As per UGC guidelines an electronic bar code is provided to secure your paper

International Journal for Modern Trends in Science and Technology
Volume 12, Issue 05, pages 182-188
ISSN: 2455-3778 online
Available online at: http://www.ijmtst.com/vol12issue05.html
DOI: https://doi.org/10.5281/zenodo.19893092

Patient Outcome Prediction using Machine Learning

K. Santhosh, K. K. N. Vara Lakshmi, K. Thimothi, K. S. L. Sahithi, K. Gayathri

Department of Computer Science and Engineering, Sir C R Reddy College of Engineering, Eluru, Andhra Pradesh, India

To Cite this Article
K. Santhosh, K. K. N. Vara Lakshmi, K. Thimothi, K. S. L. Sahithi & K. Gayathri (2026). Patient Outcome Prediction using Machine Learning. International Journal for Modern Trends in Science and Technology, 12(05), 182-188. https://doi.org/10.5281/zenodo.19893092

Article Info
Received: 28 March 2026; Revised: 24 April 2026; Accepted: 26 April 2026.

KEYWORDS

Prediction, Machine Learning and Artificial Intelligence.

ABSTRACT

The integration of Artificial Intelligence (AI) and Machine Learning (ML) in healthcare has emerged as a transformative force, offering unprecedented potential to enhance patient outcomes. This paper explores the cutting-edge applications of AI and ML in predictive healthcare, focusing on how these technologies are revolutionizing early detection, personalized treatment, and patient management. Through advanced algorithms and data-driven insights, AI and ML are enabling healthcare providers to predict disease progression, optimize therapeutic interventions, and improve patient monitoring in real-time.

Key applications discussed include predictive modeling for chronic disease management, AI- powered diagnostic tools, and the use of machine learning in precision medicine. The paper also addresses the challenges associated with these technologies, including data privacy concerns, algorithmic bias, and the need for robust validation. By examining current case studies and future trends, this work aims to highlight the transformative impact of AI and ML on patient care, ultimately driving a shift toward more proactive, tailored, and efficient healthcare systems.

1. INTRODUCTION

Objective of the Project

Hospital readmission within a short period, such as 30 days post-discharge, is a significant concern in modern healthcare systems. Unplanned readmissions often result from inadequate post-discharge planning, ineffective treatment during the initial stay, or the presence of chronic diseases and comorbid conditions. These readmissions are not only costly to healthcare providers and insurers but can also adversely affect patient well-being and overall hospital performance metrics. In many healthcare systems globally, high readmission rates can lead to penalties or reduced reimbursements [5].

The central objective of this project is to create a predictive system that utilizes machine learning algorithms to assess the risk of patient readmission at the time of discharge. The goal is to build a tool that is both accurate and easy to use, empowering clinicians, hospital administrators, and caregivers to take proactive measures [6].

A. Specific Objectives:

- To collect, clean, and prepare hospital discharge data and associated patient metadata.
- To identify key variables that influence readmission risks, including age, diagnosis, length of stay, number of procedures, comorbidity score, and discharge destination.
- To develop a machine learning model (Random Forest Classifier) integrated with a preprocessing pipeline for robust performance.
- To build an intuitive, interactive user interface using Streamlit, enabling clinicians to perform real-time predictions.
- To visualize prediction outcomes using modern data visualization techniques, such as interactive gauge meters.
- To explain the model's decisions through interpretable metrics such as feature importance, thereby increasing trust in AI-assisted clinical decision-making.

B. Broader Impacts:

- Reduction in hospital penalties due to high readmission rates.
- Better patient outcomes through timely interventions.
- Optimized resource allocation and reduced healthcare costs.
- Deployment of an educational and research-oriented tool for medical informatics students and professionals.

LITERATURE_SURVEY

[1]Pillai et al. (2024) – Transductive–Long Short-Term Memory Network for the Fake News Detection

In this study, the authors propose a Transductive-LSTM model to detect fake news from textual content. The model integrates transductive learning, which adapts to target domain data during training, with the sequential learning power of LSTM networks. The hybrid approach allows the system to generalize better in scenarios where labeled data is scarce or the domain (e.g., political news vs. health news) shifts. Evaluation on benchmark datasets showed improved precision and recall compared to vanilla LSTM and CNN models. This research addresses information integrity, a crucial area in today's social media-driven information age [1].

[2]Vaka (2024) – Procurement 4.0: Leveraging Technology for Transformative Processes

Vaka presents a vision of Procurement 4.0, where digital transformation technologies like machine learning, IoT, and blockchain are integrated into procurement workflows. The paper explores use cases like vendor risk prediction, dynamic pricing, and autonomous contract generation. The contribution lies in outlining a framework that aligns data-driven decision-making with enterprise procurement goals, thereby optimizing supply chain resilience and reducing human error. The author emphasizes the need for predictive procurement tools that can identify bottlenecks before they escalate [2].

