Content

Introduction

The **Health AI Suite** is a healthcare analytics project that uses machine learning and deep learning techniques to analyze patient-related data and generate meaningful insights. The goal of this project is to support clinical decision-making through early disease detection, estimation of hospital length of stay, grouping of similar patients, mining of medical co-occurrence patterns, analysis of medical images, modeling patient health trends over time, understanding patient feedback.

The system is built using synthetic and publicly available healthcare data, ensuring data privacy and ethical compliance. By combining data preprocessing, model development, evaluation, and deployment through interactive dashboards and APIs, this project demonstrates how AI can be applied to solve real-world healthcare problems in a practical and scalable manner.

The Health AI Suite is designed as a modular system where each model focuses on a specific healthcare task while working as part of a single pipeline. It combines machine learning and deep learning techniques to handle different types of healthcare data such as tabular records, medical images, time-series data, and patient feedback. This approach helps in generating useful predictions and insights that can assist doctors and healthcare teams in day-to-day decision-making. By integrating all components into one platform, the project shows how AI-based solutions can be practically applied in real healthcare scenarios.

Requirement Specification

Hardware Requirements

- 64-bit system (Windows / macOS)
- Minimum 8 GB RAM (4 GB minimum for basic execution)
- Minimum 10 GB free disk space

Software Requirements

- Operating System: Windows 10/11 or macOS
- Python 3.8 or above
- Jupyter Notebook / JupyterLab
- Visual Studio Code or any Python IDE
- Libraries: NumPy, Pandas, Scikit-learn, TensorFlow / Keras, PyTorch (optional)
- FastAPI
- Streamlit

Existing System

In the existing healthcare system, most clinical data analysis and decision-making processes rely on manual evaluation or basic rule-based systems. Patient records, medical reports, and feedback are often reviewed individually, making it difficult to extract meaningful insights from large volumes of data. This approach increases the time required for diagnosis, reporting, and decision-making, while also limiting the ability to identify patterns or predict future health outcomes.

Since the system lacks automation and advanced analytics, it is challenging to track patient trends, group similar patients, or provide timely predictive insights. Manual handling of healthcare data also increases the chances of errors and makes it difficult to manage historical records efficiently. Therefore, an automated and intelligent system is required to improve accuracy, efficiency, and scalability in healthcare data analysis.

Health AI Suite — Intelligent Analytics for Patient Care

1. Project Objective

Health AI Suite is an end-to-end AI/ML system designed to analyze heterogeneous healthcare data—including structured EHR data, medical images, time-series vitals, and patient feedback—to support clinical decision-making, patient monitoring, and hospital operations.

The goal of this project is not to rely on real patient data, but to demonstrate system design, Modeling choices, preprocessing rigor, and deployment readiness using synthetic and publicly available datasets, as permitted in healthcare AI workflows.

---

2. Problem Statement

Healthcare systems generate diverse data streams that are often analyzed in isolation. This project addresses the challenge of integrating multiple AI paradigms to:

- Predict patient risk categories (classification)
- Estimate hospital length of stay (regression)
- Identify patient cohorts with similar clinical profiles (clustering)
- Detect medical patterns and associations
- Analyze medical images for diagnostic support
- Monitor patient deterioration using time-series modelling
- Extract sentiment from patient feedback

By combining traditional machine learning, deep learning, and NLP models into a unified system, Health AI Suite demonstrates how AI can enhance clinical insight, operational efficiency, and patient engagement.

---

3. System Architecture Overview

The system follows a modular, service-oriented architecture:

- Data Layer
  Synthetic and publicly available datasets (tabular, image, text, time-series)
- Modeling Layer
  Independent ML/DL models trained per task:

    - Classification, regression, clustering
    - CNN for imaging
    - LSTM for time-series
    - CNN/LSTM for sentiment analysis
- Inference Layer (Fast API)
    - Centralized REST API
    - Task-specific endpoints
    - Consistent preprocessing aligned with training
- Presentation Layer (Streamlit)
    - Interactive dashboards
    - Real-time monitoring simulation
    - Visualization of predictions and probabilities

This separation ensures reproducibility, scalability, and maintainability.

---

4. Implemented Modules (Scope Alignment)

The project implements multiple modules from the problem specification, exceeding the mandatory two-module requirement and demonstrating advanced initiative.

Machine Learning Modules

1. Risk Stratification (Classification)
   Predicts patient risk levels using demographic and laboratory features.
2. Length of Stay Prediction (Regression)
   Estimates hospitalization duration based on clinical and diagnostic indicators.
3. Patient Segmentation (Clustering)
   Groups patients into clinically interpretable cohorts using K-Means clustering.
4. Medical Associations (Association Rules)
   Discovers co-occurrence patterns among diseases and clinical conditions.

