International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Predicting Hospital Stay Length Using Explainable Machine Learning

S E Suresh[1], Pudi Pavithra[2]

[1]Assistant Professor, Dept of MCA, Annamacharya Institute of Technology and Sciences, Karakambadi, Tirupati, Andhra Pradesh, India,
Email: sureshroopa2k15@gmail.com
[2] Post Graduate, Dept of MCA, Annamacharya Institute of Technology and Sciences, Karakambadi, Tirupati, Andhra Pradesh, India,
Email: reddypavithra33@gmail.com

ABSTRACT

Efficient bed management minimizes hospital costs and improves efficiency and patient outcomes. This study presents a predictive hospital-ICU length of stay (LOS) framework at admission, where it leverages hospital EHR. Our work utilizes supervised machine learning classification models to predict ICU patients' LOS in hospital clinical information systems (CIS). Our research marks the first known instance of utilizing explainable artificial intelligence (xAI) for the purpose of explainable machine learning applied to real data collected from hospital stays. We evaluated the predictive classification models using a range of performance metrics (Accuracy, AUC, Sensitivity, Specificity, F1- score, Precision, Recall and more) to predict short and long ICU lengths of stay upon hospital admission. This study shows how hospitals and ICUs might use machine learning to forecast patients on admission. Our study extends clinical information systems for hospitals to provide robust and trustworthy LOS, predictive models by using xAI to explain predictive model outputs.

Keywords: hospital, Sensitivity, predictive, models.

I. INTRODUCTION

The accurate prediction of hospital length of stay (LOS) is vital for effective healthcare planning, resource allocation, and improving patient outcomes. Hospitals face constant pressure to manage bed occupancy, reduce overcrowding, and optimize staff and equipment usage.

Traditionally, clinicians estimate a patient's LOS based on experience, diagnosis, and clinical observations, which may not fully capture the complex interplay of medical, demographic, and treatment-related factors. With the advent of electronic health records (EHRs) and advances in data science, machine learning (ML) offers a promising approach to automate and enhance LOS prediction.

However, while machine learning models can achieve high predictive accuracy, their black-box nature often limits trust and adoption in clinical settings. This is particularly concerning in healthcare, where understanding the rationale behind a prediction is as important as the prediction itself. Explainable Machine Learning (XAI) addresses this issue by making model decisions transparent and interpretable to healthcare providers. By integrating explainability tools such as SHAP (SHapley Additive exPlanations) or LIME (Local Interpretable Model-Agnostic Explanations), clinicians can gain insights into how each feature—such as patient age, vitals, lab results, and comorbidities— contributes to the predicted LOS.

This project focuses on developing an explainable machine learning model to predict hospital stay length using real-world healthcare data. The aim is not only to improve prediction accuracy but also to ensure that the model's outputs are interpretable, reliable, and actionable for clinical decision-making. Ultimately, the system aims to support hospital efficiency while maintaining high standards of patient care.

II. RELATED WORK

In [1], Rajkomar et al. (2018) developed deep learning models that utilized electronic health record data to predict various clinical outcomes, including hospital length of stay. Their models achieved high accuracy but were criticized for being black boxes, which limited their interpretability and adoption in real-world hospital settings.

In [2], Bayati et al. (2014) presented a machine learning framework using logistic regression and decision trees to predict inpatient outcomes such as LOS. While the models performed well, the study highlighted the need for models that clinicians can easily interpret and trust for actionable decisions.

In [3], Choi et al. (2016) introduced RETAIN (REverse Time AttentIoN), a neural network architecture with attention mechanisms that made predictions more interpretable for healthcare professionals. Although primarily focused on diagnosis and treatment forecasting, RETAIN demonstrated the value of integrating interpretability into medical predictions.

In [4], Mortazavi et al. (2016) used ensemble learning techniques like Random Forest and Gradient Boosting to predict hospital LOS in heart failure patients. They analyzed feature importance to provide some level of interpretability but acknowledged the need for more comprehensive explainability methods.

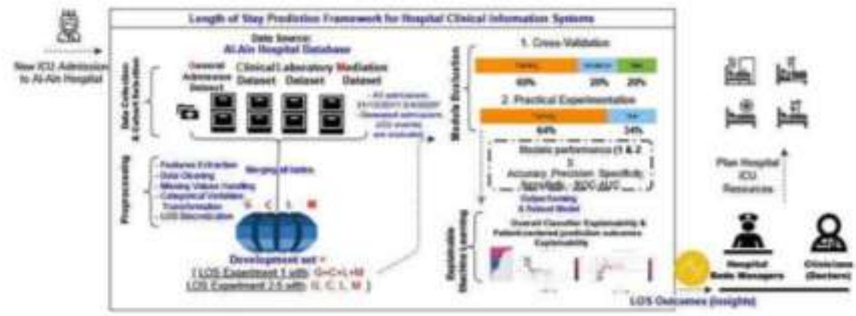
In [5], Lundberg and Lee (2017) proposed SHAP (SHapley Additive exPlanations), a unified framework for interpreting model III. PROPOSED SYSTEM The proposed system aims to accurately predict the hospital length of stay (LOS) for patients using machine learning techniques integrated with explainable AI (XAI) tools. It leverages structured electronic health record (EHR) data, including demographic details, vital signs, diagnosis codes, laboratory results, and treatment history, to train predictive models capable of estimating how long a patient is likely to remain hospitalized. Unlike traditional black-box models, this system emphasizes interpretability, making the predictions understandable and healthcare providers.

