ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

ESTIMATING THE LENGTH OF HOSPITAL STAY USING EXPLAINABLE MACHINE LEARNING

M.Anitha[1],M.Priyanka[2],S.Lakshmi Sriya[3]

#1 Assistant Professor & Head of Department of MCA, SRK Institute of Technology, Vijayawada.

#2 Assistant Professor in the Department of MCA,SRK Institute of Technology, Vijayawada.

#3 Student in the Department of MCA, SRK Institute of Technology, Vijayawada.

Abstract: In hospitals, effective bed management reduces expenses while enhancing patient outcomes. This study uses data from electronic health records (EHRs) to propose a prediction paradigm for ICU length of stay (LOS) at admission. The study assesses many machine learning methods, such as Logistic Regression, Random Forest, MLP, Gradient Boosting, XGBoost, and an extension using CatBoost, using the hospital stay dataset from the Kaggle repository. AUC, accuracy, precision, recall, and F1-score are used to evaluate the algorithms. Among the conventional algorithms, XGBoost had the best accuracy, but the extended CatBoost method beat them all with an accuracy of 98.25%. The interpretation of feature contributions was done using Explainable AI (XAI) techniques like SHAP. The study shows how advanced machine learning models and patient EHR data may be used to forecast intensive care unit stays, improving hospital resource allocation.

Keywords: Hospital bed management, ICU length of stay, electronic health records (EHR), machine learning, XGBoost, CatBoost, SHAP, explainable artificial intelligence (XAI), prediction models, healthcare analytics.

INTRODUCTION

In medical institutions, the length of stay serves as a common indicator of efficacy. It has a significant effect on healthcare expenses and resource allocation. According to a study conducted by the Australian National Health Performance Authority, shorter hospital stays are seen to be more effective, which makes it easier to quickly make beds available for new patients. However, overly brief stays might compromise the standard of care and have a detrimental effect on the patient. On the other hand, prolonged hospital stays—often brought on by complications—may raise the risk of adverse health consequences. Hospital stays may be prolonged by delays in healthcare coordination that are unrelated to the patient's clinical condition. According to the poll, delays in moving patients to alternative care providers—like rehabilitation facilities, community care initiatives, or senior care facilities—may result in longer stays. Addressing intensive care unit (ICU) problems such patient congestion, infections, mortality risk, and medical difficulties requires effective management of hospital bed availability. A shorter ICU stay combined with excellent care is crucial to reducing these risks and improving resource use, especially during unforeseen events like pandemics. This ensures better patient outcomes in addition to lowering hospital costs. Therefore, maintaining the quality of treatment depends on having enough beds and transferring patients to other wards as soon as possible.

LITERATURE SURVEY

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

2.1 An explainable machine learning framework for lung cancer hospital length of stay prediction:

https://www.nature.com/articles/s41598-021-04608-7

ABSTRACT: This study presents a machine learning (ML) model-based prediction Length of Stay (LOS) framework for patients with lung cancer. The proposed system uses electronic health records (EHR) to handle unbalanced datasets for classification-based methods. With the MIMIC-III dataset, we have applied supervised machine learning techniques to forecast the length of stay (LOS) of patients with lung cancer while they are in the intensive care unit. Throughout the three framework phases, the Random Forest (RF) Model performed better than the other models and produced the expected outcomes. Over-sampling techniques (SMOTE and ADASYN) produced the greatest AUC findings (98% with CI 95%: 95.3–100%, and 100%, respectively) when clinical importance characteristics were selected. The second-highest AUC findings (98%, with CI 95%: 95.3–100%) and 97%, with CI 95%: 93.7–100% SMOTE-Tomek and SMOTE-ENN, respectively) were obtained by combining over-sampling with under-sampling. For both (ENN and Tomek-Links), under-sampling techniques produced the least significant AUC findings (50%, with CI 95%: 40.2–59.8%). In order to identify the most important clinical characteristics that helped the RF model predict lung cancer LOS, we used the SMOTE class balancing approach to explain the results of the RF using the machine learning explainable technique known as SHAP. Our promising architecture enables us to forecast lung cancer admissions to intensive care units (ICUs) using machine learning techniques in-hospital clinical information systems.

