2026 Volume: 02 Issue: 06 www.ijrpa.com ISSN 2456-9995 Research Article

International Journal Research Publication Analysis

Page: 01-17

AI DRIVEN HOSPITAL READMISSION RISK SCORING USING TEMPORAL MODELS

*[1]Shaheer Mustafa, [1]Shubh Agarwal, [1]Ritik Vats, [1]Rishikesh K. Singh, [2]Shivani Chauhan

[1]Department of Computer Science and Engineering Quantum University, Roorkee.

[2]AssistantProfessor, Department of Computer Science and Engineering, Quantum University, Roorkee.

Article Received: 1 May 2026

Article Revised: 21 May 2026

Published on: 11 June 2026

***Corresponding Author: Shaheer Mustafa**

Department of Computer Science and Engineering Quantum University, Roorkee.

DOI: https://doi-doi.org/101555/ijrpa.3991

ABSTRACT

Hospital readmission remains a major challenge in modern healthcare systems due to its impact on patient outcomes, healthcare costs, and hospital resource utilization. Early identification of patients at risk of readmission enables healthcare providers to implement preventive interventions and improve discharge planning. This research presents an AI Driven Hospital Readmission Risk Scoring System that utilizes temporal healthcare indicators and machine learning techniques to predict patient readmission risk.

The proposed framework employs historical healthcare data including hospitalization duration, laboratory procedures, medication usage, outpatient visits, emergency visits, and previous inpatient admissions. Data preprocessing and feature engineering techniques were applied to prepare the dataset for predictive modeling. A Random Forest Classifier was utilized as the primary prediction algorithm because of its robustness, ability to handle structured healthcare data, and strong classification performance.

The trained model was deployed through a Streamlit-based web application that allows healthcare professionals to perform real-time readmission risk assessment. Experimental evaluation produced an accuracy of 86.57%, demonstrating the effectiveness of the proposed framework in identifying readmission patterns from historical healthcare records. The system provides an interactive and user-friendly decision-support environment for healthcare analytics and patient risk management.

KEYWORDS: Artificial Intelligence, Machine Learning, Hospital Readmission Prediction,

Random Forest, Healthcare Analytics, Predictive Modeling, Streamlit, Risk Scoring System.

I. INTRODUCTION

Hospital readmission has become one of the most significant challenges faced by modern healthcare systems worldwide. A hospital readmission occurs when a patient is admitted to a healthcare facility again within a specific period after discharge, often indicating incomplete recovery, inadequate follow-up care, or the progression of underlying medical conditions. Frequent readmissions not only increase healthcare expenditure but also place additional pressure on hospital resources and negatively impact patient outcomes. As healthcare organizations strive to improve the quality of care while reducing operational costs, the ability to predict potential readmissions has become an important area of research.

The rapid advancement of Artificial Intelligence (AI) and Machine Learning (ML) has transformed the healthcare sector by enabling data-driven decision-making and predictive analytics. Healthcare institutions generate large volumes of patient-related data through electronic health records, laboratory reports, medication histories, and clinical assessments. These datasets contain valuable patterns that can be leveraged to identify patients who are at a higher risk of readmission. Predictive modeling techniques allow healthcare providers to analyze historical patient information and develop proactive intervention strategies before complications arise.

Traditional approaches to identifying high-risk patients often rely on manual assessments and clinical judgment, which may be subjective and time-consuming. Machine learning algorithms provide a more systematic and scalable solution by learning complex relationships between healthcare variables and patient outcomes. By utilizing historical hospitalization data, machine learning models can detect hidden trends and risk factors that may not be immediately apparent through conventional analysis.

In this research, an AI Driven Hospital Readmission Risk Scoring System is proposed using temporal healthcare indicators and machine learning techniques. The system focuses on predicting whether a patient is likely to be readmitted based on factors such as length of hospital stay, number of laboratory procedures, medication usage, outpatient visits, emergency visits, and inpatient admissions. These variables capture important aspects of patient healthcare utilization and provide meaningful insights into future readmission risks.

A Random Forest Classifier is employed as the primary predictive model due to its ability to handle structured healthcare data efficiently, reduce overfitting, and provide robust classification performance. The model is trained on a publicly available healthcare dataset and integrated into a Streamlit-based web application that enables real-time risk assessment. Healthcare professionals can utilize the system to evaluate patient readmission probabilities and support discharge planning decisions.

