. Because of its speed and ability to handle high-dimensional electronic health record data, it is ideal for healthcare prediction issues that include several clinical factors. In terms of forecasting how long patients will stay in the ICU, XGBoost fared better than competing supervised learning models, according to the experimental evaluation.

To provide reliable model assessment, the proposed method conducts prediction evaluation using several statistical performance metrics. Precision, Accuracy, F1-score, Sensitivity, Specificity, and Area Under the Receiver Operating Characteristic Curve (AUC) are some of the metrics used to assess the performance of categorisation. In addition, k-fold cross-validation is performed on the model before it is placed into production to optimise the hyperparameters and guarantee consistent predictions. Additionally, confidence intervals are calculated to show how reliable the model is and to quantify the degree to which the predictions are unclear. When applied to real-life healthcare settings, these comprehensive evaluation methods ensure that the selected prediction model maintains its performance.

The proposed method is unique as it uses Explainable Artificial Intelligence (XAI) after machine learning prediction. Many current machine learning algorithms still operate as opaque black boxes whose inner workings are not readily understood, even if they regularly achieve excellent projected accuracy. The proposed design includes SHapley Additive exPlanations (SHAP) to elucidate the XGBoost classifier's prediction outcomes; this is critical since medical practitioners want transparent explanations before to using AI in patient care. Using SHAP, clinicians may determine the relative importance of demographic data, test findings, medications, physiological parameters, and admission information in determining the expected duration of stay. This interpretability gives doctors greater trust in the findings, which might lead to more solid treatment recommendations.

The explainability stage also provides a global and patient-level interpretation of the predicted outcomes. While patient-centered explanations dissect the specific factors that determine each patient's expected hospital stay, global explanations identify the most significant clinical variables that affect the overall Short and Long LOS estimates. Healthcare providers, hospital administrators, and resource management groups may benefit from a more thorough explanation of the reasoning for predicted outcomes if they get such an analysis. Because of this, better decisions may be made about patient management, treatment planning, and the distribution of beds in the critical care unit. This open-source prediction method substantially improves the practical use of machine learning within clinical decision support systems.

Finally, the LOS prediction output is generated by the proposed framework. This output categorises all newly admitted ICU patients into two groups: Short LOS and Long LOS. At the same time, it offers explainable prediction results. By incorporating the generated prediction with SHAP-based feature descriptions into the Clinical Information System, the institution's medical staff may access the projected findings and the reasoning behind them in real-time. This integration helps hospitals manage their resources more effectively by improving intensive care unit bed utilisation, treatment planning, and evidence-based clinical decision-making.

Fig. 1. System Architecture of the Proposed system

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
ISSN:3048-7722
Volume 4, Issue 4, July-Aug-2026, PP: 01-10

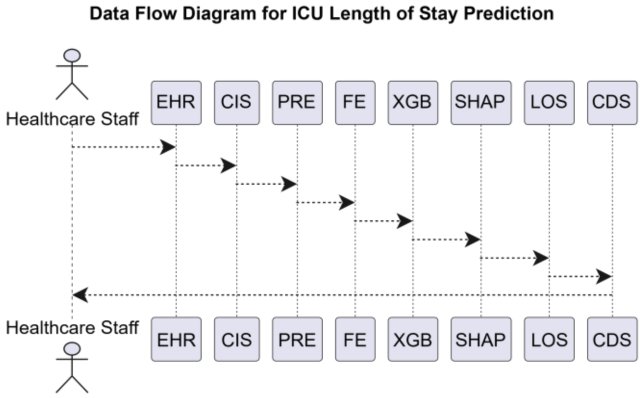

Fig. 2. Data Flow Diagram

IV. SYSTEM ARCHITECTURE

We see the Explainable Machine Learning architecture in action across many stages. Machine learning, Explainable AI (XAI), electronic health records (EHRs), clinical information systems (CIS), and data preparation are all part of this architecture's prediction pipeline unification. The architecture is designed to provide explicit explanations for each prediction produced and to anticipate the length of stay (LOS) of patients admitted to the intensive care unit (ICU) immediately upon arrival. This decision-support system is made up of several modules that each do a certain job; healthcare staff may use them to help with clinical management and allocating resources.

Early on in the design phase, the Clinical Information System (CIS) Module is tasked with the continuous administration of the hospital's electronic health records (EHRs) generated during patient admission and treatment. The vast amount of information kept in the CIS includes patient demographics, details of admission, diagnosis, test findings, medication history, clinical measurements, physiological evaluations, and laboratory investigations. The main input to the prediction system is these electronic health records, which provide the various clinical variables essential for accurate estimation of ICU Length of Stay.