Deep Learning Modules

5. Imaging Diagnosis (CNN)
   Analyzes chest images to support disease detection.
6. Sequence Modeling (LSTM)
   Models time-series vital signs to detect early signs of patient deterioration.

7. Sentiment Analysis (Deep Learning NLP)
   Classifies patient feedback as positive or negative to support quality assurance.

---

5. Data Strategy & Ethics

- No real patient-identifiable data was used.
- Data sources include:
    - Synthetic tabular and time-series data
    - Public medical image datasets
    - Publicly available patient feedback text
- All datasets are anonymized, simulated, or open, complying with ethical AI and healthcare data guidelines.

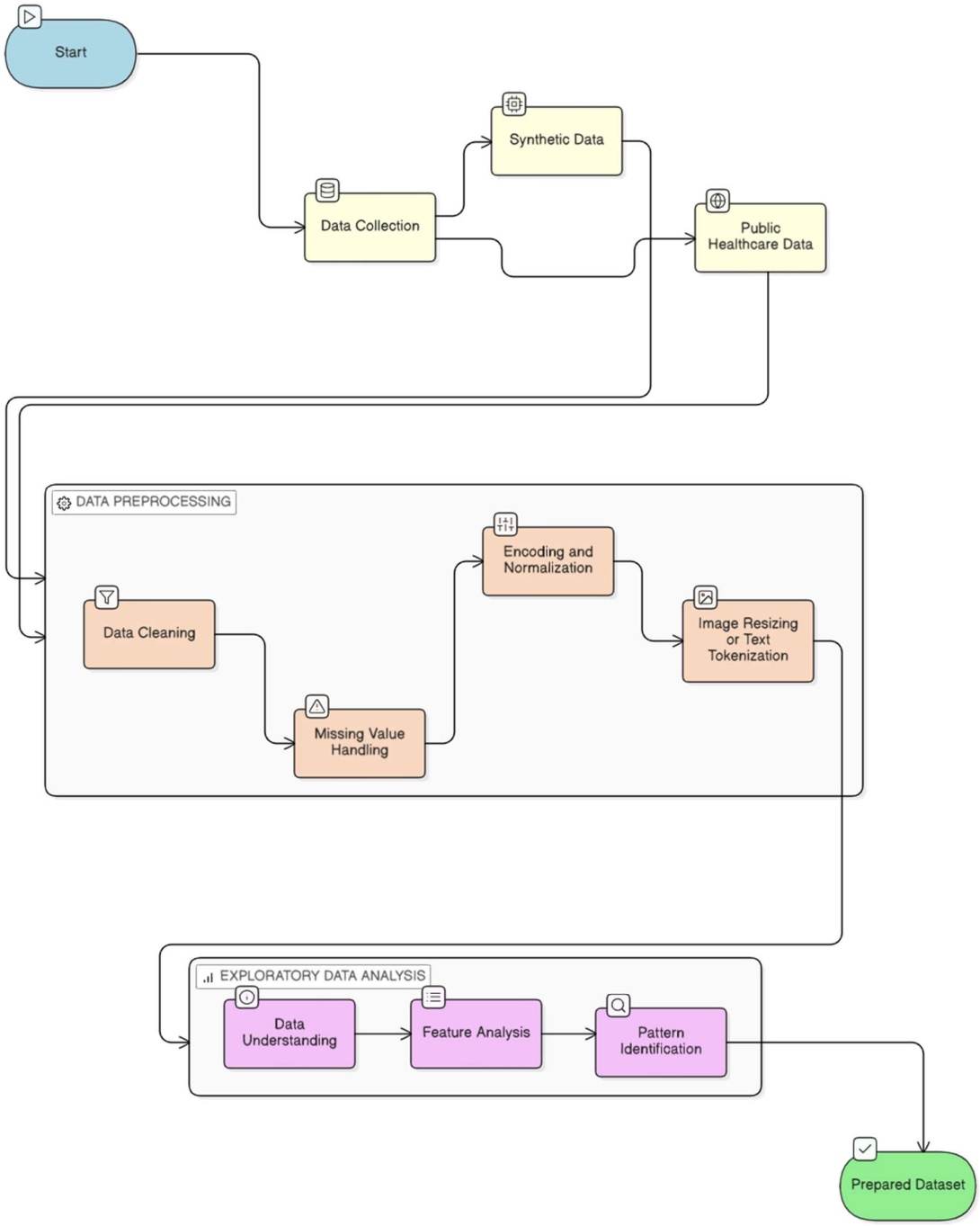
The project explicitly prioritizes methodology and system design over data volume, reflecting real-world healthcare AI constraints.