2.2 Predictive analytics framework for electronic health records with machine learning advancements: optimising hospital resources utilisation with predictive and epidemiological models:

https://researchdirect.westernsydney.edu.au/islandora/object/uws:67523/

ABSTRACT: This thesis's main goal was to examine the viability and resilience of predictive machine-learning models in relation to enhancing the use of hospital resources through data-driven strategies and forecasting hospitalisation using hospital quality assessment metrics like length of stay. The validity of the suggested methodological predictive framework on the data source of each hospital's electronic health records is part of the length of stay forecasts. Electronic health records (EHRs) were used in this thesis to power a data-driven prediction inpatient length of stay (LOS) research methodology that is appropriate for the most resource-demanding medical settings. The feasibility of methodological predicting length of stay techniques in dynamic and demanding hospital settings, including emergency rooms and intensive care units, was the main emphasis of the thesis. The thesis took into account (external) factors that are beyond hospital control, such as predicting future hospitalisations due to the spread of infectious communicable diseases during pandemics, even though the hospital length of stay predictions are based on (internal) healthcare inpatient outcomes assessment at the time of admission to discharge. The primary contributions of the thesis are the internal and external splits. As a result, the thesis assessed public health interventions during uncertain times (like pandemics) and quantified the impact of non-pharmaceutical interventions during outbreaks on subsequent hospitalised patients. To our knowledge, this method is the first in the literature to use simulation models to study the impact of epidemiological curves on future hospitalisations, which have a significant potential to affect the availability of hospital beds and put stress on hospital staff and workflow. The value of ensemble learning models in the context of LOS for hospital resource utilisation is the primary study similarity among chapters. By merging many base models to create an ideal predictive model, the ensemble learning models forecast improved predictive performance. Using data-driven methods, these prediction models investigated the internal LOS for a range of acute and chronic illnesses in order to identify the most potent and accurate predicted outcomes. For healthcare personnel working in hospital environments, this ultimately aids in achieving desired results.

2.3 Machine learning combining CT findings and clinical parameters improves prediction of length of stay and ICU admission in torso trauma:

https://link.springer.com/article/10.1007/s00330-020-07534-w

ABSTRACT: The goal is to create machine learning (ML) models that can use clinical and/or imaging data to predict longer length of stay (LOS) and intensive care unit (ICU) admission following torso (chest, abdomen, or pelvic) trauma. Materials and techniques: This study examined 840 adult patients who were hospitalised to a level 1 trauma centre during a one-year period following torso injuries. Age, sex, vital signs, clinical ratings, and test results were all considered clinical factors. Any injuries seen on CT were included in the imaging data. ICU admission and prolonged length of stay (defined as higher than the dataset's median length of stay) were the two outcomes of interest. We created and tested models for support vector machines (SVM) and artificial neural networks (ANN), and we measured the prediction performance using the area under the receiver operating characteristic (ROC) curve (AUC). Findings: SVM and ANN models have AUCs of up to 0.87 ± 0.03 and 0.78 ± 0.02 for predicting ICU admission, respectively. For predicting prolonged LOS, the AUCs of the SVM and ANN models were up to 0.80 ± 0.04 and 0.81 ± 0.05, respectively. Clinical parameter-only predictions were consistently less accurate than imaging-only or imaging-with-clinical-parameters-based predictions. Conclusions: The top-performing models beat those that just used clinical results by including imaging findings. By combining clinical and imaging data, ML models may be able to assist in trauma outcome prediction; however, more study may be required to maximise their effectiveness.

