

HealthAI – Final Report

Generated: 2025-10-22 23:33 UTC

1. Project Overview

End-to-end synthetic healthcare ML system: tabular risk/LOS models, CNN imaging, LSTM forecasting

2. Data & Preprocessing

- Tabular: data/raw/patients.csv → train/valid/test splits, preprocessing pipeline (impute + scale + one-hot)
- Images: data/raw/images (synthetic normal vs pneumonia).
- Timeseries: data/raw/vitals_timeseries.csv (HR, SBP, SpO2).
- Text: patient_feedback.csv and doctor_notes.csv.

3. Models & Metrics

Tabular Classification (Readmission)

- Accuracy: 0.9373333333333334
- F1: 0.0
- ROC AUC: 0.6868141318930819

Tabular Regression (Length of Stay)

- RMSE: 1.1144895807049833
- R²: 0.43019144663930864

CNN Pneumonia: see reports/model_cards/cnn_pneumonia.md

LSTM HR Forecasting: see reports/model_cards/timeseries_lstm.md

NLP Sentiment: see reports/model_cards/nlp_sentiment.md

4. Interpretability & Ethics

SHAP and LIME explanations for tabular models.

Ethical considerations: synthetic-only data, non-clinical use, potential biases from generation rules, transparency

5. Deployment & Usage

- API: uvicorn apps.api.main:app --host 0.0.0.0 --port 8000
- UI: streamlit run apps/ui/app.py --server.port 8501
- Docker: docker-compose up --build

6. Reproducibility

- Experiment tracking: MLflow stored under mlruns/
- Artifacts: artifacts/
- Data generators: src/data/
- Training: scripts/train_all.sh
- Docs: reports/model_cards/, reports/final/FINAL_REPORT.md