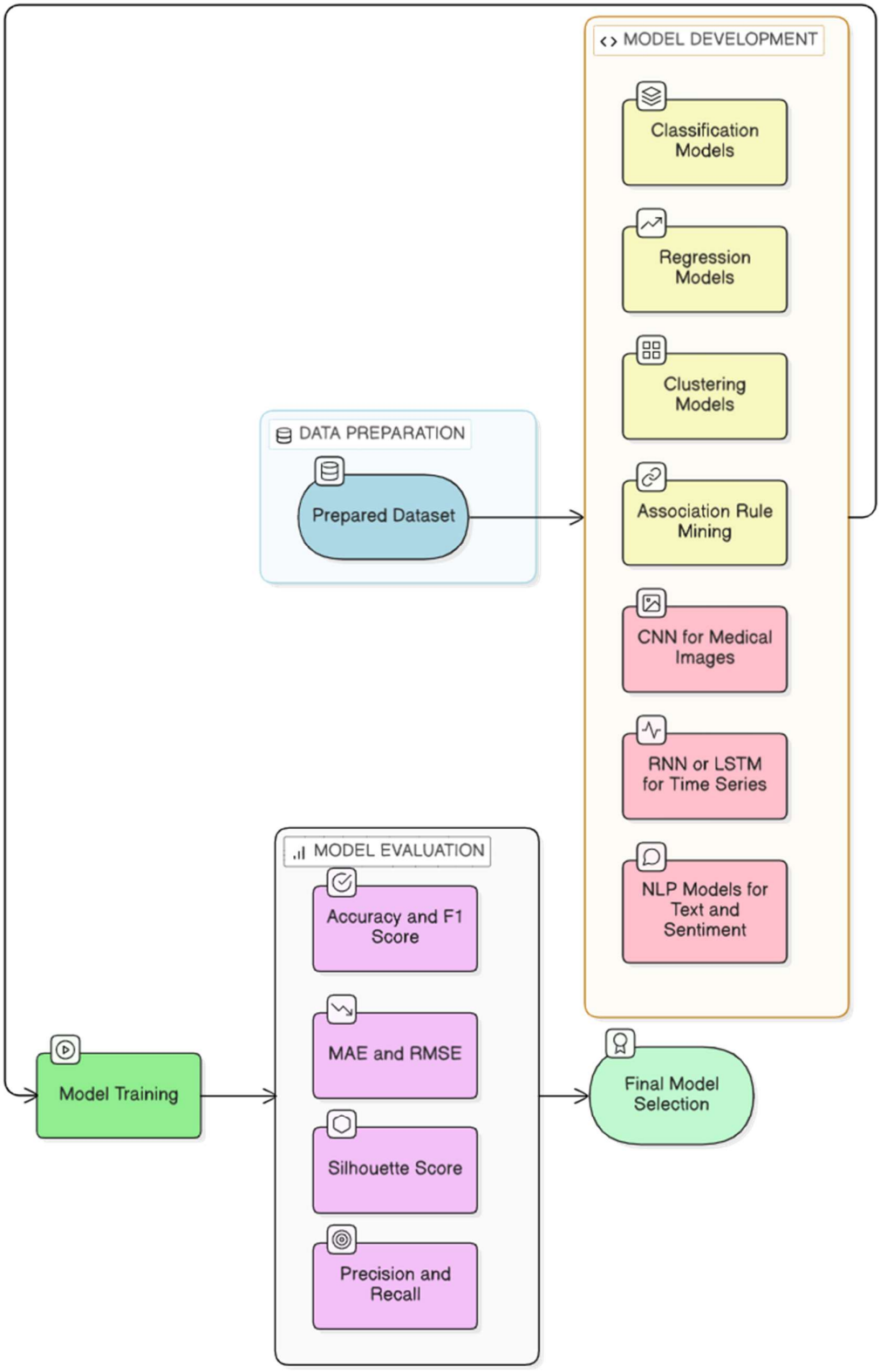
---

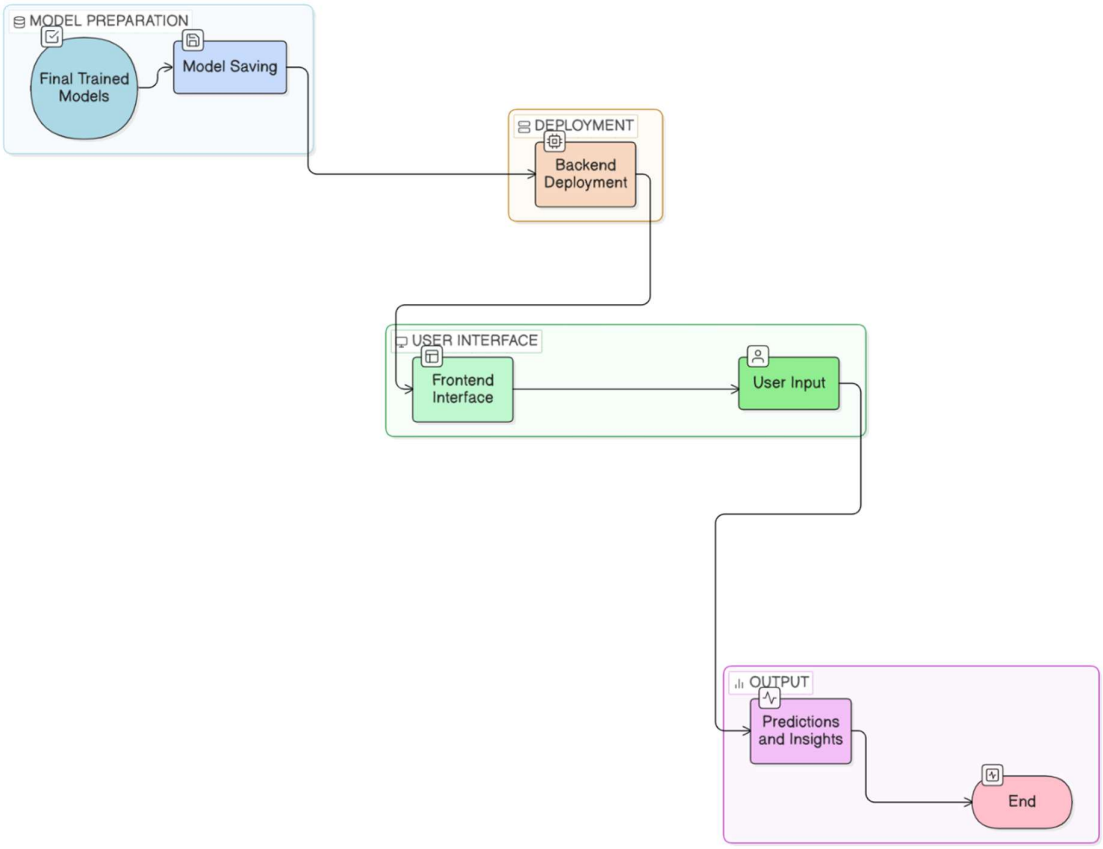
6. Deployment & Integration

- Fast API is used to expose trained models as RESTful endpoints.
- Streamlit provides an interactive dashboard for clinicians and users.
- Preprocessing at inference time strictly mirrors training pipelines, ensuring training–inference parity.
- The system is designed to be containerizable and extensible for future MLOps integration.

Project Workflow




MODEL DEVELOPMENT
Classification Models
Regression Models