2.4 Early Prediction of Mortality, Severity, and Length of Stay in the Intensive Care Unit of Sepsis Patients Based on Sepsis 3.0 by Machine Learning Models:

https://www.frontiersin.org/journals/medicine/articles/10.3389/fmed.2021.664966/full

ABSTRACT: Background: It is crucial to predict the clinical fate of sepsis patients early on since this information can direct treatment and lower patient mortality. For physicians, nevertheless, it is therapeutically challenging. Techniques: Over the course of three years (2016–2018), 2,224 sepsis patients were treated in the intensive care unit (ICU) at Peking Union Medical College Hospital. Three machine learning models—logistic regression, random forest, and XGBoost—were utilised to forecast mortality, severity (sepsis/septic shock), and length of ICU stay (LOS) (>6 days, < 6 days) using all the important medical data from the first six hours in the intensive care unit. Before the dataset was included to the models, oversampling and missing data imputation were finished. Findings: Based on the area under the operational receiver characteristics (AUC) using the random forest classifier, the severity prediction outperformed the mortality and LOS predictions in terms of classification results (sensitivity = 0.65, specificity = 0.73, F1 score = 0.72, AUC = 0.79). Additionally, out of the three models, the random forest model performed the best overall (mortality prediction: sensitivity = 0.50, specificity = 0.84, F1 score = 0.66, AUC = 0.74; LOS prediction: sensitivity = 0.79, specificity = 0.66, F1 score = 0.69, AUC = 0.76). The three models mentioned above have better prediction power than the SOFA score alone. Conclusions: A thorough early warning of sepsis may be obtained within the first six hours of intensive care unit admission by using the random forest classifier. This will help with clinical decision-making, resource allocation, and management.

3.5 Predicting Intensive Care Unit Length of Stay and Mortality Using Patient Vital Signs: Machine Learning Model Development and Validation:

https://medinform.jmir.org/2021/5/e21347/

ABSTRACT: Background: At every step of care, patient monitoring is essential. Patient monitoring in the intensive care unit (ICU) in particular has the potential to lower complications and morbidity while also improving the quality of treatment by allowing hospitals to provide better, more affordable patient care and enhance the standard of medical services in the ICU. Our goal is to present the creation and verification of models for predicting mortality and duration of stay in intensive care units. When unfavourable medical circumstances are anticipated, the models will be utilised in an intelligent intensive care unit (ICU) patient monitoring module of an Intelligent Remote Patient Monitoring (IRPM) framework, which tracks patients' health status and promptly provides alarms, manoeuvre recommendations, or reports. Techniques: In order to create two prediction models—one for mortality prediction and another for ICU length of stay—we extracted ICU stay data for adult patients from the publicly accessible Medical Information Mart for Intensive Care (MIMIC) database. Six popular machine learning (ML) binary classification techniques were utilised for the mortality model in order to predict the discharge status (survived or not). Using the median ICU stay of 2.64 days for the patient group, we used the same six machine learning techniques for binary classification in the length of stay model. We employed two machine learning techniques to estimate the number of days for the regression-based categorisation. We constructed two versions of each prediction model: one based on our suggested quantiles approach, which uses 21 additional features derived from the baseline vital sign features, such as their modified means, standard deviations, and quantile percentages, and the other utilising 12 baseline demographic and vital sign features. Results: By employing the quantiles technique, we were able to execute predictive modelling with a small number of characteristics while still achieving respectable performance. Using the random forest approach, the mortality model's highest accuracy was almost 89%. Using the random forest approach, the length of stay model's maximum accuracy, based on the population median ICU stay (2.64 days), was almost 65%. Conclusions: The novel aspect of our strategy is that we developed models to reasonably predict ICU length of stay and mortality using a combination of machine learning (ML) and the quantiles approach, which relies solely on the patient's profile vital signs and does not require the use of any external features. In order to improve

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

the prediction potential of our models, this method is based on feature engineering of the vital signs by including their changed means, standard deviations, and quantile percentages of the original features. This gave us a richer dataset.

3. METHODOLOGY

a) Proposed Work:

The proposed system introduces a robust machine learning framework for predicting ICU length of stay (LOS) by leveraging comprehensive Electronic Health Records (EHR) data. Unlike traditional systems that rely mainly on vital signs, this approach incorporates a broader range of clinical information, including patient history and health conditions, to improve predictive accuracy. Various machine learning algorithms such as Logistic Regression, Random Forest, Multi-Layer Perceptron (MLP), Gradient Boosting, and XGBoost are applied to classify ICU stays as "short" or "long." The system is evaluated using performance metrics like accuracy, precision, recall, F1-score, and AUC. Furthermore, Explainable AI (XAI) tools such as SHAP are integrated to interpret the model and highlight the most influential features contributing to the prediction, enhancing transparency and decision support in clinical settings.