The major objectives of this research are:

1. To develop an intelligent healthcare analytics system capable of predicting hospital readmission risk.
2. To identify significant temporal healthcare indicators associated with patient readmission.
3. To apply machine learning techniques for accurate classification of readmission outcomes.
4. To deploy the predictive model through a user-friendly web-based interface for real-time decision support.
5. To evaluate the effectiveness of the proposed system using standard machine learning performance metrics.

The proposed framework contributes to healthcare analytics by combining predictive modeling with practical deployment capabilities. Early identification of high-risk patients can assist healthcare providers in implementing preventive measures, improving patient monitoring, optimizing resource allocation, and ultimately reducing avoidable hospital readmissions. The integration of artificial intelligence into healthcare decision-support systems demonstrates the growing potential of data-driven technologies in enhancing clinical outcomes and operational efficiency.

The remainder of this paper is organized as follows. Section II presents the literature review of existing hospital readmission prediction techniques. Section III describes the proposed methodology and system architecture. Section IV discusses the experimental setup and implementation details. Section V presents the results and performance evaluation. Section VI outlines future research directions, and Section VII concludes the study.

II. LiteratureSurvey

Hospital readmission prediction has attracted considerable attention in recent years due to its potential to improve patient care and reduce healthcare costs. Researchers have explored various machine learning and artificial intelligence techniques to identify patients at risk of readmission using electronic health records and clinical datasets.

In a study conducted by K. K. Khaing et al., machine learning models were applied to healthcare datasets to predict hospital readmission outcomes. The researchers compared multiple classification algorithms and found that ensemble-based methods achieved better predictive performance than traditional statistical approaches. Their findings highlighted the importance of feature selection and data preprocessing in healthcare analytics.

A. Rajkomar et al. proposed a deep learning framework for healthcare prediction tasks using electronic health records. The study demonstrated that deep neural networks could effectively learn complex relationships among clinical variables and improve prediction accuracy for various medical outcomes, including patient readmission. The authors emphasized the growing role of artificial intelligence in clinical decision support systems.

Research by Futoma et al. investigated early hospital readmission prediction using machine learning models trained on patient history and hospitalization records. Their work revealed that healthcare utilization indicators such as previous admissions and emergency visits are strong predictors of future readmission risk. The study also emphasized the importance of temporal healthcare information in predictive modeling.

Shickel et al. explored deep learning applications in healthcare analytics and reviewed various predictive models used for disease diagnosis, patient monitoring, and readmission prediction. Their findings suggested that machine learning-based systems can significantly enhance healthcare decision-making when integrated with electronic medical records.

Another notable study by Huang et al. utilized Random Forest and Gradient Boosting algorithms to predict hospital readmissions. The authors reported that ensemble learning techniques provided stable and reliable results due to their ability to handle high-dimensional healthcare data and complex feature interactions. Their work demonstrated the suitability of tree-based methods for structured clinical datasets.

Bertsimas et al. developed a predictive healthcare analytics framework using machine learning models to identify patients at risk of adverse clinical outcomes. The study concluded that predictive systems could help healthcare providers allocate resources more effectively and implement targeted intervention strategies for high-risk patients.

Recent research by Chen et al. employed Extreme Gradient Boosting (XGBoost) for readmission prediction and achieved promising results in terms of accuracy and recall. The study demonstrated the effectiveness of boosting algorithms in healthcare classification problems and encouraged further exploration of advanced ensemble learning techniques.

Although significant progress has been made in hospital readmission prediction, many existing approaches focus primarily on complex models that may require extensive computational resources and specialized deployment environments. Furthermore, several studies emphasize prediction accuracy without providing practical implementation frameworks that healthcare professionals can easily access and utilize.

To address these limitations, the present research proposes an AI Driven Hospital Readmission Risk Scoring System using a Random Forest Classifier and temporal healthcare indicators. The proposed framework not only performs readmission risk prediction but also integrates the trained model into a Streamlit-based web application, enabling real-time clinical decision support. This combination of predictive analytics and practical deployment contributes toward the development of accessible and efficient healthcare intelligence systems.