Clustering Models
Association Rule Mining
CNN for Medical Images
RNN or LSTM for Time Series
NLP Models for Text and Sentiment
DATA PREPARATION
Prepared Dataset
MODEL EVALUATION
Accuracy and F1 Score
MAE and RMSE
Silhouette Score
Precision and Recall
Model Training
Final Model Selection


MODEL PREPARATION
Final Trained Models
Model Saving
DEPLOYMENT
Backend Deployment
USER INTERFACE
Frontend Interface
User Input
OUTPUT
Predictions and Insights
End

Source Codes

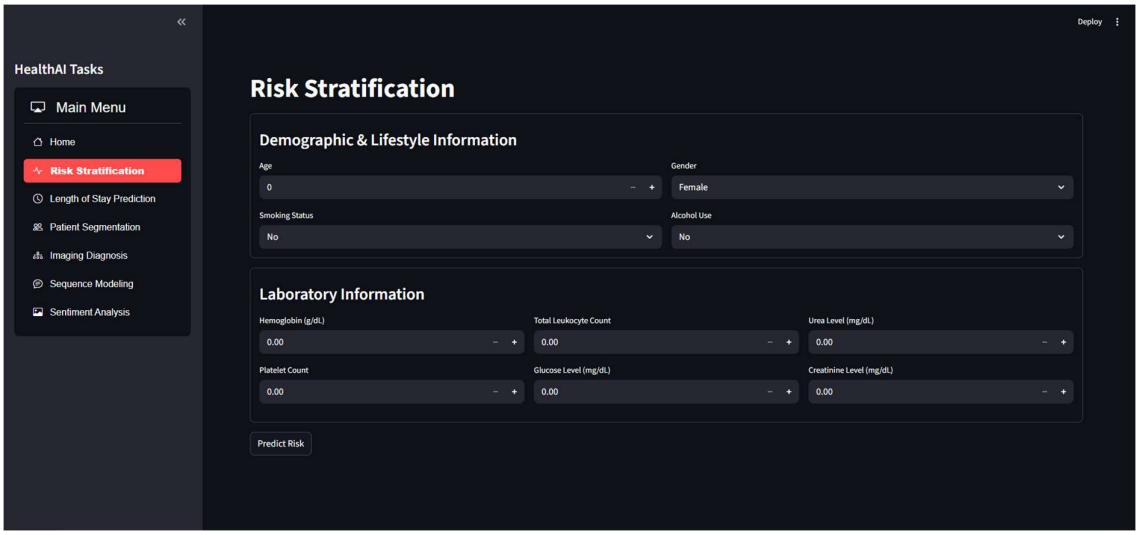
Data Preprocessing and Model Development

