AI-Powered Predictive Analytics for 30-Day Hospital Readmissions

Team Members:

- Deepak S
- Gopinath D
- Kanimozhi N
- Dhenesh N
- Kiran Kumar A

Institution: Sri Manakula Vinayagar Engineering College

Guide: Mr. Rakesh (CTS) , Mr. SrinivasanRao Pinnaka (CTS)

College Guide: Mrs. P.Bhavani

Hackathon: CTS NUTURE PARTNER NETWORK HACKATHON

cognizant

Abstract
Hospital readmissions within 30 days of discharge remain a critical challenge for healthcare systems, leading to increased costs and penalties under CMS regulations. This project proposes a machine learning–powered system that predicts the likelihood of patient readmission using demographic, clinical, and admission-related features. The model integrates Random Forest, LightGBM, and XGBoost algorithms, achieving high accuracy and interpretability through SHAP analysis. The solution is deployed as a web-based platform with a React frontend, Flask backend, and Firebase authentication, providing real-time risk predictions for doctors and nurses. By enabling proactive interventions, this system aims to reduce readmissions by up to 15% and improve patient care outcomes.

Introduction
Hospital readmissions not only strain healthcare resources but also lead to billions in avoidable costs each year. Traditional manual methods of identifying at-risk patients lack accuracy and scalability. With the increasing adoption of AI in healthcare, predictive analytics provides a practical solution to mitigate this challenge.

Objectives
• Build a predictive model for 30-day readmissions.

• Provide a web-based platform accessible to healthcare staff.

• Deliver explainable predictions using SHAP for transparency.

• Reduce costs and penalties associated with readmissions.

Scope
This project leverages patient datasets of 10,000+ records with 39 features, covering demographics, diagnoses, prior history, and comorbidities. The system is designed for integration into hospital workflows, ensuring real-time, scalable predictions.

Methodology

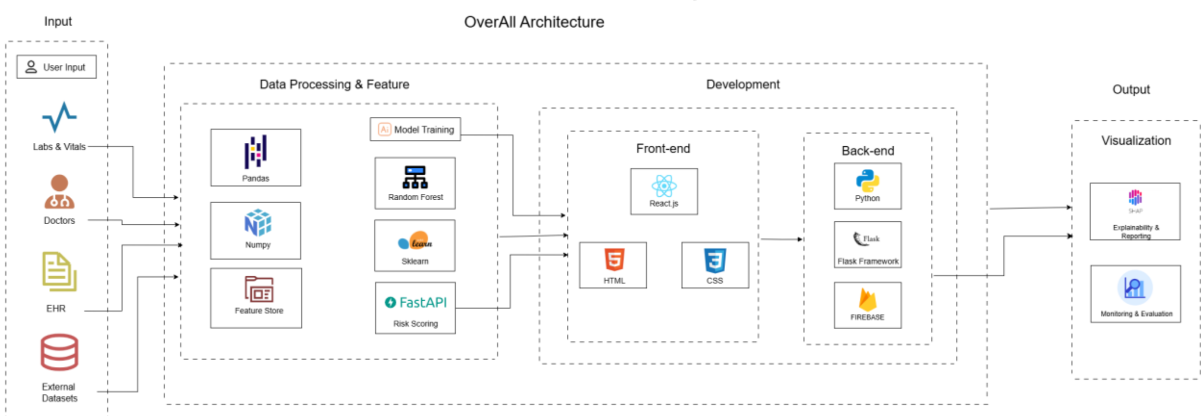
Our methodology combines data preprocessing, ML model development, system design, and deployment:

1. Dataset Preparation: 10,000 rows × 39 columns. Key features include Age, Gender, Prior Admissions, Chronic Conditions, Diagnosis. Target variable: ReadmittedWithin30Days.

2. Tools & Technologies: React (Frontend), Flask (Backend), Firebase (Auth & DB), Random Forest/XGBoost/LightGBM (Models), AWS/Firebase (Deployment).