III. METHODOLOGY

The proposed AI Driven Hospital Readmission Risk Scoring System utilizes machine learning techniques to predict whether a patient is likely to be readmitted after discharge. The methodology consists of data collection, preprocessing, feature selection, model training, prediction generation, and deployment through a web-based interface. The overall framework is designed to support healthcare professionals in identifying high-risk patients and facilitating timely intervention.

A. Dataset Description

The study utilizes the Diabetes 130-US Hospitals dataset obtained from the UCI Machine Learning Repository. The dataset contains patient encounter records collected from multiple hospitals across the United States. It includes demographic information, hospitalization records, laboratory procedures, medication usage, and admission history.

For this research, the target variable is patient readmission status, which indicates whether a patient was readmitted after discharge. To improve model efficiency and interpretability, a subset of highly relevant healthcare utilization features was selected.

The selected input features are:

- Time in Hospital
- Number of Laboratory Procedures
- Number of Medications
- Number of Outpatient Visits
- Number of Emergency Visits
- Number of Inpatient Admissions

These features represent temporal healthcare indicators that capture patient interaction with healthcare services and are considered important factors influencing readmission risk.

B. Data Preprocessing

Raw healthcare datasets often contain inconsistencies, missing values, and categorical variables that must be transformed before model training. Therefore, several preprocessing steps were performed.

1. Removal of irrelevant attributes.
2. Handling of missing and inconsistent values.
3. Conversion of categorical readmission labels into binary classes.
4. Feature extraction and selection.
5. Data normalization and formatting.
6. Splitting of data into training and testing datasets.

The preprocessing stage ensured data quality and improved the reliability of machine learning predictions.

C. Feature Engineering

Feature engineering plays a critical role in predictive healthcare analytics. The selected features were analyzed using exploratory data analysis techniques to identify their relationship with readmission outcomes.

Temporal healthcare indicators were prioritized because they provide insights into patient healthcare utilization patterns. Patients with frequent inpatient admissions, emergency visits, or extended hospital stays often demonstrate a higher probability of future readmission.

The engineered dataset was structured into feature vectors suitable for machine learning model training and evaluation.

D. Random Forest Classification Model

A Random Forest Classifier was selected as the primary prediction algorithm due to its robustness, scalability, and effectiveness in handling structured healthcare datasets.

Random Forest is an ensemble learning technique that constructs multiple decision trees during training and combines their predictions through majority voting. This approach reduces overfitting and improves model generalization.

The algorithm operates through the following steps:

1. Generate multiple bootstrap samples from the training dataset.
2. Construct individual decision trees using random subsets of features.

3. Train each tree independently.
4. Aggregate predictions from all trees.
5. Produce the final classification based on majority voting.

The advantages of Random Forest include:

- High predictive accuracy.
- Resistance to overfitting.
- Ability to handle large datasets.
- Efficient feature importance estimation.
- Robust performance on healthcare data.

E. System Architecture

The proposed system consists of four major components:

1) Data Layer

Responsible for storing and managing healthcare records and processed datasets.

2) Machine Learning Layer

Performs data preprocessing, feature extraction, model training, and prediction generation using the Random Forest algorithm.

3) Application Layer

Implements the prediction engine and business logic for readmission risk assessment.

4) User Interface Layer

Provides an interactive environment for healthcare professionals through a web-based dashboard.

The architecture enables seamless communication between data processing modules and prediction services.

IV. SystemArchitecture

The proposed AI Driven Hospital Readmission Risk Scoring System is designed using a modular and layered architecture that enables efficient healthcare data processing, accurate prediction generation, and real-time deployment. The architecture integrates healthcare analytics, machine learning techniques, and a web-based application to assist healthcare professionals in identifying patients who are at risk of hospital readmission. The layered design improves scalability, maintainability, reliability, and flexibility for future healthcare applications.

The first layer of the architecture is the **Healthcare Data Ingestion Layer**. This component is responsible for collecting and managing patient healthcare information obtained from electronic health records and healthcare datasets. The system accepts patient-related attributes including hospitalization duration, laboratory procedures, medication count, outpatient visits, emergency visits, and inpatient admissions. Since healthcare data may originate from multiple sources and formats, this layer ensures proper organization and structured storage before further processing.