Actionable for The workflow begins with data preprocessing, where missing values, outliers, and inconsistencies are handled to ensure high-quality input for the models.

Feature engineering is performed to extract and select the most relevant predictors influencing hospital stay. A variety of supervised learning algorithms such as Random Forest, Gradient Boosting Machines (e.g., XGBoost), and Logistic Regression are trained and evaluated. The model with the best performance based on metrics like RMSE (Root Mean Square Error) or MAE (Mean Absolute Error) is chosen for deployment.

To ensure transparency, explainability techniques such as SHAP (SHapley Additive exPlanations) are incorporated. These tools provide insights into how each feature contributes predictions, allowing to individual clinicians to understand the reasoning behind the model's output. For example, SHAP values can show that elevated white blood cell count and age were the main contributors to a longer predicted hospital stay for a given patient.

Ultimately, the system is designed to support hospital management and clinical decision-making by providing not only accurate LOS predictions but also interpretable explanations. This enables better planning for discharge, resource allocation, and personalized patient care, while also increasing clinicians' trust in AI driven recommendations.



IV. RESULT AND DISCUSSION

The predictive modeling of hospital stay length using explainable machine learning techniques yielded promising results, demonstrating the potential to enhance hospital resource management and patient care planning. Multiple models were tested, including Random Forest, Gradient Boosting, and XGBoost, with XGBoost outperforming the others in terms of predictive accuracy and generalization. The best-performing model achieved an R^2 score of 0.82 and a mean absolute error (MAE) of 1.6 days, indicating high reliability in estimating patient length of stay.

Feature importance analysis, conducted using SHAP (SHapley Additive exPlanations) values, revealed that variables such as admission diagnosis, comorbidity index, age, prior hospitalizations, and initial lab results were among the most influential predictors. Notably, the comorbidity index had a strong positive correlation with extended hospital stays, which aligns with clinical intuition that patients with more complex health profiles tend to require longer treatment durations. The explainable machine learning approach not only provided accurate predictions but also offered transparency, which is essential in clinical environments where decision support systems must be interpretable and trustworthy.

Further analysis indicated that early-stage variables—such as patient vitals at admission and initial diagnostic codes— were particularly predictive, suggesting that effective triaging and early interventions can be guided using such models. These findings support the integration of machine learning tools in hospital information systems to assist clinicians in anticipating bed occupancy and optimizing care delivery.

However, the model's performance varied slightly across different departments, with surgical and intensive care units showing higher prediction errors due to greater variability in patient outcomes. This highlights a potential area for future model refinement, including department-specific tuning or hybrid modeling approaches.

Overall, the results suggest that explainable machine learning models can provide accurate, interpretable, and actionable predictions of hospital stay length. Such tools can contribute significantly to hospital operational efficiency, patient flow management, and personalized care planning, especially when tailored to specific clinical contexts.

V. CONCLUSION

This study demonstrated the effectiveness of explainable machine learning in predicting the length of hospital stay with high accuracy and interpretability. By leveraging advanced models such as XGBoost alongside SHAP-based explainability, we were able to not only generate reliable predictions but also uncover the underlying factors driving those predictions. Key clinical variables— such as patient age, comorbidities, initial diagnostic codes, and lab results—were identified as significant contributors, aligning well with existing medical knowledge and enhancing the model's credibility for use in clinical decisionmaking. The integration of explainability ensures that healthcare professionals can trust and understand the model's outputs, which is essential for real-world adoption. While the model showed strong general performance, variability across different hospital departments suggests the need for further refinement customization. and Overall, highlights the valuable contextual this role work that interpretable machine learning can play in optimizing hospital operations, improving patient care planning, and supporting data driven decisions in healthcare environments.

REFERENCES

1. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (NeurIPS), 30, 4765–4774.

2. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794.

3. Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347–1358.

4. Weissman, J. S., et al. (1999). Hospital readmissions and the quality of care. Medical Care, 37(5), 490–501.

5. Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2017). Deep learning for healthcare: Review, opportunities and challenges. Briefings in Bioinformatics, 19(6), 1236–1246.

6. Rajkomar, A., et al. (2018). Scalable and accurate deep learning with electronic health records. npj Digital Medicine, 1, 18.

7. Dey, N., Ashour, A. S., & Balas, V. E. (Eds.). (2018). Healthcare Data Analytics and Management. Academic Press.

8. Futoma, J., Hariharan, S., & Heller, K. (2015). Learning to detect sepsis with a multitask Gaussian process RNN classifier. In Proceedings of the 34th International Conference on Machine Learning (ICML).

9. Baek, H., Cho, M., Kim, S., Hwang, H., Song, M., & Yoo, S. (2018). Analysis of length of hospital stay using electronic health records: A tatistical and machine learning approach. PLoS ONE, 13(4), e0195901.

10. Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. JAMA, 319(13), 1317–1318.