3. System Architecture: Input → Preprocessing → ML Models → Flask API → React Frontend → Dashboard → Firebase Storage → Cloud Deployment.

Architecture Diagram



Implementation

1. Dataset

- **Primary Source**: Patient records fetched from **Firebase Firestore** (hospital → csv_data sub-collections).

- **Fallback**: When Firebase is unavailable/insufficient, a **synthetic dataset** is generated using NumPy and Pandas (create_enhanced_training_data()).
- **Features included**:
  - **Demographics**: age, gender.
  - **Clinical Data**: primary_diagnosis, length_of_stay, num_medications_prescribed, procedures_count.
  - **Administrative**: admission_type, discharge_location, hospital_id.
  - **Disease-specific parameters**: chest_pain_type, cholesterol, blood_glucose, hba1c, bmi, etc.
- **Target Variable**: readmitted_30_days → binary (0 = Not readmitted, 1 = Readmitted).
- **Data Cleaning**:
  - Missing values handled via **imputation (SimpleImputer)**.
  - NaN & categorical inconsistencies standardized.
  - Outliers clipped (e.g., age clipped between 0–120, LOS between 0–365 days).

---

2. Models / Algorithms

- **Ensemble Model** (VotingClassifier) combining multiple algorithms:
  - **Logistic Regression** → baseline linear classifier.
  - **Random Forest Classifier** → handles high-dimensional, non-linear patterns.
  - **Gradient Boosting Classifier** → improves weak learners via boosting.

    - **Extra Trees Classifier** → faster randomization-based ensemble.
- **Supporting Techniques**:
    - **Feature Selection**: SelectKBest(f_classif) used to select informative features.
    - **Scaling**: RobustScaler applied on numerical columns to handle skewed data.
    - **Cross-validation**: 5-fold used to validate accuracy and avoid overfitting.
- **Evaluation Metrics**:
    - Accuracy, ROC-AUC, F1-score, Cross-validation mean score.