After data acquisition, the information is passed to the **Data Preprocessing Layer**, which plays a critical role in improving data quality and consistency. Healthcare datasets often contain missing values, duplicate records, noisy information, and inconsistencies that may negatively affect prediction performance. Therefore, preprocessing operations are applied to clean the data, remove irrelevant information, and transform records into a machine-readable format suitable for machine learning algorithms.

The next component is the **Feature Engineering and Transformation Layer**. This layer identifies and prepares the most relevant healthcare indicators contributing to readmission risk prediction. The selected features include time in hospital, number of laboratory procedures, number of medications, outpatient visits, emergency visits, and inpatient admissions. These variables are transformed into feature vectors that serve as inputs to the predictive model. Feature engineering improves prediction accuracy by reducing unnecessary complexity and emphasizing the most informative patient attributes.

The core component of the proposed architecture is the **Machine Learning Inference Layer**, which contains the Random Forest classification engine. The Random Forest model processes the transformed patient feature vectors and predicts the likelihood of hospital readmission. Random Forest is selected because of its ability to handle structured healthcare datasets efficiently, reduce overfitting through ensemble learning, and provide stable classification performance. The model generates prediction outcomes that classify patients according to their readmission risk level. These predictions help healthcare professionals identify vulnerable patients who may require additional monitoring or preventive care after discharge.

To ensure practical usability, the trained Random Forest model is integrated into the **Deployment Layer** using the Streamlit framework. Streamlit provides an interactive graphical user interface that allows healthcare professionals to interact with the prediction system in real time. Through the web interface, clinicians can enter patient healthcare indicators and instantly obtain readmission risk predictions. The interactive nature of the application improves accessibility and allows non-technical healthcare staff to utilize the system effectively.

The final component is the **User Interface Layer**, which serves as the interaction point between healthcare professionals and the prediction framework. This layer presents prediction outcomes in a simple and understandable format using risk classifications and prediction results. Doctors and hospital administrators can use these outputs to support discharge planning, patient monitoring, resource allocation, and follow-up treatment strategies.

One of the major strengths of the proposed architecture is its modular design. Each layer operates independently while maintaining seamless communication with other system components. This modularity improves maintainability because individual modules can be upgraded or replaced without affecting the entire framework. For example, the Random Forest model can be replaced with advanced machine learning or deep learning algorithms in future research while retaining the same deployment infrastructure.

The architecture also supports future scalability and expansion. As healthcare datasets continue to grow, the framework can be integrated with cloud computing platforms, distributed machine learning systems, and advanced healthcare analytics solutions. Additional functionalities such as Explainable Artificial Intelligence (XAI), real-time patient monitoring, and temporal deep learning models can also be incorporated into the existing architecture. This flexibility makes the proposed system suitable for modern intelligent healthcare environments and future clinical decision-support applications.

V. Implementation

The implementation of the proposed AI Driven Hospital Readmission Risk Scoring System is carried out using the Python programming language due to its simplicity, extensive machine learning ecosystem, and strong support for healthcare analytics applications. Python provides a wide range of libraries and frameworks that simplify data preprocessing, model development, evaluation, and deployment. The complete implementation process includes dataset handling, preprocessing, feature engineering, model training, model serialization, backend integration, and deployment through a web-based application.

The first phase of implementation focuses on healthcare dataset processing and manipulation. The Diabetes 130-US Hospitals dataset is imported and analyzed using the Pandas and NumPy libraries. Pandas provides efficient data structures such as DataFrames that simplify healthcare data management, cleaning, filtering, and transformation. NumPy supports high-performance numerical computations required during preprocessing and machine learning tasks.

During preprocessing, multiple operations are performed to improve dataset quality and prepare the data for machine learning algorithms. Missing values are identified and handled appropriately, duplicate records are removed, and irrelevant attributes are excluded from the analysis. The target variable, readmitted, is converted into a binary classification format suitable for predictive modeling.

After preprocessing, feature engineering techniques are applied to select the most relevant healthcare indicators associated with hospital readmission. The final feature set used in the implementation includes:

- Time in Hospital
- Number of Laboratory Procedures
- Number of Medications
- Number of Outpatient Visits
- Number of Emergency Visits
- Number of Inpatient Admissions

These selected features provide important information regarding patient healthcare utilization patterns and historical clinical interactions.