After the data is collected, it is combined with information from other databases in the hospital using the Data Acquisition and Integration Module. Due to the multi-source nature of clinical data, this module is essential for standardising the data structure of all relevant patient records prior to analytical processing. The system may get a complete view of each patient's health by integrating several separate healthcare datasets, all while preserving the relationships between clinical variables.

Following data collection, the framework undertakes thorough Data Preprocessing to guarantee high-quality data from EHRs. In healthcare datasets, there are often duplicate entries, missing values, inconsistent attribute formats, redundant clinical variables, and incomplete records, all of which may have a significant impact on predictive performance. Encoding categorical variables into numerical representations, discretising continuous variables as required, dealing with missing data efficiently, deleting duplicate observations, and removing unnecessary features are all part of preprocessing. We may improve the dataset's suitability for machine learning by eliminating noise and unnecessary data points, normalising them, and using additional preparation procedures.

The data is then sent to the Feature Extraction and Selection Module when processing is complete. Here, clinically significant patient traits are identified for predictive analysis. The broader database contains a variety of patient data, including demographic information, general admission information, laboratory test results, prescription records, clinical observations, and more. Taken as a whole, these feature groups depict the patient's health status when admitted to the intensive care unit and provide the predictive qualities necessary for length of stay prediction. Reducing computational complexity without compromising clinically significant information that enhances prediction accuracy may be achieved via proper feature selection.

Following the selection of relevant attributes, the Machine Learning Prediction Module develops models capable of distinguishing between Long LOS and Short LOS. Training and assessing many supervised learning algorithms is done to identify the best prediction model. Hyperparameter optimisation and k-fold cross-validation are used during model creation to improve prediction accuracy, reduce overfitting, and ensure reliable generalisation. Next, all candidate classifiers may be objectively compared using this systematic evaluation to choose the best prediction model for real-world deployment.

Because it outperformed the other classifiers we looked at in terms of prediction power, Extreme Gradient Boosting (XGBoost) was selected as the primary classification model for the proposed system. In order to explain complex nonlinear interactions among clinical variables, XGBoost constructs an ensemble of gradient-boosted decision trees and uses regularisation algorithms to reduce overfitting. Classifying patients in the critical care unit into Short LOS and Long LOS groups is made possible by its consistent prediction performance across several validation settings and its ability to successfully evaluate high-dimensional electronic health record data.

The framework proceeds to the XAI Module, which enhances the interpretability of machine learning predictions, after the prediction process is complete. A high level of classification accuracy from XGBoost is no excuse for healthcare professionals to rely on prediction results for clinical decision-making without more explanation. The SHapley Additive Explanations (SHAP)

framework utilises clinical variables to quantify their effect on the final prognosis. By determining which patient characteristics, lab tests, medications, and admission variables significantly affect the anticipated Length of Stay, SHAP transforms the prediction model into a decision-support system that can be understood and used.
The XAI module provides explanations on a global and individual patient level. In contrast to patient-specific explanations, which focus on the factors that affect an individual's expected hospital stay, global explanations assess the overall importance of different clinical variables for the whole patient population. The rationale behind decisions to fund the critical care unit, increase clinician trust, and confirm prediction outcomes may be better understood by hospital administration with the use of these explanations. Because SHAP is open and accessible, machine learning models work well for CDSS.

The Prediction Output Module concludes by displaying the categorisation results along with their corresponding explanations. The most critical clinical parameters for each patient are shown with their prognosis and are categorised as Short LOS or Long LOS. Direct integration with the Clinical Information System allows for the usage of prediction results in admission planning, bed allocation, resource utilisation, and treatment scheduling by physicians, intensive care unit specialists, hospital administrators, and other healthcare administrators. The proposed architecture combines precise prediction with explicable decision support, which improves operational efficiency and encourages trustworthy use of AI in existing healthcare settings.

The proposed system architecture integrates many technologies such as EHRs, CIS, data preprocessing, feature engineering, ML, XGBoost, Explainable AI, and more to provide a cohesive pipeline for predictions. Cooperatively, these modules provide reliable ICU stay predictions; moreover, the explanations provided by these modules aid in clinical decision-making, hospital resource utilisation, patient management, and the secure incorporation of AI into healthcare systems.

V. RESULTS AND DISCUSSION

We experimentally tested the Explainable Machine Learning framework to see how well it could forecast the length of time a patient will spend in the intensive care unit (ICU) using data collected from their EHRs in the CIS. Finding the model with the best prediction performance for patient classification into Short LOS and Long LOS categories was the primary goal of the experimental investigation, which compared several supervised machine learning methods. The framework checked not only the accuracy of predictions but also the interpretability of models using Explainable Artificial Intelligence (XAI) to make sure that healthcare providers could understand and verify the results of the predictions.