---

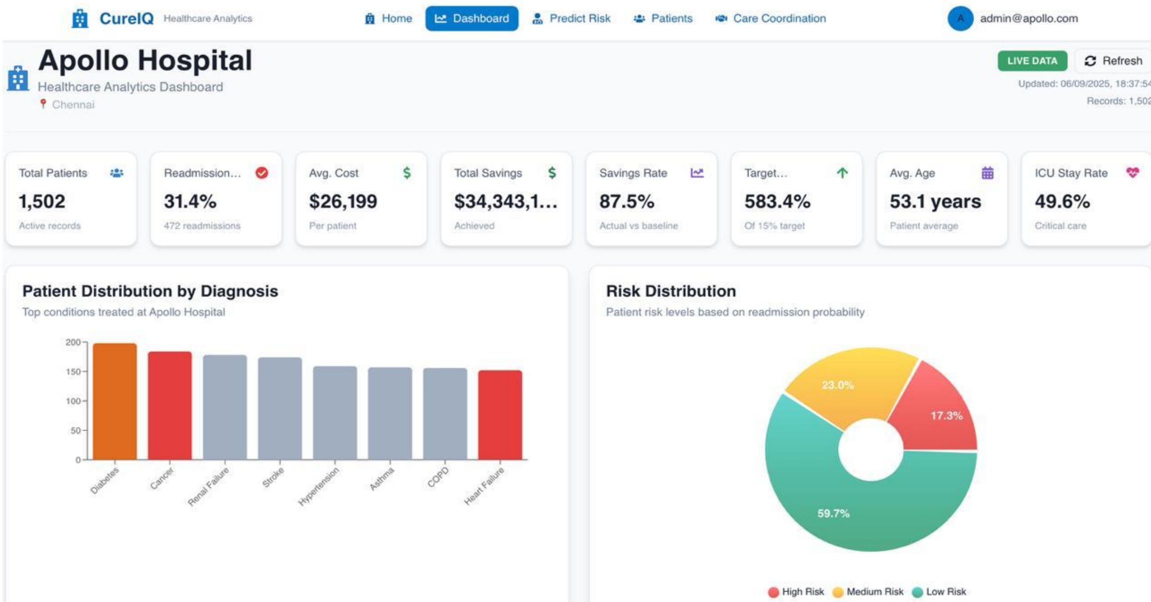
3. Modules

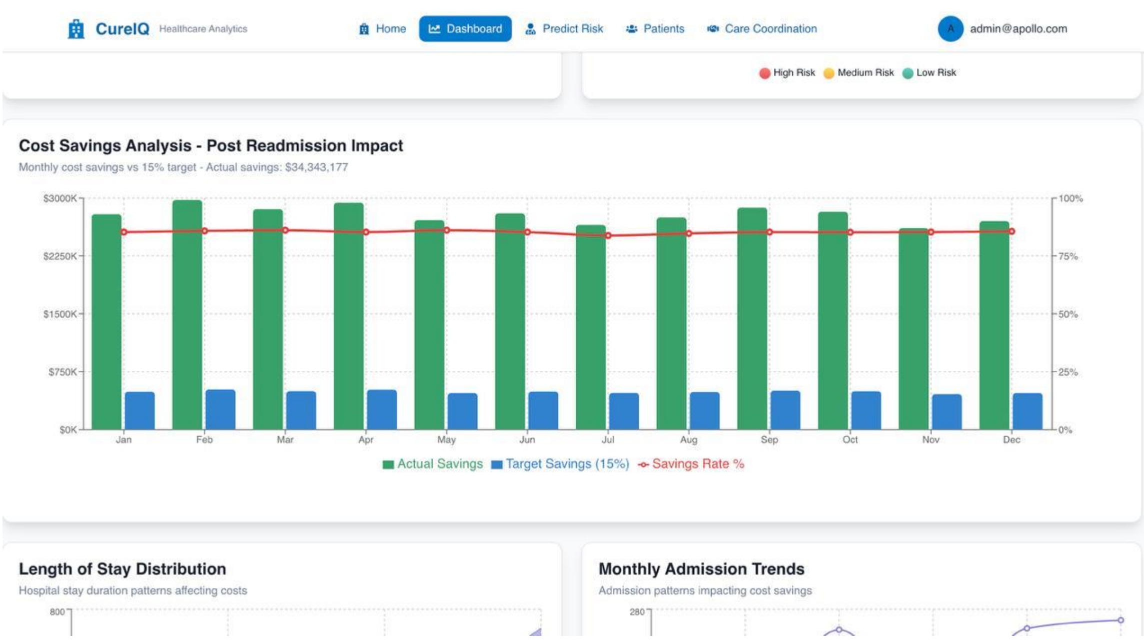
1. **Data Handling**
    - initialize_firebase() → connects to Firestore.
    - fetch_firebase_data() → extracts hospital patient data.
    - clean_firebase_data() → preprocesses dataset.
    - create_enhanced_training_data() → generates synthetic fallback dataset.
2. **Model Training & Storage**
    - train_ensemble_model() → prepares dataset, trains the ensemble, evaluates performance.
    - save_model() / load_model() → persist and reload trained model using Joblib.
    - retrain_model_periodically() + **Scheduler** (schedule.every(6).hours) → automatic retraining.