The processed dataset is divided into training and testing datasets using the Scikit-learn library. Train-test splitting ensures that the machine learning model is evaluated on unseen patient records and can generalize effectively. Approximately 80% of the dataset is utilized for training purposes, while the remaining 20% is reserved for testing and validation. Scikit-learn also provides evaluation utilities and performance metrics such as accuracy, precision, recall, F1-score, and confusion matrix analysis.

The primary predictive model implemented in this framework is the Random Forest Classifier. Random Forest is selected because of its high predictive accuracy, resistance to overfitting, and ability to handle structured healthcare datasets efficiently. The algorithm constructs multiple decision trees during training and combines their outputs through majority voting to generate final predictions. This ensemble learning approach improves classification reliability and robustness.

During model training, Random Forest parameters such as the number of estimators, maximum tree depth, minimum samples split, and minimum samples leaf can be adjusted to optimize prediction performance. The model learns patterns from historical healthcare records and establishes relationships between selected healthcare indicators and readmission outcomes.

After successful training, the model is evaluated using standard machine learning performance metrics. The experimental results obtained from the testing dataset are as follows:

Performance Metric	Obtained Value
Accuracy	86.57%
Precision	15.51%
Recall	4.42%
F1-Score	6.88%

The confusion matrix generated during evaluation is shown below:

Actual / Predicted	Negative	Positive
Negative	17,519	550
Positive	2,184	101

These results indicate that the model achieves strong overall classification accuracy while providing predictive insights regarding patient readmission behavior.

Once the training process is completed successfully, the trained Random Forest model is serialized using the Joblib library. Model serialization allows the trained machine learning model to be saved and reused without requiring retraining. This improves efficiency and reduces computational overhead during real-time prediction tasks.

For deployment purposes, the serialized model is integrated into the Streamlit framework. Streamlit enables rapid development of interactive machine learning web applications with minimal coding complexity. The frontend interface is designed to be simple, user-friendly, and accessible for healthcare professionals who may not possess extensive technical expertise.

The deployed application contains input fields corresponding to the selected healthcare indicators. Clinicians can enter patient information such as hospitalization duration, medication count, laboratory procedures, outpatient visits, emergency visits, and inpatient admissions. The backend processing pipeline receives the inputs, performs necessary transformations, and forwards the feature vector to the Random Forest prediction engine.

The prediction engine analyzes the supplied patient information and generates readmission risk predictions in real time. The application displays prediction outcomes that help healthcare professionals identify patients who may require additional monitoring, follow-up treatment, or preventive intervention after discharge.

The implementation successfully transforms the machine learning model into a practical healthcare decision-support system capable of assisting hospitals in proactive patient management and readmission risk assessment.

VI. RESULTS AND DISCUSSION

The proposed AI-driven hospital readmission risk scoring system was evaluated using the UCI Diabetes 130-US Hospitals dataset. A Random Forest Classifier was trained on selected temporal healthcare indicators, including time in hospital, number of laboratory procedures, number of medications, outpatient visits, emergency visits, and inpatient visits. The performance of the model was assessed using standard classification metrics such as Accuracy, Precision, Recall, and F1-Score.

A. Model Performance Evaluation

The experimental results obtained from the Random Forest model are presented in Table 1.

Table 1: Performance Metrics of the Proposed Model.

Metric	Value (%)
Accuracy	86.57
Precision	15.51
Recall	4.42
F1-Score	6.88

The model achieved an overall accuracy of 86.57%, indicating that it correctly classified a large proportion of patient records. However, the relatively low recall value suggests that a significant number of actual readmission cases were not identified by the model. This outcome can be attributed to the class imbalance present in the healthcare dataset, where non-readmitted patients considerably outnumber readmitted patients.

Although the precision value of 15.51% indicates that only a limited proportion of predicted positive cases were actual readmissions, the model still demonstrates its ability to identify high-risk patients based on temporal healthcare indicators.

B. Confusion Matrix Analysis

The confusion matrix generated from the test dataset is shown in Table 2.

Table 2: Confusion Matrix of the Random Forest Model.

Actual Class	Predicted Negative	Predicted Positive
Negative	17,519	550
Positive	2,184	101

The confusion matrix reveals that the model successfully classified 17,519 non-readmitted patients and correctly identified 101 readmitted patients. However, 2,184 readmitted cases were incorrectly classified as non-readmitted, which contributed to the lower recall score.