To further optimize the prediction results, the system is extended by incorporating the CatBoost algorithm, which is highly effective in handling categorical healthcare data and minimizing overfitting through ordered boosting techniques. CatBoost outperforms other models by achieving higher accuracy and offering better generalization across diverse patient records. A web-based interface is built using the Flask framework, allowing real-time prediction of ICU stays and providing healthcare professionals with an intuitive platform for decision-making. The seamless backend integration ensures that predictions are delivered quickly and accurately, thereby improving hospital resource allocation and supporting timely clinical interventions.

b) System Architecture:

The system architecture consists of several key components that work together to predict ICU length of stay. First, patient Electronic Health Records (EHR) are collected and preprocessed to handle missing values, encode categorical variables, and normalize numerical features. The cleaned data is then passed through multiple machine learning models including Logistic Regression, Random Forest, MLP, Gradient Boosting, XGBoost, and CatBoost for training and evaluation. Among these, CatBoost is used in the extended model due to its superior handling of categorical data. The trained model is integrated into a Flask-based web application, which serves as the user interface for clinicians to input patient data and receive real-time predictions. Additionally, Explainable AI techniques like SHAP are employed to visualize feature importance, ensuring transparency and interpretability in the decision-making process.

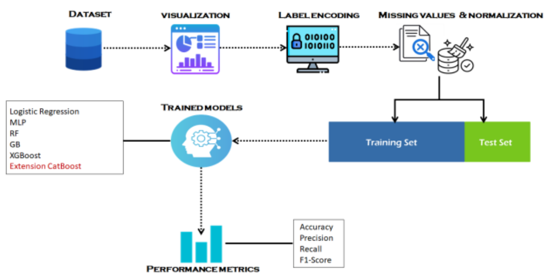

Fig: proposed architecture

c) Modules:

i. Dataset Loading

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

a. The dataset is loaded using Python libraries such as Pandas and NumPy.

b. It enables access to structured patient EHR data for further analysis and modeling.

ii. Visualization

a. Data visualization is performed using libraries like Matplotlib and Seaborn.

b. Helps identify trends, data distributions, and potential outliers related to ICU stay.

iii. Label Encoding

a. Converts categorical (non-numeric) data into numerical format using label encoding.

b. Ensures that machine learning algorithms can effectively process the input features.

iv. Handling Missing Values & Normalization

a. Missing values are imputed using suitable techniques to maintain data quality.

b. Normalization standardizes feature ranges, improving model accuracy and training efficiency.

v. Data Splitting

a. The dataset is split into 80% training and 20% testing subsets.

b. This supports accurate model training and reliable performance evaluation.

vi. Model Generation

a. Multiple ML models like Logistic Regression, Random Forest, MLP, XGBoost, and CatBoost are implemented.

b. Each model is evaluated using metrics such as accuracy, recall, F1-score, precision, and AUC.

vii. Admin Login

a. Admin is provided with a secure login interface to access the system.

b. Enables authorized access and control over the prediction process.

viii. Predict Hospital Stay

a. Users can input new patient data for ICU stay prediction via the web interface.

b. Provides real-time prediction results using the trained ML model.

ix. Logout

a. Allows users or admin to securely log out after usage.

b. Ensures user session management and system security.

e) Algorithms:

i. Logistic Regression: Based on patient characteristics, the likelihood of an ICU duration of stay is modelled using logistic regression. It is appropriate for binary classification problems, including forecasting brief or extended ICU stays, since it offers interpretable coefficients that show the significance of each variable.

ii. MLP (Multi-Layer Perceptron): Because of its layered construction, the Multi-Layer Perceptron is used to capture intricate correlations within the information. MLP can model nonlinear patterns by processing inputs across many layers, improving the prediction accuracy for estimating the length of an intensive care unit stay based on a variety of patient health markers.

iii. Random Forest (RF): Because Random Forest can capture feature significance and is resilient when dealing with overfitting, it is employed. This ensemble learning approach effectively predicts ICU length of stay based on a variety of patient variables by combining numerous decision trees to increase classification accuracy.