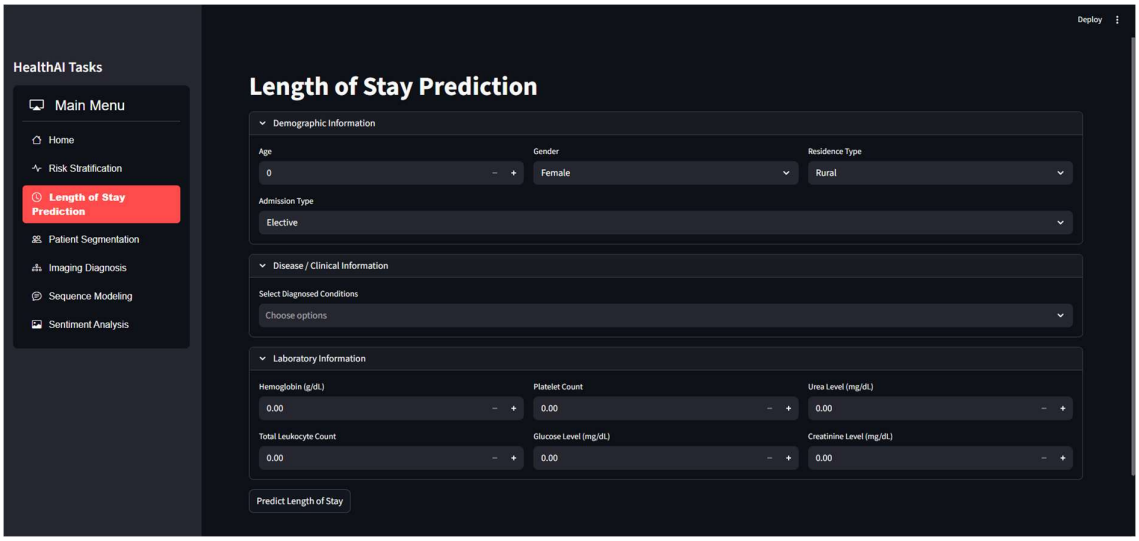
Module	Notebook (HTML)
Data Cleaning	data_cleaning.html
Risk Stratification	risk_stratification.html
Length of Stay Prediction	length_of_stay.html
Patient Segmentation	patient_segmentation.html
Medical Associations	medical_associations.html
Imaging Diagnosis (CNN)	imaging_diagnosis.html
Sequence Modeling	sequence_modeling.html
Sentiment Analysis	sentiment_analysis.html