The results indicate that while the model performs effectively in recognizing patients with a low risk of readmission, further optimization techniques such as class balancing, feature expansion, and advanced ensemble methods may improve the detection of actual readmission cases.

C. Impact of Temporal Healthcare Indicators

The selected temporal healthcare indicators played an important role in the prediction process. Variables such as hospitalization duration, frequency of inpatient visits, emergency admissions, and medication count provided valuable information regarding patient health conditions and healthcare utilization patterns.

The Random Forest algorithm effectively captured the relationships among these indicators and generated risk predictions based on historical patient behavior. This demonstrates the importance of temporal healthcare data in identifying potential readmission risks.

D. Deployment and User Interface Evaluation

To improve practical usability, the trained model was deployed using the Streamlit framework. The web-based interface allows healthcare professionals to input patient-related information and instantly obtain readmission risk predictions.

The deployment offers several advantages:

- Real-time prediction capability.
- User-friendly graphical interface.
- Fast and efficient model inference.
- Easy integration into healthcare decision-support environments.
- Accessibility without requiring technical expertise.

The deployment phase demonstrates how machine learning models can be translated from research environments into practical healthcare applications that support clinical decision-making.

E. DISCUSSION

The experimental findings confirm that machine learning techniques can be effectively utilized for hospital readmission risk prediction. The Random Forest model achieved strong overall classification accuracy and successfully identified patterns within patient healthcare utilization records.

Despite the high accuracy, the relatively low recall and F1-score indicate opportunities for improvement in detecting readmitted patients. Future enhancements may include advanced feature engineering, data balancing approaches, hyperparameter optimization, and the incorporation of additional clinical variables.

Overall, the proposed system establishes a reliable foundation for AI-driven healthcare analytics and demonstrates the potential of predictive modeling for supporting hospital resource planning and patient care management.

VII. Future Scope

The proposed AI-driven hospital readmission risk scoring system demonstrates the potential of machine learning in supporting healthcare decision-making. However, several enhancements can be incorporated in future versions to improve prediction performance, scalability, and real-world applicability.

A. Integration of Additional Clinical Features

The current model utilizes selected temporal healthcare indicators for prediction. Future work can incorporate additional clinical attributes such as laboratory test results, diagnosis codes, medication history, demographic information, and patient comorbidities. The inclusion of richer healthcare data may improve the model's ability to identify high-risk patients more accurately.

B. Advanced Machine Learning and Deep Learning Models

Although the Random Forest algorithm provides reliable performance and interpretability, future research may explore advanced machine learning and deep learning techniques such as Gradient Boosting, XGBoost, LightGBM, Artificial Neural Networks (ANNs), and Long Short-Term Memory (LSTM) networks. These approaches may capture more complex relationships within healthcare data and improve readmission detection rates.

C. Handling Class Imbalance

The experimental results indicate that the dataset contains an imbalance between readmitted and non-readmitted patients. Future studies can apply techniques such as SMOTE (Synthetic Minority Over-sampling Technique), undersampling, oversampling, and cost-sensitive learning to improve recall and overall predictive effectiveness for minority-class patients.

D. Explainable Artificial Intelligence (XAI)

Healthcare applications require transparency and trust in machine learning predictions. Future implementations can integrate Explainable AI techniques such as SHAP (SHapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations) to provide explanations for individual predictions and assist healthcare professionals in understanding model decisions.

E. Real-Time Healthcare Monitoring

The current system operates using historical patient information. Future versions can be connected with real-time healthcare monitoring systems and Electronic Health Records (EHRs) to continuously assess patient risk and generate timely alerts for healthcare providers.

F. Cloud-Based Deployment and Scalability

The Streamlit-based application can be extended into a cloud-hosted healthcare analytics platform. Cloud deployment would support large-scale data processing, remote accessibility, multi-user environments, and seamless integration with hospital information systems.

G. Decision Support Integration

Future development can transform the system into a comprehensive Clinical Decision Support System (CDSS) that assists healthcare professionals in treatment planning, discharge management, resource allocation, and patient follow-up strategies.