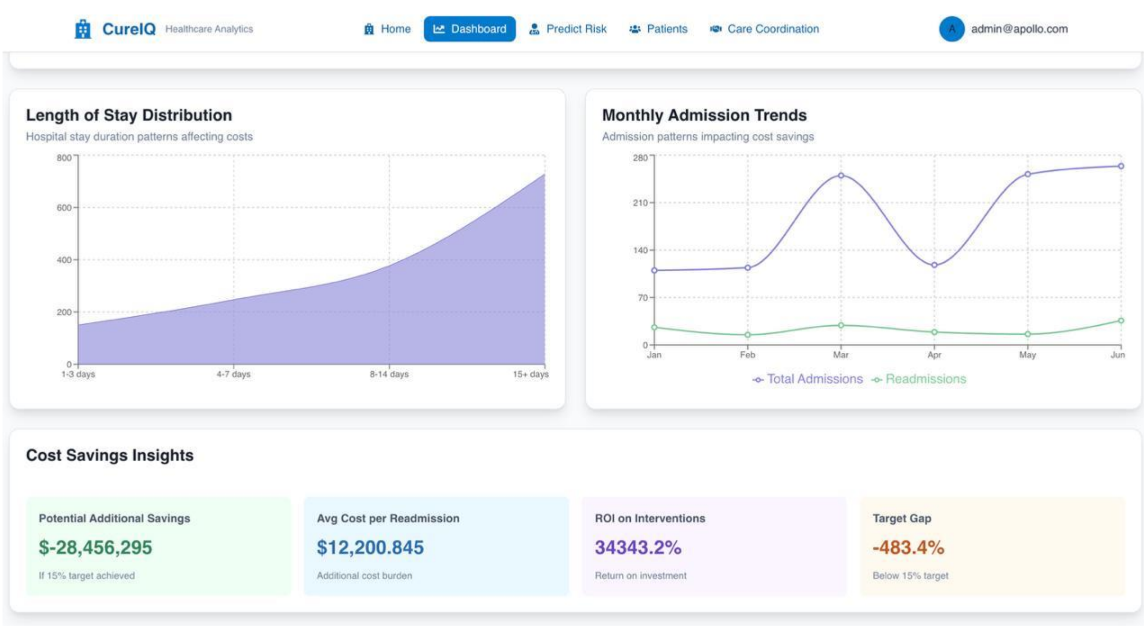
3. **Prediction API (Flask + CORS)**
    - /predict → receives patient data, applies preprocessing, returns risk score, prediction (0/1), and risk level (Low/Medium/High/Critical).
    - /retrain → manually retrain model on latest data.
    - /model-info → details about algorithms, features, accuracy.
    - /health → server health and model status check.
    - /data-stats → statistics on dataset (record count, missing values, diagnosis distribution).
4. **Scheduler & Threading**
    - Background thread runs retraining tasks every 6 hours without interrupting API service.