iv. Gradient Boosting (GB): By making incremental improvements, gradient boosting is used to increase prediction accuracy. This algorithm focusses on fixing mistakes from earlier iterations as it creates models one after the other. Accurately forecasting intensive care unit stays using patient EHR data is made possible by its effectiveness in managing intricate data linkages.

v. XGBoost: In classification jobs, XGBoost is used because of its excellent speed and scalability. In addition to providing outstanding prediction accuracy for ICU length of stay, this sophisticated boosting technique effectively manages computing resources during model training and performs very well when working with big datasets that contain a variety of characteristics.

vi. Extension CatBoost: An enhanced extension called CatBoost is used to enhance categorisation outcomes. By using gradient boosting and categorical characteristics, this technique improves prediction accuracy for intensive care unit duration of stay while streamlining the modelling process without requiring a lot of data preparation.

4. EXPERIMENTAL RESULTS

The experimental analysis was conducted on a hospital stay dataset sourced from the Kaggle repository, using six machine learning algorithms: Logistic Regression, Random Forest, MLP, Gradient Boosting, XGBoost, and an extended model with CatBoost. The dataset was preprocessed through label encoding, normalization, and missing value imputation before being split into training and testing sets. Each model was evaluated using key performance metrics such as accuracy, precision, recall, F1-score, and AUC. Among the traditional models, XGBoost delivered the highest accuracy, showcasing its effectiveness with complex data. However, the extended CatBoost model outperformed all others with an accuracy of 98.25%, demonstrating superior handling of categorical features and reduced overfitting. Additionally, Explainable AI techniques like SHAP were integrated to interpret model predictions, enhancing transparency and aiding clinicians in understanding feature importance in ICU stay prediction.

Accuracy: How well a test can differentiate between healthy and sick individuals is a good indicator of its reliability. Compare the number of true positives and negatives to get the reliability of the test. Following mathematical:

Accuracy = TP + TN /(TP + TN + FP + FN)

$$Accuracy = \frac{(TN + TP)}{T}$$

Precision: The accuracy rate of a classification or number of positive cases is known as precision. The formula is used to calculate precision:

Precision = TP/(TP + FP)

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

Recall: The ability of a model to identify all pertinent instances of a class is assessed by machine learning recall. The completeness of a

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

model in capturing instances of a class is demonstrated by comparing the total number of positive observations with the number of precisely predicted ones.

$$Recall = \frac{TP}{(FN + TP)}$$

F1-Score: A high F1 score indicates that a machine learning model is accurate. Improving model accuracy by integrating recall and precision. How often a model gets a dataset prediction right is measured by the accuracy statistic.

$$F1 - Score = 2 * \frac{(\text{Precision} \ * \ \text{Recall})}{((\text{Precision} \ + \ \text{Recall})\ )}$$

mAP: Assessing the level of quality Precision on Average (MAP). The position on the list and the number of pertinent recommendations are taken into account. The Mean Absolute Precision (MAP) at K is the sum of all users' or enquiries' Average Precision (AP) at K.