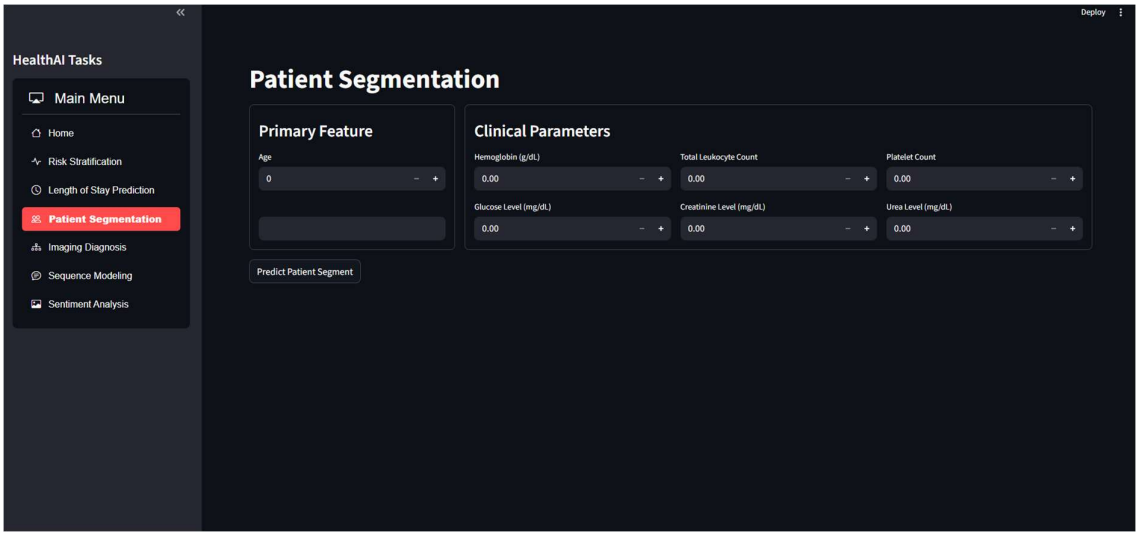
Model deployment

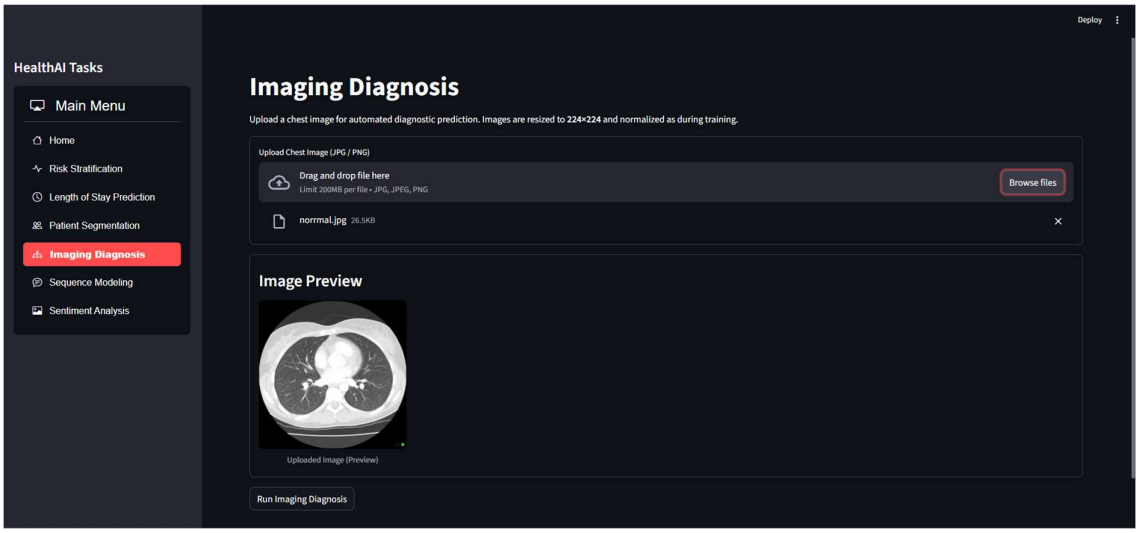
Module	Notebook (HTML)
Streamlit	
app.py	app.html
Fast API	
main.py	main.html
loaders.py	loaders.html
Infrence.py	inference.html