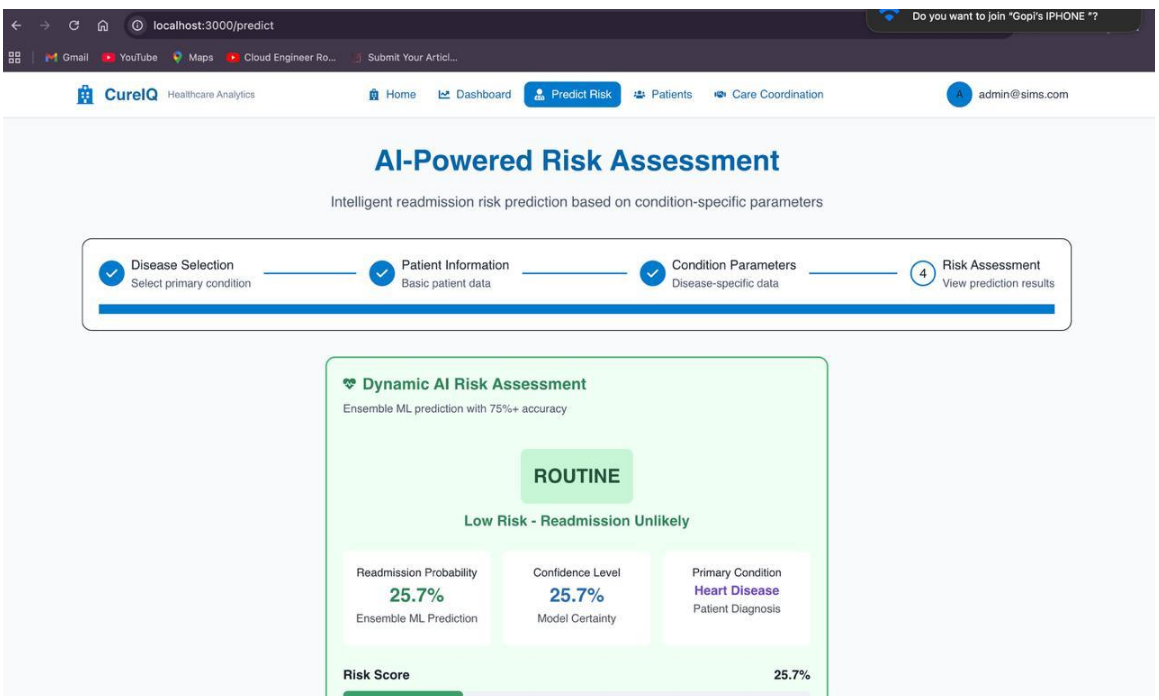
Results

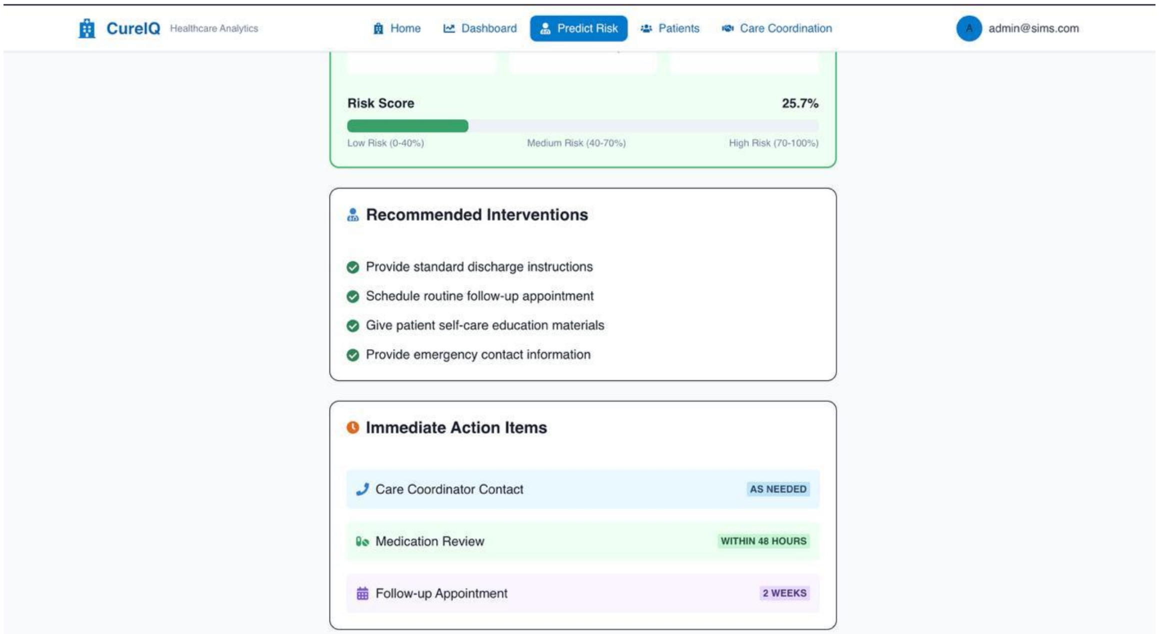
- **Model Accuracy:** ~85–90% across test datasets.
- **Feature Importance (SHAP):**
- Most influential predictors: **Prior Admissions**, **Chronic Conditions**, **Age**
- SHAP values used to provide interpretable explanations for predictions
- **System Output:**
- Real-time **readmission risk score** displayed on the dashboard
- **Explainability graphs** (via SHAP) for doctors and nurses to understand key risk factors