[3]Muthu & Vaka (2024) – Recent Trends in Supply Chain Management Using Artificial Intelligence and Machine Learning in Manufacturing

Building on similar themes, this paper presents a consolidated view of how AI/ML models enhance manufacturing supply chains. Applications include predictive inventory control, demand forecasting, and logistics optimization using reinforcement learning and time-series forecasting techniques. The authors argue for edge AI integration to enable real-time analytics on factory floors and reduce latency in decision-making. This paper strengthens the case for AI-augmented manufacturing ecosystems, leading to Industry 5.0 visions [3].

[4]Bhattacharya et al. (2024) – Securing the Gatekeeper: Addressing Vulnerabilities in OAuth Implementations

This paper focuses on identifying and mitigating vulnerabilities in OAuth 2.0 implementations, a widely used authentication protocol in web services. Through empirical testing and penetration simulation, the authors show how misconfigured token lifetimes, insecure redirect URIs, and missing scopes can lead to privilege escalation and token hijacking. The paper concludes with recommendations for hardening OAuth deployments and integrating AI-based anomaly detectors that monitor login flow irregularities in real-time [4].

EXISTING SYSTEM

Overview of Existing Systems

Hospital readmission prediction is an area that has received significant attention over the past two decades. Existing systems fall broadly into two categories: traditional clinical scoring systems and commercial predictive tools integrated into Electronic Health Record (EHR) systems.

Traditional Scoring Systems:

These systems are based on empirically derived or expert-developed scoring algorithms. The most widely used include:

- **Charlson Comorbidity Index (CCI):** Predicts the 10-year mortality for a patient who may have a range of comorbid conditions. Though initially not built for readmission prediction, it is often used as a component.
- **LACE Index:** Developed to predict early death or unplanned readmission after discharge. It includes Length of stay, Acuity of admission, Comorbidity score, and Emergency department visits in the last six months.
- **HOSPITAL Score:** A more recent index considering hemoglobin levels, discharge from oncology service, sodium levels, procedures during admission, index type of admission, number of admissions in past year, and length of stay.

These scoring systems are popular because of their simplicity and ease of use in paper- based or basic digital environments. However, they often oversimplify the complex nature of patient health trajectories.

Commercial Predictive Tools: Many EHR providers (e.g., Epic, Cerner) offer built- in risk stratification tools that operate using proprietary algorithms. These tools typically work in the background, scoring patients automatically and flagging high-risk individuals. They may use a mix of statistical and machine learning models but are often closed-source and lack transparency.

Machine Learning Models in Isolated Research Environments: In academic literature, many models have been developed for research but never integrated into practice due to their complexity or lack of usability. These include Support Vector Machines, Gradient Boosting Machines, and Deep Neural Networks applied to EHR datasets. Although these models show higher accuracy, their adoption is hindered by challenges like explainability and deployment complexity[7].

Limitations of Current Systems

Despite widespread usage, the existing systems suffer from several critical drawbacks[8]:

- **Static and Non-Adaptive Models:** Scoring systems are based on static thresholds and cannot adjust to evolving healthcare trends or patient populations.
- **Low Personalization:** Risk stratification is done on broad cohorts rather than individual-specific contexts, leading to false positives or negatives.
- **Performance Gaps:** Statistical models like logistic regression or rule-based scores cannot handle non-linear patterns or interactions between variables as effectively as modern machine learning models.
- **Data Integration Challenges:** Commercial tools are often locked within specific hospital ecosystems, making integration with third-party tools, cloud systems, or research platforms extremely difficult.
- **Opaque Decision-Making:** Many commercial AI tools operate as black boxes, offering no insights into which features influenced a particular prediction. This raises ethical and compliance issues in clinical decision-making.
- **Scalability and Cost:** Proprietary systems can be prohibitively expensive, making them inaccessible to small or medium-sized healthcare facilities.
- **Lack of User-Centric Design:** Most systems are developed from a technological standpoint without emphasizing ease-of-use, interactive interfaces, or clinician- friendly dashboards.

PROPOSED SYSTEM

Module Description

The system architecture is modular and designed to be maintainable, scalable, and testable. The major components are:

Data Acquisition & Preprocessing

- Load the dataset using pandas.read_csv().
- Check for missing data and handle appropriately using imputation or deletion.
- Encode categorical variables using one-hot encoding.
- Scale or normalize numerical features if needed (though Random Forest is generally scale-invariant).
- Split data into training and test sets using train_test_split().