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k$$

$$AP_k = \text{the AP of class } k$$

$$n = \text{the number of classes}$$

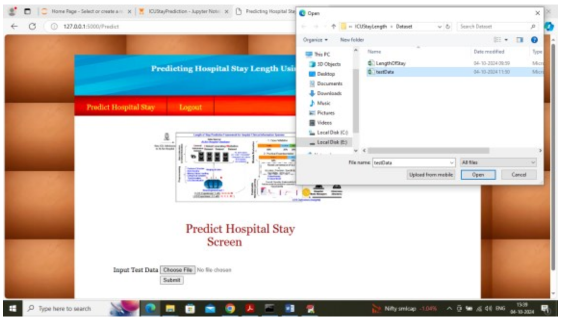

Fig.2 upload dataset

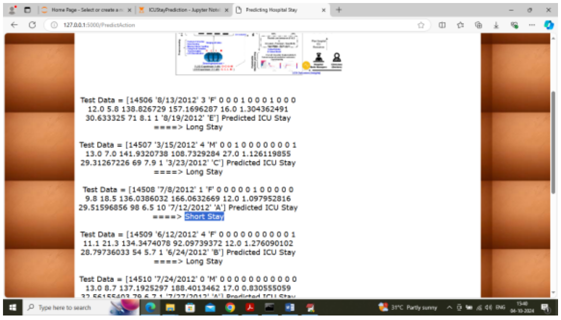

Fig.3. predicted results

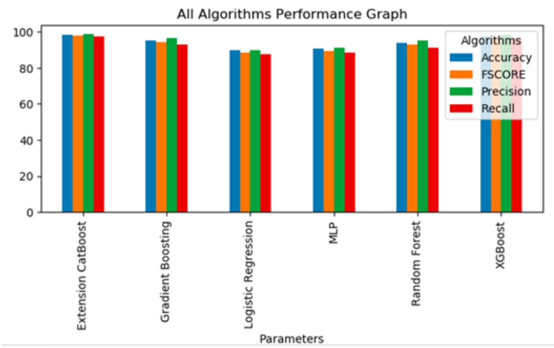

Fig: results graph

5. CONCLUSION

To sum up, the suggested solution effectively tackles the crucial problem of using patient Electronic Health Records (EHR) to forecast ICU length of stay (LOS). The study shows how applying different machine learning models may improve patient care and hospital resource management by making precise predictions about intensive care unit stays. With the greatest accuracy of 98.25% of the algorithms tested, the CatBoost model was the most effective approach. When compared to other conventional models, its capacity to efficiently handle categorical data and make use of gradient boosting greatly enhanced the prediction outcomes. By highlighting important characteristics that go into the prediction, Explainable AI (XAI) approaches like SHAP also contributed value by bringing transparency and understanding into the decision-making process. All things considered, the system emphasises how crucial it is to combine explainability with sophisticated machine learning algorithms to

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

maximise ICU resource allocation, which can eventually enhance patient outcomes and hospital efficiency.

6. FUTURE SCOPE

The research can be improved in the future by investigating deep learning methods for more intricate feature extraction and sequence modelling from EHR data, such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM). To improve prediction accuracy, hybrid models and ensemble learning techniques like stacking can also be used. To improve performance and lower computing complexity, feature engineering and dimensionality reduction methods like PCA might be investigated.

REFERENCES

[1] B. Alsinglawi, O. Alshari, M. Alorjani, O. Mubin, F. Alnajjar, M. Novoa, and O. Darwish, "An explainable machine learning framework for lung cancer hospital length of stay prediction," Sci. Rep., vol. 12, no. 1, pp. 1–10, Jan. 2022.

[2] B. Alsinglawi, "Predictive analytics framework for electronic health records with machine learning advancements: Optimizing hospital resources utilization with predictive and epidemiological models," Ph.D. thesis, Western Sydney Univ., Penrith, NSW, Australia, 2022. [Online]. Available: islandora/object/uws:67523/

[3] P. V. Staziaki, D. Wu, J. C. Rayan, I. D. D. O. Santo, F. Nan, A. Maybury, N. Gangasani, I. Benador, V. Saligrama, J. Scalera, and S. W. Anderson, "Machine learning combining CT findings and clinical parameters improves prediction of length of stay and ICU admission in torso trauma," Eur. Radiol., vol. 31, no. 7, pp. 5434–5441, Jul. 2021.

[4] L. Su, Z. Xu, F. Chang, Y. Ma, S. Liu, H. Jiang, H. Wang, D. Li, H. Chen, X. Zhou, N. Hong, W. Zhu, and Y. Long, "Early prediction of mortality, severity, and length of stay in the intensive care unit of sepsis patients based on sepsis 3.0 by machine learning models," Frontiers Med., vol. 8, p. 883, Jun. 2021

[5] K. Alghatani, N. Ammar, A. Rezgui, and A. Shaban-Nejad, "Predicting intensive care unit length of stay and mortality using patient vital signs: Machine learning model development and validation," JMIR Med. Informat., vol. 9, no. 5, May 2021, Art. no. e21347.

