

Project Title	HealthAI Suite — Intelligent Analytics for Patient Care
Skills take away From This Project	<p>Data cleaning & preprocessing</p> <p>Feature engineering</p> <p>Exploratory Data Analysis (EDA)</p> <p>Classification modeling</p> <p>Regression modeling</p> <p>Clustering</p> <p>Association rule mining</p> <p>Model evaluation metrics</p> <p>Neural networks (MLP)</p> <p>Convolutional Neural Networks (CNN)</p> <p>Recurrent Neural Networks (RNN)</p> <p>Long Short-Term Memory (LSTM)</p> <p>Transfer learning (BioBERT/ClinicalBERT)</p> <p>Text preprocessing</p> <p>Sentiment analysis</p> <p>Healthcare chatbot development</p>

	Machine translation Version control (Git) Experiment tracking (MLflow/W&B) Model deployment (FastAPI) Dashboard creation (Streamlit) Containerization (Docker) Healthcare data domain knowledge Model interpretability (SHAP/LIME) Ethical AI in healthcare
Domain	Healthcare

Problem Statement:

Design an end-to-end AI/ML system that analyzes patient health data (EHR, diagnostics, medical text, and patient feedback) to:

1. Predict outcomes (regression).
2. Classify disease risk categories.
3. Discover patient subgroups (clustering).
4. Mine medical associations (associative learning).
5. Build and compare deep learning models (NN, CNN, RNN, LSTM).
6. Leverage pretrained models (e.g., BioBERT, ClinicalBERT).
7. Develop a healthcare chatbot for patient queries.

8. Build a translator for multilingual medical communication.
9. Perform sentiment analysis on patient feedback.

The aim is to demonstrate how multiple AI paradigms improve **clinical decision support, patient engagement, and operational efficiency**.

Business Use Cases:

Risk Stratification (Classification): Early detection of diseases like diabetes, heart disease, cancer staging.

Length of Stay Prediction (Regression): Forecast patient hospitalization duration for resource planning.

Patient Segmentation (Clustering): Group patients into cohorts (e.g., chronic vs. acute, lifestyle-driven vs. genetic-risk).

Medical Associations (Association Rules): Discover patterns like “(high BMI \wedge hypertension) \Rightarrow increased risk of diabetes”.

Imaging Diagnostics (CNN): Automate radiology analysis for X-rays, CT, or MRI (e.g., pneumonia detection).

Sequence Modeling (RNN/LSTM): Track patient vitals over time to forecast deterioration or readmission.

Pretrained Models: Use **BioBERT/ClinicalBERT** for clinical notes, discharge summaries, drug side effects.

Healthcare Chatbot: Patient triage bot for symptoms, appointment scheduling, and FAQs.

Translator: Bridge doctor–patient communication in regional languages.

Sentiment Analysis: Capture patient experience from feedback/reviews to improve hospital services.

Approach:

Data Preparation: Clean patient data, handle missing values, normalize vitals, tokenize notes.

EDA & Feature Engineering: Clinical indicators (BMI, BP, cholesterol, blood sugar, history of medication).

Modeling by Module:

- Classification → logistic regression, XGBoost, NN.
- Regression → LOS prediction via linear models & LSTM.
- Clustering → k-means/HDBSCAN on patient features.
- Association → Apriori for comorbidities.
- Imaging → CNN on chest X-rays.
- Time Series → LSTM on vitals.
- NLP → BioBERT for diagnosis notes.
- Translator → MarianMT for English ↔ regional.

- Sentiment → Finetuned BERT on patient reviews.
- Chatbot → RAG pipeline with FAQs, guidelines, medical corpus.

Evaluation: Cross-validation, ROC/AUC, BLEU for translation, human evaluation for chatbot.

Integration: Build dashboard (Streamlit) + API (FastAPI).

Results:

Accurate disease classification (>80% F1).

Regression: MAE < baseline for hospital stay predictions.

Meaningful patient clusters (distinct profiles).

Association rules interpretable by clinicians.

CNN detects pathologies at \geq human-level baseline accuracy.

RNN/LSTM captures patient deterioration patterns.

Pretrained BioBERT improves clinical NLP tasks.

Chatbot gives relevant answers with <5% error.

Translator achieves BLEU score > baseline; usable for patient-doctor communication.

Sentiment model detects dissatisfaction trends for hospital QA teams.