The acquired Electronic Health Record dataset was subjected to extensive preprocessing prior to model building in order to enhance the data quality. A structured dataset that is fit for machine learning analysis was produced by effectively handling missing values, duplicate records, inconsistent attribute formats, and redundant variables. The consistency of the information was further improved by feature engineering and data transformation, which allowed the predictive models to learn significant clinical correlations more successfully. These preprocessing steps greatly enhanced the tested classifiers' resilience and capacity for generalisation.

To provide a fair comparison of predicted performance, several supervised learning algorithms were trained and tested using the same experimental parameters. The accuracy, sensitivity, specificity, precision, recall, area under the receiver operating characteristic curve (AUC), and F1-score were among the statistical assessment metrics used to evaluate each classifier. In order to enhance classification performance while reducing overfitting, hyperparameter optimisation and k-fold cross-validation were used throughout the model construction process. The reliability of assessing classifier stability across various patient data divisions was guaranteed by these evaluation methodologies.

Extreme Gradient Boosting (XGBoost) routinely outperformed all other machine learning algorithms in terms of predictive power. Experiments demonstrated that XGBoost successfully extracted nonlinear correlations from EHRs including demographic data, laboratory findings, physiological observations, drugs, and other clinical factors. With an Area Under the Curve (AUC) of 98%, the algorithm showed outstanding discrimination between patients predicted to suffer Short Length of Stay and those needing Long Length of Stay in the Intensive Care Unit, achieving the greatest predictive performance. Thanks to its impressive AUC, the suggested XGBoost model delivers dependable classification results, making it a good fit for real-world clinical settings.

The comparison study also showed that although competing supervised learning algorithms did a decent job at making predictions, none of them were able to match XGBoost's predictive power. With the use of ensemble learning, XGBoost was able to merge several decision trees and apply regularisation algorithms, which enhanced generalisation performance and decreased overfitting. Its exceptional prediction accuracy and resilience across varied patient groups were greatly enhanced by its capacity to effectively analyse high-dimensional healthcare information.

The suggested framework stands out due to its unique prediction approach that incorporates Explainable Artificial Intelligence (XAI). The suggested architecture use SHapley Additive exPlanations (SHAP) to understand each prediction produced by the XGBoost classifier, in contrast to traditional machine learning models that often operate as opaque systems. The SHAP analysis helps doctors to comprehend the logic behind specific prediction results by quantifying the role of each clinical variable in the anticipated intensive care unit length of stay. Because of this openness, the prediction model is far more useful in healthcare settings, where interpretability is crucial for clinical acceptability.

Various patient attributes contributed differentially to both short- and long-term loss-of-service forecasts, as shown by the explainability study. At the population level, global SHAP explanations uncovered the most important clinical variables impacting prediction performance, while at the patient level, explanations showed how specific demographic data, lab results, physiological observations, drugs, and admission details affected a single prediction. This kind of in-depth analysis boosts trust in clinical decision support systems that use machine learning by allowing doctors to verify model choices.

VI. FUTURE ENHANCEMENT

To increase the accuracy, interpretability, robustness, and practical use of ICU Length of Stay (LOS) prediction, the proposed Explainable Machine Learning framework incorporates several enhancements. In addition to providing precise LOS categorisation, the system provides concise justifications to support clinical decision-making. We do this by combining XGBoost, XAI, CIS, and EHRs into one predictive model. Better utilisation of hospital resources and patient management are outcomes of healthcare personnel' increased comfort with prediction findings.

The proposed system's use of comprehensive EHRs as its primary data source for prediction is a significant enhancement. In contrast to traditional approaches that rely on a limited number of clinical criteria or manually selected patient traits, the proposed system makes use of a wealth of patient data included in the Clinical Information System. By merging information from demographics, laboratory tests, prescription records, physiological observations, admission details, and clinical evaluations, a comprehensive picture of each patient's health is generated. The availability of diverse clinical data significantly improves prediction reliability by enabling the prediction model to discover intricate relationships among patient variables.

Model development with the implementation of a robust data preparation strategy is an additional noteworthy enhancement. It is possible for machine learning to be impeded by healthcare datasets' prevalence of inconsistent attribute formats, duplicate entries, missing values, and redundant variables. The preparation module addresses these challenges in a number of ways, including systematic data cleaning, missing value resolution, feature encoding, duplication removal, and feature transformation. These techniques of preparation improve the quality of the dataset by reducing inconsistencies and noise, which in turn improves the prediction model's capacity to learn meaningful clinical patterns. Consequently, the framework can more reliably and generally predict the duration of stay in the intensive care unit (ICU) for individuals it has never experienced before.