[6] M. C. Blom, K. Erwander, L. Gustafsson, M. Landin-Olsson, F. Jonsson, and K. Ivarsson, "The probability of readmission within 30 days of hospital discharge is positively associated with inpatient bed occupancy at discharge—A retrospective cohort study," BMC Emergency Med., vol. 15, no. 1, pp. 1–6, Dec. 2015.

[7] E. Rocheteau, P. Liò, and S. Hyland, "Temporal pointwise convolutional networks for length of stay prediction in the intensive care unit," 2020, arXiv:2007.09483.

[8] C. W. Hanson, C. S. Deutschman, H. L. Anderson, P. M. Reilly, E. C. Behringer, C. W. Schwab, and J. Price, "Effects of an organized critical care service on outcomes and resource utilization: A cohort study," Crit. Care Med., vol. 27, no. 2, pp. 270–274, Feb. 1999.

[9] S. Siddiqui, S. Ahmed, and R. Manasia, "Apache II score as a predictor of length of stay and outcome in our ICUs," J. Pakistan Med. Assoc., vol. 55, no. 6, p. 253, 2005.

[10] W. A. Knaus, J. E. Zimmerman, D. P. Wagner, E. A. Draper, and D. E. Lawrence, "APACHE—Acute physiology and chronic health eval uation: A

physiologically based classification system," Crit. Care Med., vol. 9, no. 8, pp. 591–597, Aug. 1981.

[11] A. Knaus, D. P. Wagner, E. A. Draper, J. E. Zimmerman, M. Bergner, P. G. Bastos, C. A. Sirio, D. J. Murphy, T. Lotring, A. Damiano, and F. E. Harrell Jr., "The APACHE III prognostic system: Risk prediction of hospital mortality for critically III hospitalized adults," Chest, vol. 100, no. 6, pp. 1619–1636, 1991.

[12] M. T. Keegan, O. Gajic, and B. Afessa, "Comparison of APACHE III, APACHE IV, SAPS 3, and MPM0III and influence of resuscitation status on model performance," Chest, vol. 142, no. 4, pp. 851–858, Oct. 2012.

[13] C.-C. Yeh, Y.-A. Chen, C.-C. Hsu, J.-H. Chen, W.-L. Chen, C.-C. Huang, and J.-Y. Chung, "Quick-SOFA score ≥ 2 predicts prolonged hospital stay in geriatric patients with influenza infection," Amer. J. Emergency Med., vol. 38, no. 4, pp. 780–784, Apr. 2020.

[14] C. Li, L. Chen, J. Feng, D. Wu, Z. Wang, J. Liu, and W. Xu, "Prediction of length of stay on the intensive care unit based on least absolute shrinkage and selection operator," IEEE Access, vol. 7, pp. 110710–110721, 2019.

[15] M. M. Islam, T. N. Poly, and Y.-C. Li, "Recent advancement of clinical information systems: Opportunities and challenges," Yearbook Med. Informat., vol. 27, no. 1, pp. 83–90, Aug. 2018.

[16] E. Levesque, E. Hoti, D. Azoulay, P. Ichai, D. Samuel, and F. Saliba, "The implementation of an intensive care information system allows shortening theICUlengthofstay,"J.Clin.Monitor.Comput.,vol.29,no.2, pp. 263–269, Apr. 2015.

[17] S. Calloway, H. A. Akilo, and K. Bierman, "Impact of a clinical decision support system on pharmacy clinical interventions, documentation efforts, and costs," Hospital Pharmacy, vol. 48, no. 9, pp. 744–752, Sep. 2013.

[18] S. T. McMullin, T. P. Lonergan, C. S. Rynearson, T. D. Doerr, P. A. Veregge, and E. S. Scanlan, "Impact of an evidence-based computerized decision support system on primary care prescription costs," Ann. Family Med., vol. 2, no. 5, pp. 494–498, Sep. 2004.

[19] R. T. Sutton, D. Pincock, D. C. Baumgart, D. C. Sadowski, R. N. Fedorak, and K. I. Kroeker, "An overview of clinical decision support systems: Benefits, risks, and strategies for success," NPJ Digit. Med., vol. 3, no. 1, pp. 1–10, Feb. 2020.