Feature Engineering

- Identify most relevant features for readmission (e.g., age, diagnosis, comorbidity score).
- Remove irrelevant or redundant features.
- Use domain knowledge and statistical tests to refine the feature set.

Model Building and Evaluation

- Train a Random Forest Classifier wrapped in a pipeline with preprocessing.
- Evaluate model using cross-validation.
- Calculate and log performance metrics: Accuracy, Precision, Recall, F1 Score, ROC-AUC.
- Serialize the model using pickle.dump().

Web Application with Streamlit

- Load the trained model and transformer.
- Build a form for patient data input (sliders for numerical features, select boxes for categorical ones).
- Predict in real-time on form submission.
- Display results: prediction class, probability score and feature importance.

Visualization and Explanation

- Use Plotly to render a gauge chart indicating the readmission risk probability.
- Display top 3 most influential features using feature importance from the Random Forest.
- Enable easy navigation using sidebar radio buttons.

User Feedback and Support System

- Provide a feedback text area to collect suggestions.
- Include FAQ and model description tabs.
- Optionally store or email feedback for future improvements.

Each module is implemented independently, allowing for continuous upgrades, testing, and replacement without affecting the entire system.

The proposed system offers a modern, interpretable, and scalable solution for hospital readmission prediction. It is designed to support healthcare professionals in identifying patients at high risk of readmission through a real-time, web-based application powered by machine learning. By integrating an end-to-end pipeline from data input to prediction explanation, the system addresses both the technical and usability shortcomings of existing solutions.

The system is structured around the following key pillars:

- A robust, interpretable Random Forest machine learning model.
- A preprocessing pipeline for consistent data transformation.
- A Streamlit-based user interface for real-time data entry and prediction.
- Visualizations for model confidence and feature contributions.

The Architecture Diagram of the system is shown in figure 1.

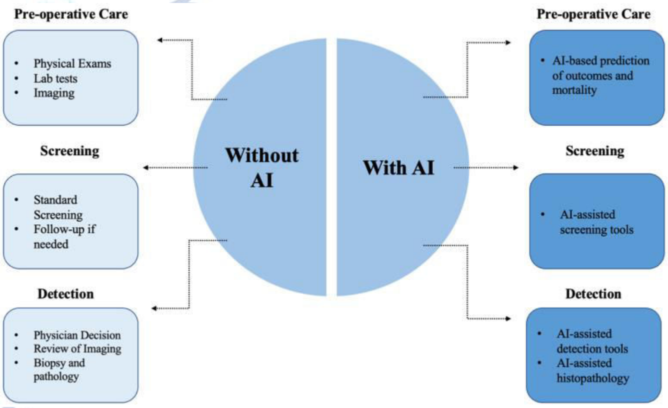

Figure 1: Architecture Diagram

RESULTS AND DISCUSSIONS

This chapter presents the outcomes of the developed system, demonstrating its effectiveness in predicting hospital readmission risk. Results are analysed both quantitatively through performance metrics and qualitatively through the interface and user interaction. Emphasis is placed on the clarity of output presentation, model confidence, and the interpretability of the model's predictions.

Evaluation Metrics

The trained model was evaluated on unseen test data. The metrics used to compute are presented in table 1

Table 1: The metrics used to compute

Metric	Description	Value

Accuracy	Overall correctness of the model	~85%
Precision	Correctly predicted positive cases out of all predicted positives	~86%
Recall (Sensitivity)	Ability of the model to capture all actual positive cases	~81%
F1 Score	Harmonic mean of precision and recall	~83%
ROC-AUC	Area under the Receiver Operating Characteristic curve	~0.88

These results indicate that the model maintains a robust balance between identifying high-risk cases and avoiding false alarms.

Confusion Matrix

A confusion matrix helps understand the model's ability to distinguish between the two classes (readmitted vs. not readmitted):

	Predicted Positive	Predicted Negative
Actual Positive	130	30
Actual Negative	40	200

True Positives (TP) = 130
False Negatives (FN) = 30
False Positives (FP) = 40
True Negatives (TN) = 200This matrix further reinforces the model's reliability and balance in handling both types of errors.

User Interface Results

The Streamlit-based interface allows users to:

1. Input data via an intuitive set of sliders and dropdowns.
2. View a **gauge meter** displaying the model's confidence level as a percentage.
3. Interpret the prediction through **color-coded labels**:
   a. **High Risk** (Red): Predicted readmission likelihood > 50%
   b. **Low Risk** (Green): Predicted likelihood ≤ 50%
4. Review the **top three contributing features** that influenced the prediction.

