

PROJECT OUTLINE

12-label Multi-Morbidity Prediction System across the three categories (Brain, Cardiovascular, Multi-Morbidity Indicators), ensuring the content is formatted for high academic and professional value.



Project Report: Interpretable Multi-Morbidity Risk Stratification on the Cardio-Neuro Axis

1. Introduction: The Multi-Morbidity Challenge

The project addresses a critical gap in preventative care: the inability of traditional models to assess the simultaneous risk of multiple, interconnected chronic diseases. The **Cardio-Neuro-Metabolic Axis** highlights the strong correlations between vascular health, metabolic disorders, and neuro-cognitive decline. This system moves beyond single-disease prediction to offer a holistic, personalized **12-label risk profile**.

Project Goal: Develop, train, and deploy an Explainable Machine Learning (XAI) system capable of predicting high-risk status (binary 0/1) for 12 chronic diseases across three critical categories within a single model.

2. Top 12 Chronic Targets for Multi-Label Prediction

The project focuses on 12 clinically relevant conditions, ensuring both data availability (via fusion) and high impact on patient outcomes.

Category	High-Risk Prediction Targets (The 12 Labels)	Rationale & Clinical Importance
I. Primary Brain Disease Risk (Neuro-Cognitive)	1. Alzheimer's Disease (AD) (MCI to AD conversion)	Focus on predicting pre-clinical conversion using cognitive and imaging biomarkers.
	2. Vascular Dementia (VaD)	Links cognitive decline directly to vascular and ischemic damage.

Category	High-Risk Prediction Targets (The 12 Labels)	Rationale & Clinical Importance
	3. Parkinson's Disease (PD)	A separate neurological condition with unique feature drivers (e.g., motor scores).
	4. Epilepsy/Seizure Disorder (Recurrence Risk)	Requires high-quality monitoring of medication adherence and event history.
II. Cardiovascular Risk (Vascular System)	5. Ischemic Stroke (Recurrence Risk)	High-risk outcome, the primary acute neurovascular event.
	6. Uncontrolled Hypertension (HTN)	The leading modifiable risk factor for both stroke and heart disease.
	7. Coronary Artery Disease (CAD) / MI Risk	Major systemic vascular endpoint, requires lipid and BP features.
	8. Atrial Fibrillation (AFib) (New Onset)	Key source of embolic stroke risk; often silent in EHRs.
III. Multi- & Bi-Morbidity Indicators (Systemic Comorbidities)	9. Type 2 Diabetes Mellitus (T2DM) (Onset Risk)	Essential metabolic driver of vascular damage and cognitive decline.
	10. Chronic Kidney Disease (CKD) (Stage 3+)	Indicator of widespread systemic damage due to HTN and T2DM.

Category	High-Risk Prediction Targets (The 12 Labels)	Rationale & Clinical Importance
	11. Chronic Major Depressive Disorder (MDD)	High comorbidity with chronic pain and neurodegenerative conditions.
	12. Uncontrolled Hyperlipidemia (High Cholesterol)	A core modifiable risk factor, managed via EHR prescriptions.

3. Project Phases and Execution Flow

The project followed a three-phase structure, enabling concurrent work by the three team members and ensuring a robust, integrated final system.

Phase	Duration	Primary Focus	Key Handoff/Integration Point
Phase 1: Parallel Setup	3 Weeks	Data, Model, and System Interface Definition	Data Expert hands off the Clean Feature Table (the EHR proxy) to the Modeler.
Phase 2: Integration	3 Weeks	Feature Engineering, Model Optimization, and XAI Integration	ML Modeler hands off the Final Production Model to the Project Lead.
Phase 3: Convergence	3 Weeks	Deployment, Final Validation, and Documentation	All roles contribute to the Final Report and Deployed API .

4. Task Division and Role Contributions

The project was executed by three specialized members, with tasks divided to achieve a highly balanced workload (\$\approx 33\%\$ each) and maximize specialized output.

4.1 Data & Domain Expert (The Data Architect)

- **Phase 1 Focus:** Data Fusion (ADNI + Synthetic features) and the complex generation of the **12 Binary Target Labels**.
- **Phase 2 Focus:** Creation of **Advanced Longitudinal Features** (e.g., rate of change/slopes) and the single, unifying **Comorbidity Index Score**.
- **Phase 3 Focus:** Clinical Validation of the model's output and drafting the initial **Report Sections** (Data, Feature Engineering).

4.2 ML Modeler & Analyst (The Algorithm Specialist)

- **Phase 1 Focus:** Setting up the **ML Pipeline** and establishing the **Baseline Model** to validate the data and metrics (Macro F1/AUPRC).
- **Phase 2 Focus:** Implementing and Tuning the final **Deep Neural Network (DNN) Multi-Label Classifier** and exporting the production model file.
- **Phase 3 Focus:** Generating the comprehensive **Performance Visualization** (all charts) and writing the **Results and Methodology** sections of the report.

4.3 Project Lead & Deployer (The System Architect)

- **Phase 1 Focus:** Building the **Dockerized FastAPI Shell** and defining the system's input/output contract (JSON format).
- **Phase 2 Focus:** Integrating the **SHAP library** and implementing the logic to convert numerical SHAP values into a **readable text explanation** for clinical use.
- **Phase 3 Focus:** Final System Deployment, creating the **Showcase Interface** (Jupyter/Web Demo), and documenting the **MLOps and System Architecture** sections.

5. Conclusion: High-Value Deliverables

This project successfully transitioned from a complex clinical problem to a robust, technically advanced solution. The core value propositions for the resume and publication are:

1. **Technical Complexity:** Successful application of **Deep Multi-Label Classification (MLC)** to predict 12 correlated clinical endpoints simultaneously, showcasing advanced ML proficiency.

2. **Data Proficiency:** Mastery of **Data Fusion** and **Longitudinal Feature Engineering** on complex EHR-like data (ADNI), a critical skill in health tech.
3. **Clinical Actionability:** The integration of **SHAP Explanations** provides the necessary transparency for clinical decision support, making the model a highly practical and trustworthy tool.
4. **MLOps Maturity:** Delivery of the entire system as a **Dockerized API**, demonstrating readiness for production environments.

This system provides a framework for future research into multi-morbidity management and preventative intervention.