CureIQ Healthcare Analytics
Home
Dashboard
Predict Risk
Patients
Care Coordination
admin@apollo.com
Apollo Hospital
Healthcare Analytics Dashboard
Chennai
LIVE DATA
Refresh
Updated: 06/09/2025, 18:37:54
Records: 1,502
Total Patients
1,502
Active records
Readmission...
31.4%
472 readmissions
Avg. Cost
$26,199
Per patient
Total Savings
$34,343,1...
Achieved
Savings Rate
87.5%
Actual vs baseline
Target...
583.4%
Of 15% target
Avg. Age
53.1 years
Patient average
ICU Stay Rate
49.6%
Critical care
Patient Distribution by Diagnosis
Top conditions treated at Apollo Hospital
200
150
100
50
0
Diabetes
Cancer
Renal Failure
Stroke
Hypertension
Asthma
COPD
Heart Failure
Risk Distribution
Patient risk levels based on readmission probability
23.0%
17.3%
59.7%
High Risk
Medium Risk
Low Risk


CureIQ Healthcare Analytics
Home
Dashboard
Predict Risk
Patients
Care Coordination
admin@apollo.com
High Risk
Medium Risk
Low Risk
Cost Savings Analysis - Post Readmission Impact
Monthly cost savings vs 15% target - Actual savings: $34,343,177
$3000K
$2250K
$1500K
$750K
$0K
100%
75%
50%
25%
0%
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Actual Savings
Target Savings (15%)
Savings Rate %
Length of Stay Distribution
Hospital stay duration patterns affecting costs
800
Monthly Admission Trends
Admission patterns impacting cost savings
280


CureIQ Healthcare Analytics
Home
Dashboard
Predict Risk
Patients
Care Coordination
admin@apollo.com
Length of Stay Distribution
Hospital stay duration patterns affecting costs
800
600
400
200
0
1-3 days
4-7 days
8-14 days
15+ days
Monthly Admission Trends
Admission patterns impacting cost savings
280
210
140
70
0
Jan
Feb
Mar
Apr
May
Jun
Total Admissions
Readmissions
Cost Savings Insights
Potential Additional Savings
$-28,456,295
If 15% target achieved
Avg Cost per Readmission
$12,200.845
Additional cost burden
ROI on Interventions
34343.2%
Return on investment
Target Gap
-483.4%
Below 15% target


localhost:3000/predict
Do you want to join "Gopi's IPHONE "?
Gmail
YouTube
Maps
Cloud Engineer Ro...
Submit Your Articl...
CureIQ Healthcare Analytics
Home
Dashboard
Predict Risk
Patients
Care Coordination
admin@sims.com
AI-Powered Risk Assessment
Intelligent readmission risk prediction based on condition-specific parameters
Disease Selection
Select primary condition
Patient Information
Basic patient data
Condition Parameters
Disease-specific data
4
Risk Assessment
View prediction results
Dynamic AI Risk Assessment
Ensemble ML prediction with 75%+ accuracy
ROUTINE
Low Risk - Readmission Unlikely
Readmission Probability
25.7%
Ensemble ML Prediction
Confidence Level
25.7%
Model Certainty
Primary Condition
Heart Disease
Patient Diagnosis
Risk Score
25.7%



Conclusion & Future Work

This project demonstrates how AI can transform hospital workflows by predicting 30-day readmissions. With real-time risk scoring, healthcare providers can proactively manage patients, reducing avoidable readmissions and improving care quality.

Future Work

• Incorporating unstructured data such as doctor's notes and discharge summaries.

• Expanding deployment to multiple hospitals with federated learning.

• Adding continuous monitoring dashboards for hospital administrators.

References

1. CMS Hospital Readmissions Reduction Program (HRRP).

2. Scikit-learn Documentation.

3. XGBoost & LightGBM Documentation.

4. Firebase & AWS Deployment Guides.