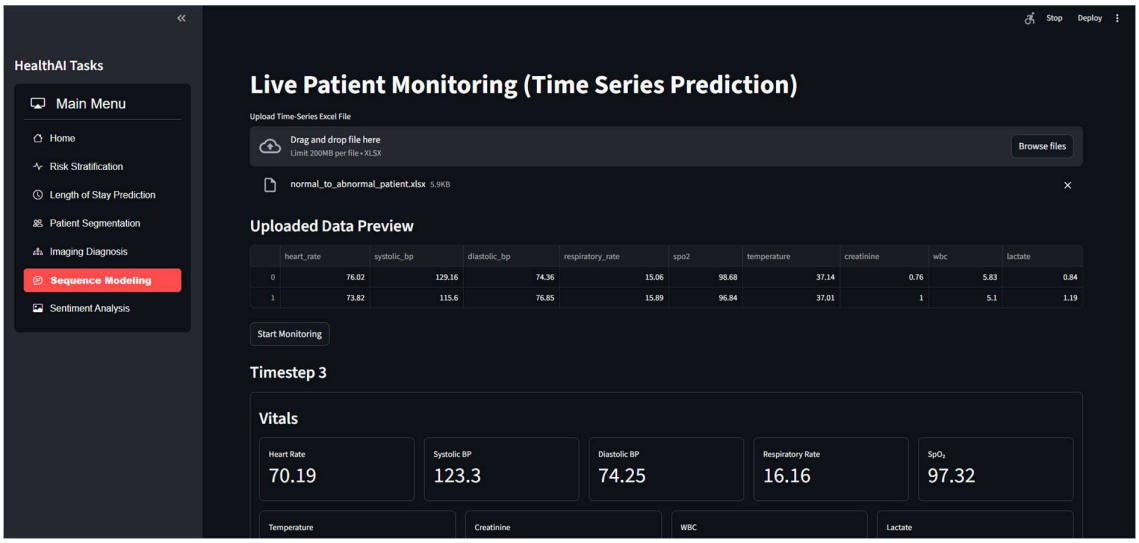
Screenshorts

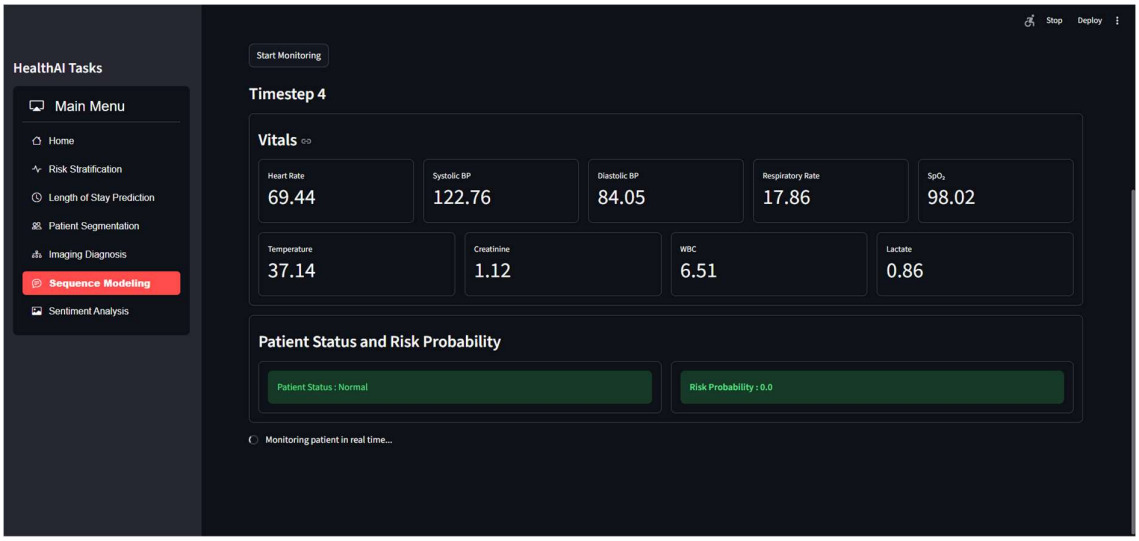


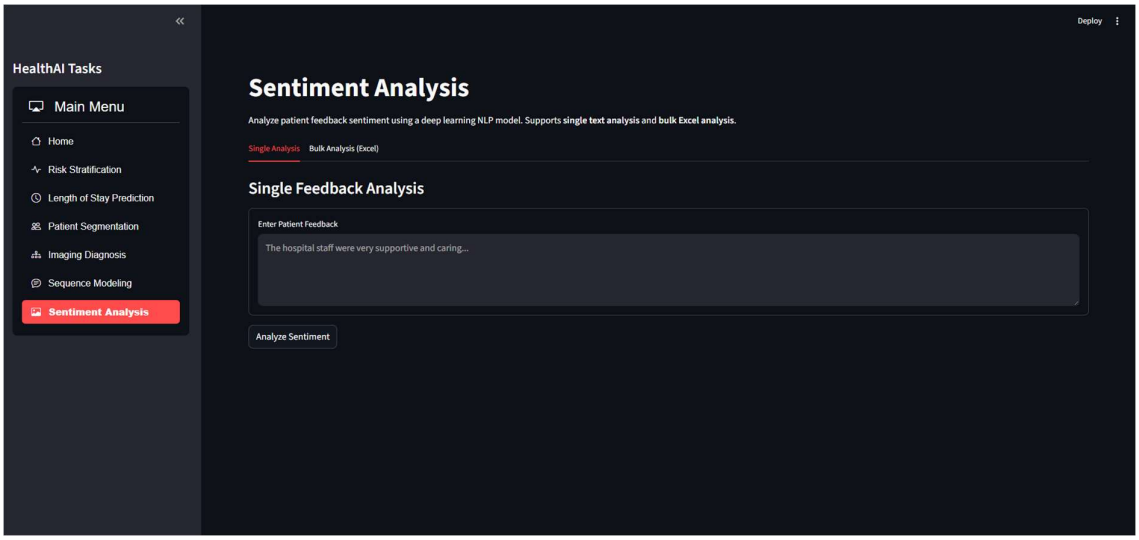



HealthAI Tasks
Main Menu
Home
Risk Stratification
Length of Stay Prediction
Patient Segmentation
Imaging Diagnosis
Sequence Modeling
Sentiment Analysis
Deploy
Patient Segmentation
Primary Feature
Age
0
Clinical Parameters
Hemoglobin (g/dL)
0.00
Total Leukocyte Count
0.00
Platelet Count
0.00
Glucose Level (mg/dL)
0.00
Creatinine Level (mg/dL)
0.00
Urea Level (mg/dL)
0.00
Predict Patient Segment