H. Personalized Healthcare Recommendations

In addition to predicting readmission risk, future systems may provide personalized recommendations for patient care, follow-up schedules, medication adherence, and preventive interventions. Such capabilities can contribute to reducing avoidable hospital readmissions and improving patient outcomes.

In summary, future enhancements focused on advanced analytics, explainable AI, real-time integration, and healthcare decision support can significantly increase the effectiveness and practical value of the proposed hospital readmission prediction system.

VIII. CONCLUSION

Hospital readmissions represent a significant challenge for healthcare institutions due to their impact on patient outcomes, hospital resources, and overall healthcare costs. Early identification of patients at risk of readmission can assist healthcare providers in implementing preventive measures and improving the quality of care.

This research presented an AI-driven hospital readmission risk scoring system using temporal healthcare indicators and a Random Forest Classification model. The study utilized the UCI Diabetes 130-US Hospitals dataset and selected key healthcare utilization features, including hospitalization duration, laboratory procedures, medication count, outpatient visits, emergency visits, and inpatient visits. Through data preprocessing, feature selection, model training, and evaluation, a predictive framework was developed to estimate the likelihood of patient readmission.

The experimental results demonstrated that the proposed model achieved an accuracy of 86.57%, indicating strong overall classification performance. The confusion matrix analysis further highlighted the model's effectiveness in identifying non-readmitted patients while also revealing challenges associated with detecting minority-class readmission cases. These findings emphasize the importance of addressing class imbalance and enhancing predictive sensitivity in future work.

To improve practical applicability, the trained model was deployed through a Streamlit-based web application that enables real-time risk prediction using patient healthcare indicators. The deployment demonstrates how machine learning models can be transformed into accessible decision-support tools for healthcare professionals.

Overall, the proposed system successfully illustrates the potential of artificial intelligence and machine learning techniques in healthcare analytics. By leveraging historical patient information and temporal healthcare indicators, the framework provides a foundation for intelligent readmission risk assessment and supports data-driven clinical decision-making. Future enhancements involving advanced algorithms, explainable AI, and real-time healthcare integration can further strengthen the effectiveness and adoption of such predictive healthcare systems.

REFERENCES

1. M. K. Khaing, "Machine Learning Based Predictive Models for Hospital Readmission," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 9, no. 3, pp. 45–52, 2019.
2. A. Rajkomar, E. Oren, K. Chen, A. M. Dai, N. Hajaj, P. Hardt, et al., "Scalable and Accurate Deep Learning with Electronic Health Records," *npj Digital Medicine*, vol. 1, no. 18, pp. 1–10, 2018.
3. J. Futoma, S. Morris, and J. Lucas, "A Comparison of Models for Predicting Early Hospital Readmissions," *Journal of Biomedical Informatics*, vol. 56, pp. 229–238, 2015.
4. 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 Analysis," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 5, pp. 1589–1604, 2018.
5. C. Huang, Y. Wang, and L. Zhang, "Hospital Readmission Prediction Using Machine Learning Approaches," *Healthcare Analytics Journal*, vol. 4, pp. 100–112, 2021.
6. D. Bertsimas, M. V. Bjarnadottir, M. A. Kane, J. C. Kryder, R. Pandey, S. Vempala, and

G. Wang, “Algorithmic Prediction of Health-Care Costs with Applications to Hospital Readmissions,” *Operations Research*, vol. 56, no. 6, pp. 1382–1392, 2008.

7. T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, USA, 2016, pp. 785–794.
8. L. Breiman, “Random Forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
9. F. Pedregosa et al., “Scikit-learn: Machine Learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
10. D. Dua and C. Graff, “UCI Machine Learning Repository,” University of California, Irvine, School of Information and Computer Sciences, 2019.
11. T. M. Mitchell, *Machine Learning*. New York, NY, USA: McGraw-Hill, 1997.
12. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
13. J. R. Quinlan, “Induction of Decision Trees,” *Machine Learning*, vol. 1, no. 1, pp. 81–106, 1986.
14. World Health Organization (WHO), “Global Strategy on Digital Health 2020–2025,” Geneva, Switzerland: WHO Publications, 2021.
15. A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, et al., “A Guide to Deep Learning in Healthcare,” *Nature Medicine*, vol. 25, no. 1, pp. 24–29, 2019.