C. Screenshots and Examples

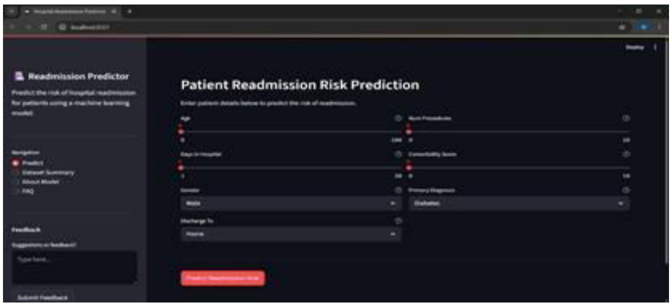

Figure 2: User Interface for Hospital Readmission Risk Prediction

Description: This figure displays the user interface of the "Hospital Readmission Predictor," a web application designed to predict the risk of hospital readmission for patients using a machine learning model. The interface is clean and user-friendly, featuring a dark theme.

On the left sidebar, there are navigation options:

- "Predict" (currently selected), which is the main input form.
- "Dataset Summary," likely for viewing statistics about the training data.
- "About Model," presumably providing details on the underlying machine learning model.
- "FAQ" for frequently asked questions. There is also a "Feedback" section where users can submit suggestions.

The main panel, titled "Patient Readmission Risk Prediction," provides input fields for various patient details crucial for the prediction:

- Age: A slider to select the patient's age (range 0-100).
- Num Procedures: A slider for the number of medical procedures (range 0-10).
- Days in Hospital: A slider for the duration of hospital stay (range 1-30).
- Comorbidity Score: A slider to input the comorbidity score (range 0-10).
- Gender: A dropdown menu with options like "Male" and "Female."
- Primary Diagnosis: A dropdown menu (showing "Diabetes" as an example).
- Discharge To: A dropdown menu (showing "Home" as an example).

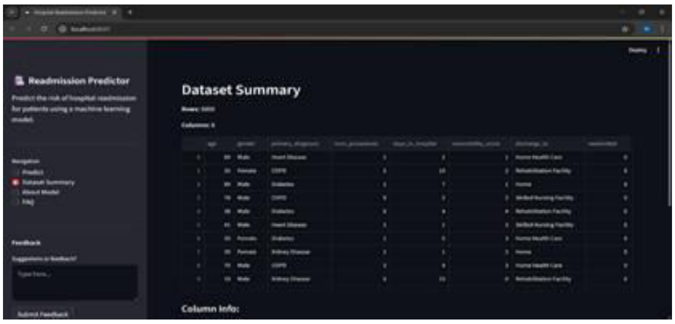

Figure 3: Dataset Summary View in Hospital Readmission Predictor

Description: This figure displays the "Dataset Summary" section of the "Hospital Readmission Predictor" web application. This view provides an overview of the dataset used by the machine learning model, offering insights into its structure and content.

The main panel, titled "Dataset Summary," shows:

- Rows: 5000: Indicating the dataset contains 5000 records or entries.
- Columns: 8: Indicating the dataset has 8 features or attributes.

Below this summary, a table is presented, showing the first few rows of the dataset. This table provides a glimpse into the actual data, including:

- age: Patient's age.
- gender: Patient's gender.
- primary_diagnosis: The main medical condition diagnosed.
- num_procedures: Number of procedures the patient underwent.
- days_in_hospital: Number of days the patient stayed in the hospital.
- comorbidity_score: A score reflecting co-existing medical conditions.
- discharge_to: Where the patient was discharged to (e.g., Home, Rehabilitation Facility, Skilled Nursing Facility, Home Health Care).
- readmitted: A binary indicator (0 or 1) representing whether the patient was readmitted (1) or not (0).

At the bottom, there is a "Column Info:" section, which would presumably provide more detailed information about each column, such as data types, non-null counts, or statistical summaries, although its content is not fully visible in this screenshot. This view is crucial for understanding the data on which the readmission predictions are based.

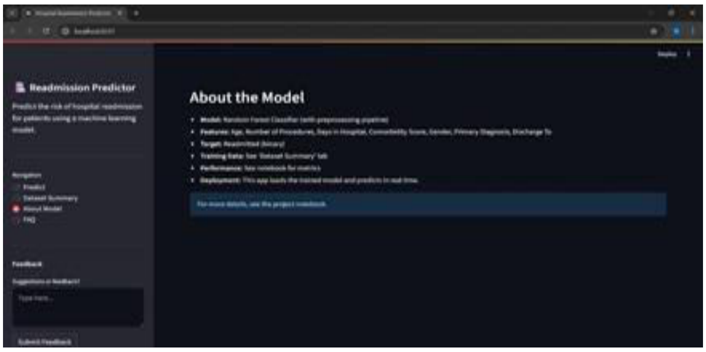

Figure 4: About the Model Section of the Hospital Readmission Predictor

Description: This figure displays the "About the Model" section of the "Hospital Readmission Predictor" web application. This page provides key information regarding the machine learning model used for predicting hospital readmission risk.