HealthAI Tasks
Main Menu
Home
Risk Stratification
Length of Stay Prediction
Patient Segmentation
Imaging Diagnosis
Sequence Modeling
Sentiment Analysis
Deploy
Imaging Diagnosis
Upload a chest image for automated diagnostic prediction. Images are resized to 224×224 and normalized as during training.
Upload Chest Image (JPG / PNG)
Drag and drop file here
Limit 200MB per file • JPG, JPEG, PNG
Browse files
norrmal.jpg 26.5KB
Image Preview
Uploaded Image (Preview)
Run Imaging Diagnosis


HealthAI Tasks
Main Menu
Home
Risk Stratification
Length of Stay Prediction
Patient Segmentation
Imaging Diagnosis
Sequence Modeling
Sentiment Analysis
Stop
Deploy
Live Patient Monitoring (Time Series Prediction)
Upload Time-Series Excel File
Drag and drop file here
Limit 200MB per file • XLSX
Browse files
normal_to_abnormal_patient.xlsx 5.9KB
Uploaded Data Preview
heart_rate systolic_bp diastolic_bp respiratory_rate spo2 temperature creatinine wbc lactate
0 76.02 129.16 74.36 15.06 98.68 37.14 0.76 5.83 0.84
1 73.82 115.6 76.85 15.89 96.84 37.01 1 5.1 1.19
Start Monitoring
Timestep 3
Vitals
Heart Rate
70.19
Systolic BP
123.3
Diastolic BP
74.25
Respiratory Rate
16.16
SpO2
97.32
Temperature
Creatinine
WBC
Lactate


Stop
Deploy
HealthAI Tasks
Main Menu
Home
Risk Stratification
Length of Stay Prediction
Patient Segmentation
Imaging Diagnosis
Sequence Modeling
Sentiment Analysis
Start Monitoring
Timestep 4
Vitals
Heart Rate
69.44
Systolic BP
122.76
Diastolic BP
84.05
Respiratory Rate
17.86
SpO2
98.02
Temperature
37.14
Creatinine
1.12
WBC
6.51
Lactate
0.86
Patient Status and Risk Probability
Patient Status : Normal
Risk Probability : 0.0
Monitoring patient in real time...


Deploy
HealthAI Tasks
Main Menu
Home
Risk Stratification
Length of Stay Prediction
Patient Segmentation
Imaging Diagnosis
Sequence Modeling
Sentiment Analysis
Sentiment Analysis
Analyze patient feedback sentiment using a deep learning NLP model. Supports single text analysis and bulk Excel analysis.
Single Analysis
Bulk Analysis (Excel)
Single Feedback Analysis
Enter Patient Feedback
The hospital staff were very supportive and caring...
Analyze Sentiment

Dataset and Data Sources

Datasets Used in the Project

Machine Learning (ML) Tasks

Tasks Covered:

- Risk Stratification
- Length of Stay Prediction
- Patient Segmentation
- Medical Associations

Dataset Name: Hospital Admissions Data
Platform: Kaggle
Dataset Link: https://www.kaggle.com/datasets/ashishsahani/hospital-admissions-data

---

Deep Learning (DL) Tasks

Task Covered:

- Imaging Diagnostics

Dataset Name: Chest CT-Scan Images Dataset
Platform: Kaggle
Dataset Link: https://www.kaggle.com/datasets/mohamedhanyyy/chest-ctscan-images

---

Sequence Modeling

Dataset Description:

- Synthetically generated time-series dataset created for experimental and modeling purposes.

Sentiment Analysis

Dataset Name: Hospital Reviews Dataset
Platform: Kaggle
Dataset Link: https://www.kaggle.com/datasets/junaid6731/hospital-reviews-dataset

Conclusion

The **HealthAI Prediction Suite** successfully demonstrates the application of machine learning and deep learning techniques in healthcare data analysis. By integrating multiple predictive modules such as risk stratification, length of stay prediction, patient segmentation, imaging diagnosis, time-series monitoring, and sentiment analysis, the system provides a unified platform for healthcare analytics.

The project reduces manual effort, improves analytical accuracy, and offers an interactive interface for data-driven decision support. Overall, the HealthAI system highlights the potential of artificial intelligence in enhancing healthcare insights and serves as a strong foundation for future improvements and real-world applications.