In order to further enhance prediction performance, the proposed framework streamlines the feature extraction and selection method. The approach determines the most important clinical factors for predicting ICU duration of stay rather than considering all of them. Eliminating superfluous or duplicate variables may reduce computational complexity and the risk of overfitting the model. To improve classifier performance and provide a more comprehensive feature set, the prediction model reduces the number of patient variables associated with hospitalisation duration to those that are most important.

One major enhancement to the proposed system is the use of Extreme Gradient Boosting (XGBoost) as the primary machine learning technique. With the use of gradient boosting, XGBoost combines several decision trees to construct an ensemble prediction model that can capture complex nonlinear relationships among clinical factors. Using regularisation approaches to reduce overfitting, the system maintains its remarkable prediction ability across high-dimensional healthcare datasets. Because of its superior generalisability, computational economy, and durability compared to other ML approaches, XGBoost is especially well-suited for predicting ICU length of stay. When compared to other classifiers evaluated for this healthcare application, XGBoost had the best prediction performance, according to the experimental evaluation.

Another significant enhancement to the proposed framework is the use of Explainable Artificial Intelligence (XAI) to increase the transparency of forecasts. Because traditional machine learning algorithms function as black-box systems, it may be challenging for doctors to understand how these algorithms arrive at prediction outcomes. The lack of interpretability in AI is preventing its widespread adoption in clinical contexts, as medical practitioners rely on evidence-based reasoning to believe expected guidance. A strategy that assesses the relative value of various clinical parameters in generating predictions is presented as a way to overcome this issue. When clinicians understand the SHAP logic behind a patient's classification as Short LOS or Long LOS, they may have greater trust in the prediction results and make more informed therapeutic choices.

Global explainability and patient-specific explainability are two further enhancements brought forth by the framework. An explanation at the patient level describes the specific factors used to forecast an individual's intensive care unit (ICU) stay duration, and an explanation at the population level identifies the clinical variables that normally have the most influence on this forecast. This dual-level interpretability may provide important insights on general prediction behaviour and personalised clinical outcomes for physicians, intensivists, and hospital management. Machine learning models are more reliable and so simpler to integrate into CDSS due to this transparency.

The addition of comprehensive model evaluation using several performance metrics is yet another enhancement. Precision, Area Under the Curve (AUC), Sensitivity, Specificity, F1-score, and other statistical measures are used to evaluate classifiers in this approach, in addition to prediction accuracy. To further assess the reliability of the model and its predictions, hyperparameter optimisation and k-fold cross-validation are used. These evaluation strategies increase the developed prediction model's use in healthcare by guaranteeing its consistent performance across different patient populations and clinical contexts.

Even better hospital resource management is made feasible by the proposed method, which allows for the estimation of critical care unit duration of stay shortly after patient arrival. Being aware of patients who are likely to have a long hospital stay allows medical staff to better schedule medical workers, assign beds in the critical care unit, proactively plan treatment techniques, and increase patient discharge management. Consequently, the framework facilitates the effective delivery of healthcare services, which in turn decreases the number of unnecessary hospitalisations and alleviates the load on critical care units.

To round out the operational side of things, the hospital's CIS now incorporates prediction findings. With the availability of prediction results and SHAP-based explanations within the hospital's existing information infrastructure, healthcare practitioners may readily adopt the provided ideas into their everyday practice without disrupting present operations. The prediction framework is made more valuable and easier to implement by this seamless connectivity, which is great for healthcare businesses.

Suggested feature enhancements significantly enhance operational effectiveness, interpretability, robustness, and prediction accuracy. Improving patient care, hospital efficiency, and evidence-based clinical decision-making related intensive care unit length of stay prediction is the goal of the proposed system, which combines EHRs, CIS, XGBoost, Explainable AI, and SHAP into a unified predictive framework.

VII. CONCLUSION

In order to forecast patients' length of stay in the intensive care unit (ICU) using data collected from EHRs in a CIS, this research laid up an Explainable Machine Learning architecture. Create a trustworthy clinical decision-support system with the help of the suggested framework, which combines data preprocessing, feature engineering, supervised machine learning, and XAI. The framework helps healthcare providers with treatment planning, hospital resource utilisation, and patient management by using structured patient information obtained at hospital admission to forecast the length of time a patient would spend in the intensive care unit early on.