Project Evaluation metrics:

- **Classification:** Accuracy, F1-score, ROC-AUC.
- **Regression:** RMSE, MAE, R^2 .
- **Clustering:** Silhouette, Calinski-Harabasz, clinical interpretability.
- **Associations:** Support, Confidence, Lift.
- **Imaging:** Accuracy, Precision, Recall, AUC.
- **RNN/LSTM:** Forecast RMSE, early warning detection rate.
- **NLP (BioBERT/Translator):** BLEU, COMET, F1 on NER tasks.
- **Sentiment:** Precision/Recall, MCC.
- **Chatbot:** Relevance score, Faithfulness (citation grounding), Response latency.

Technical Tags:

healthcare, EHR, classification, regression, clustering,
association-rules, cnn, rnn, lstm, bioBERT, clinicalBERT, transformers,
rag, mlops, pytorch, tensorflow, fastapi, streamlit, scikit-learn, spacy,
nltk, huggingface

Data Set:

- **MIMIC-III or MIMIC-IV** (clinical records, vitals, diagnoses).

- **PhysioNet** (time-series vital signs).
- **Chest X-ray 14 / NIH Dataset** (image dataset for CNN).
- **Patient feedback dataset** (e.g., hospital review portals, Kaggle).
- **Synthetic dataset** (if real-world data not available, anonymized).

Format: CSV/Parquet for tabular (EHR), JPG/PNG for images, TXT for notes.

Variables: age, gender, vitals, lab results, diagnoses, medications, procedures, outcomes.

Data Set Explanation:

Clinical Tabular Data: demographics, vitals, labs, LOS.

Imaging Data: labeled chest X-rays (Normal/Pneumonia/etc.).

Text Data: discharge summaries, physician notes, patient reviews.

Preprocessing Steps:

- Missing value imputation.
- One-hot encode categorical (gender, comorbidity).
- Normalize vitals (z-score).
- Tokenize text (BioBERT/ClinicalBERT tokenizer).
- Resize/augment medical images.

- Split train/val/test chronologically.

Project Deliverables:

Source code (organized repo with modules).

Data preprocessing scripts.

Model notebooks (EDA, ML, DL, NLP).

Trained model artifacts.

API endpoints (FastAPI).

Dashboard (Streamlit) with patient risk predictions.

Documentation (README + model cards).

Final project report + presentation slides.

Demo video (showing chatbot, translator, sentiment insights).

Project Guidelines:

- **Version Control:** Git + branching workflow.
- **Reproducibility:** Seeds fixed, config-driven experiments.
- **Data Security:** Anonymize PII, follow HIPAA/GDPR guidelines.
- **Experiment Tracking:** MLflow or Weights & Biases.
- **Coding Standards:** PEP8, unit tests, modular design.
- **Deployment:** Containerize models, provide REST APIs.

- **Ethical AI:** Ensure fairness, transparency, interpretability (e.g., SHAP for predictions).
- **Documentation:** Provide pipeline diagrams, model cards, and user guide.

Timeline:

10 days

PROJECT DOUBT CLARIFICATION SESSION (PROJECT AND CLASS DOUBTS)

About Session: The Project Doubt Clarification Session is a helpful resource for resolving questions and concerns about projects and class topics. It provides support in understanding project requirements, addressing code issues, and clarifying class concepts. The session aims to enhance comprehension and provide guidance to overcome challenges effectively.

Note: Book the slot at least before 12:00 Pm on the same day

Timing: Monday-Saturday (4:00PM to 5:00PM)

Booking link : <https://forms.gle/XC553oSbMJ2Gcfug9>

For DE/BADM project/class topic doubt slot clarification session:

Booking link : <https://forms.gle/NtkO4UV9cBV7Ac3C8>

Session timing:

For DE: 04:00 pm to 5:00 pm every saturday

For BADM 05:00 to 07:00 pm every saturday

LIVE EVALUATION SESSION (CAPSTONE AND FINAL PROJECT)

About Session: The Live Evaluation Session for Capstone and Final Projects allows participants to showcase their projects and receive real-time feedback for improvement. It assesses project quality and provides an opportunity for discussion and evaluation.

Note: This form will Open only on Saturday (after 2 PM) and Sunday on Every Week

Timing:

For BADM and DE

Monday-Saturday (11:30AM to 1:00PM)

For DS and AIML

Monday-Saturday (05:30PM to 07:00PM)

Booking link : <https://forms.gle/1m2Gsro41fLtZurRA>