[20] D. Romanow, A. Rai, M. Keil, and S. Luxenberg, "Does extended CPOE use reduce patient length of stay?" Int. J. Med. Informat., vol. 97, pp. 128–138, Jan. 2017.

[21] G. S. Birkhead, M. Klompas, and N. R. Shah, "Uses of electronic health records for public health surveillance to advance public health," Annu. Rev. Public Health, vol. 36, no. 1, pp. 345–359, Mar. 2015.

[22] B. Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi, "Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis," IEEE J. Biomed. Health Informat., vol. 22, no. 5, pp. 1589–1604, Sep. 2018.

[23] C. Combi and G. Pozzi, "Clinical information systems and artificial intelligence: Recent research trends," Yearbook Med. Informat., vol. 28, no. 1, pp. 83–94, Aug. 2019.

[24] E. Berner and T. L. Lande, Clinical Decision Support Systems (Heath Informatics). Springer, 2007. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-31913-1_1

[25] X. Ma, Y. Si, Z. Wang, and Y. Wang, "Length of stay prediction for ICU patients using individualized single classification algorithm," Comput. Methods Programs Biomed., vol. 186, Apr. 2020, Art. no. 105224.

[26] L. Su, Z. Xu, F. Chang, Y. Ma, S. Liu, H. Jiang, H. Wang, D. Li, H. Chen, X. Zhou, N. Hong, W. Zhu, and Y. Long, "Early prediction of mortality, severity, and length of stay in the intensive care unit of sepsis patients based on sepsis 3.0 by machine learning models," Frontiers Med., vol. 8, p. 883, Jun. 2021.

[27] T.Gentimis, A.J.Alnaser, A.Durante, K.Cook, and R.Steele, "Predicting hospital length of stay using neural networks on MIMIC III data," in Proc. IEEE 15th Int. Conf Dependable, Autonomic Secure Comput., 15th Int. Conf Pervasive Intell. Comput., 3rd Int. Conf Big Data Intell. Com put. Cyber Sci. Technol. Congr. (DASC/PiCom/DataCom/CyberSciTech), Nov. 2017, pp. 1194–1201.

[28] K. Alghatani, N. Ammar, A. Rezgui, and A. Shaban-Nejad, "Predicting intensive care unit length of stay and mortality using patient vital signs: Machine learning model development and validation," JMIR Med. Informat., vol. 9, no. 5, May 2021, Art. no. e21347.

[29] A. E. W. Johnson, T. J. Pollard, L. Shen, L.-W.-H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark, "MIMIC-III, a freely accessible critical care database," Sci. Data, vol. 3, no. 1, pp. 1–9, May 2016.

[30]

[31] R. J. Steele and B. Thompson, "Data mining for generalizable pre admission prediction of elective length of stay," in Proc. IEEE 9th Annu. Comput. Commun. Workshop Conf. (CCWC), Jan. 2019, pp. 127–133.

[32] H. Quan, V. Sundararajan, P. Halfon, A. Fong, B. Burnand, J.-C. Luthi, L. D. Saunders, C. A. Beck, T. E. Feasby, and W. A. Ghali, "Coding algorithms for defining comorbidities in ICD-9-CM and ICD 10 administrative data," Med. Care, vol. 43, no. 11, pp. 1130–1139, Nov. 2005.

Authors profile:

Ms.M.Anitha Working as Assistant & Head of Department of MCA ,in SRK Institute of technology in Vijayawada. She done with B .tech, MCA ,M. Tech in Computer Science .She has 14 years of Teaching experience in SRK Institute of technology, Enikepadu, Vijayawada, NTR District. Her area of interest includes Machine Learning with Python and DBMS.

Ms.M.Priyanka Completed his Masters of Computer Applications from JNTUK. Currently working as an Assistant Professor in the department of MCA at SRK Institute of Technology, Enikepadu, NTR District. His area of interest include Artificial Intelligence and Machine Learning.

Technology, Enikepadu, Vijayawada, NTR District. She has Completed Degree in B.Sc(M.P.C.) from VIDHYALAYA DEGREE COLLEGE,Gudiwada. Her area of interest are Artificial intelligence and Machine Learning with Python.