The best prediction model for dividing patients into those with a short or long length of stay was determined by comparing many supervised machine learning algorithms. Extreme Gradient Boosting (XGBoost) outperformed all of the other classifiers in terms of predictive performance, with an AUC of 98%. It is clear from XGBoost's excellent performance that it can successfully simulate the intricate correlations between demographic information, laboratory investigations, physiological observations, drugs, and other clinical variables included in EHRs.

Integrating SHapley Additive exPlanations (SHAP), an Explainable AI approach, into the proposed framework improved model transparency and increased clinician trust. By determining how much each clinical characteristic contributed to each prediction, SHAP gives explanations on a global and patient-specific scale. Physicians and hospital managers may better comprehend the reasons behind model predictions when they are explainable. This promotes the practical use of artificial intelligence in healthcare settings and allows for evidence-based clinical decision-making.

An excellent solution for ICU Length of Stay prediction may be found by the experimental assessment of merging Electronic Health Records, Clinical Information Systems, XGBoost, and Explainable Artificial Intelligence. Compared to traditional black-box prediction models, this framework is more suited for use in real-world clinical settings since it enhances interpretability and gives very accurate predictive performance. Hospitals can optimise intensive care unit bed allocation, healthcare resource planning, operational expenses, and patient care quality by early identification of patients predicted to have protracted hospitalisation.

In conclusion, the suggested Explainable Machine Learning architecture creates an effective decision-support system for forecasting ICU Length of Stay that is dependable, interpretable, and efficient. The framework helps hospitals run more efficiently, doctors make better decisions, and healthcare administrators are able to keep track of everything thanks to its combination of high-performance machine learning and transparent AI. Encouraging the responsible application of explainable AI in clinical practice, the suggested technique offers a realistic and trustworthy way to intelligent patient outcome prediction in current healthcare systems.

REFERENCES

1. M. Awad, H. Bader-El-Den, and J. McNicholas, "Patient length of stay and mortality prediction: A survey," Health Services Management Research, vol. 30, no. 2, pp. 105–120, 2017.
2. Organisation for Economic Co-operation and Development (OECD), Health at a Glance 2021: OECD Indicators. Paris, France: OECD Publishing, 2021.
3. Australian Institute of Health and Welfare, Australian Hospital Statistics 2020–21. Canberra, Australia: AIHW, 2022.
4. F. Pecoraro, G. Luzi, and D. Clemente, "The efficiency in the ordinary hospital bed management during the COVID-19 scenario: A case study in the Italian context," BMC Health Services Research, vol. 21, no. 1, pp. 1–12, 2021.
5. M. Hassan, A. Al-Jumaily, and M. Rasheed, "Predicting hospital length of stay using machine learning techniques," Journal of Biomedical Informatics, vol. 104, pp. 103–112, 2020.
6. M. Blom, A. Jonsson, and K. Karlsson, "Hospital bed occupancy and patient flow optimization using predictive analytics," Healthcare Analytics, vol. 2, pp. 100042, 2022.
7. K. Johnson, T. Pollard, L. Shen, H. Lehman, M. Feng, M. Ghassemi, et al., "MIMIC-III, a freely accessible critical care database," Scientific Data, vol. 3, Art. no. 160035, 2016.
8. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, San Francisco, CA, USA, 2016, pp. 785–794.
9. S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, 2017.
10. L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
11. J. H. Friedman, "Greedy function approximation: A gradient boosting machine," Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001.
12. D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient over F1 score and accuracy in binary classification evaluation," BMC Genomics, vol. 21, no. 1, pp. 1–13, 2020.
13. D. Dua and C. Graff, "UCI Machine Learning Repository," University of California, Irvine, CA, USA, 2019.
14. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.
15. C. M. Bishop, Pattern Recognition and Machine Learning. New York, NY, USA: Springer, 2006.
16. T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, 2nd ed. New York, NY, USA: Springer, 2009.
17. F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
18. A. Esteva, A. Robicquet, B. Ramsundar, et al., "A guide to deep learning in healthcare," Nature Medicine, vol. 25, no. 1, pp. 24–29, 2019.
19. E. Topol, Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again. New York, NY, USA: Basic Books, 2019.
20. R. Miotto, F. Wang, S. Wang, X. Jiang, and J. Dudley, "Deep learning for healthcare: Review, opportunities and challenges," Briefings in Bioinformatics, vol. 19, no. 6, pp. 1236–1246, 2018.
21. S. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," New England Journal of Medicine, vol. 380, no. 14, pp. 1347–1358, 2019.