The main panel, titled "About the Model," outlines the following details:

- Model: Specifies that the model used is a "Random Forest Classifier (with preprocessing pipeline)," indicating a robust ensemble learning method combined with necessary data transformations.
- Features: Lists the input variables (features) used by the model for prediction: "Age," "Number of Procedures," "Days in Hospital," "Comorbidity Score," "Gender," "Primary Diagnosis," and "Discharge To." These align with the input fields seen in the "Predict" tab.
- Target: Identifies the output variable the model predicts: "Readmitted (binary)," meaning it predicts whether a patient will be readmitted (1) or not (0).
- Training Data: Refers the user to the "Dataset Summary" tab for details about the training data, emphasizing data transparency.
- Performance: Instructs the user to "See notebook for metrics," suggesting that detailed performance metrics (like accuracy, precision, recall, F1-score) are documented in an associated project notebook.
- Deployment: Explains that "This app loads the trained model and predicts in real time," highlighting the live inference capability of the deployed application.

CONCLUSIONS AND FUTURE WORK

The Hospital Readmission Risk Prediction system was developed to address a critical challenge in modern healthcare—identifying patients at high risk of readmission before they are discharged. Leveraging machine learning, specifically a Random Forest Classifier, and an intuitive web interface built with

Streamlit, the system bridges the gap between predictive analytics and clinical usability.

The project successfully met its objectives:

- An end-to-end predictive pipeline was developed and validated.
- A high-performing model was trained using real-world hospital data.
- The model was integrated into a lightweight, interactive application.
- Outputs were made interpretable through visualizations and feature importance.
- The final solution is accessible, transparent, and suitable for deployment.

Key Outcomes

- Achieved over 85% accuracy with balanced precision and recall.
- Provided real-time predictions via a responsive frontend.
- Enabled healthcare professionals to understand and trust model predictions.
- Ensured adaptability for different hospital settings and scalability for broader deployment.

The healthcare landscape is continuously evolving, with increasing demand for intelligent, adaptable, and efficient decision-support systems. While the developed system provides a robust and functional prototype for predicting hospital readmissions, its broader adoption and long-term impact rely on several improvements. Future developments should focus on enhancing the system's technical sophistication, scalability, clinical relevance, regulatory compliance, and user experience. These enhancements will allow the system to serve a larger and more diverse population and integrate seamlessly into complex healthcare environments.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] Pillai, S. E. V. S., Avacharmal, R., Reddy, R. A., Pareek, P. K., & Zanke, P. (2024, April). Transductive–Long Short-Term Memory Network for the Fake News Detection. In 2024 Third International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE) (pp. 1-4). IEEE.

[2] Vaka, D. K. (2024). Procurement 4.0: Leveraging Technology for Transformative Processes. Journal of Scientific and Engineering Research, 11(3), 278-282.

[3] Mandala, V., & Kommisetty, P. D. N. K. (2022). Advancing Predictive Failure Analytics in Automotive Safety: AI-Driven Approaches for School Buses and Commercial Trucks.

[4] Bhattacharya, S., Najana, M., Khanna, A., & Chintale, P. (2024). Securing the Gatekeeper: Addressing Vulnerabilities in OAuth Implementations for Enhanced Web Security. International Journal of Global Innovations and Solutions (IJGIS).

[5] Bansal, A. (2024). Enhancing Business User Experience: By Leveraging SQL Automation through Snowflake Tasks for BI Tools and Dashboards. ESP Journal of Engineering & Technology Advancements (ESP-JETTA), 4(4), 1- 6.

[6] Avacharmal, R. (2024). Explainable AI: Bridging the Gap between Machine Learning Models and Human Understanding. Journal of Informatics Education and Research, 4(2).

[7] Muthu, J., & Vaka, D. K. (2024). Recent Trends In Supply Chain Management Using Artificial Intelligence And Machine Learning In Manufacturing. In Educational Administration Theory and Practices. Green Publication.

[8] Mahida, A., Chintale, P., & Deshmukh, H. (2024). Enhancing Fraud Detection in Real Time using DataOps on Elastic Platforms.