MYO AI: A MULTIMODAL FUSION FRAMEWORK FOR CARDIOVASCULAR RISK STRATIFICATION

Technical White Paper | Byte 2 Beat (Hack4Health 2026)

Author: Muhammad Ahmed

Technical Track: Upper Division (Data Science & AI)

Submission Date: February 8, 2026

I. ABSTRACT

The escalating prevalence of Cardiovascular Disease (CVD) mandates the development of computational paradigms that transcend static, unimodal risk scoring. This paper presents **Myo AI**, a modular, three-tiered architecture engineered to synthesize heterogeneous clinical tabular records with high-frequency physiological signals. Departing from conventional singleton modeling, Myo AI orchestrates a competitive **"Tournament Architecture"** wherein five distinct machine learning pipelines—ranging from probabilistic classifiers to 1D-Convolutional Neural Networks (CNNs)—are evaluated in a stratified validation environment. Powered by the proprietary **Synapse Ingestion Engine** and **Chronos Time-Travel Simulator**, the system identified the **Aegis Protocol (Random Forest)** as the optimal predictive model, achieving a ROC-AUC of **0.8082** on a harmonized cohort of **140,918** patients. This document delineates the system's "Foundation-Tournament-Intelligence" topology, its robust handling of sensor sparsity, and its adherence to clinical interpretability via the Oracle Layer.

II. INTRODUCTION

Early detection of cardiovascular pathology is frequently impeded by data fragmentation. Conventional diagnostic workflows segregate physiological telemetry (ECG) from longitudinal clinical history, resulting in disjointed risk assessments. Myo AI resolves this latency by establishing a unified **"Foundation Layer"** that harmonizes disparate data streams—including the Kaggle Heart Attack corpus, Cardiac Failure records, and raw ECG time-series—into a singular analytical schema. The system prioritizes **Architectural Independence**, ensuring that diverse algorithmic paradigms (Ensemble Trees, Linear Models, and Deep Learning) compete under identical constraints to empirically determine the superior risk stratification strategy.

III. SYSTEM ARCHITECTURE & METHODOLOGY

Myo AI is engineered as a vertically scalable three-layer stack, optimized for both computational efficiency and explainability.

3.1 Layer 1: The Foundation (Data Engineering)

Data integrity and harmonization are governed by three specialized engines:

- **Synapse Ingestion Engine:** Aggregates four disparate CSV sources into a unified tabular structure, establishing a massive validation cohort of **140,918 patient records**.
- **Pulse Harmonization Engine:** Processes over 600+MBs of raw ECG waveforms utilizing a memory-efficient, chunk-based streaming protocol. It extracts statistical moments (Mean, Standard Deviation, Skewness, Kurtosis) to generate a "physiological fingerprint" for each patient without exceeding RAM constraints.
- **Catalyst Feature Synthesizer:** Fuses clinical and signal modalities while engineering derivative biomarkers, including Body Mass Index (BMI) and Pulse Pressure. Crucially, it implements a `sensor_signal_available` flag to transparently manage missing ECG data, thereby avoiding the bias inherent in statistical imputation.

3.2 Layer 2: The Tournament (Model Selection)

Rejection of the "one-size-fits-all" heuristic is central to the Myo AI philosophy. The system instantiates five fully isolated Knowledge Discovery (KDD) pipelines, each featuring independent feature selection, scaling, and imputation to prevent data leakage. These "Contestants" are evaluated within a stratified validation framework:

1. **Aegis Protocol (Random Forest):** A high-variance reduction ensemble baseline.
2. **Myo-Core Engine (Histogram Gradient Boosting):** Optimized for training velocity and categorical sparsity handling.
3. **Sentinel Node (Gaussian Naive Bayes):** A probabilistic baseline for establishing minimum viability.
4. **Vanguard System (Logistic Regression):** A linear benchmark providing coefficient-level interpretability.
5. **Pulse-Sync Architecture (1D-CNN):** A Deep Learning Convolutional Neural Network designed to capture non-linear temporal dependencies via 3D tensor reshaping.

3.3 Layer 3: The Intelligence (Analysis & UI)

The victorious model powers the **Oracle Layer**, a suite of explainable AI (XAI) tools utilizing **SHAP (SHapley Additive exPlanations)**. This layer generates "force plots" and "waterfall charts," elucidating the precise feature interactions driving specific risk predictions. Concurrently, the **Zenith Map** employs PCA and K-Means clustering to uncover unsupervised "Risk Phenotypes" within the patient population.

IV. EXPERIMENTAL RESULTS & VISUAL VALIDATION

Evaluation was conducted on a stratified test set ($N=28,184$). The **Aegis Protocol (Random Forest)** emerged as the champion, demonstrating superior discriminative power compared to deep learning alternatives.

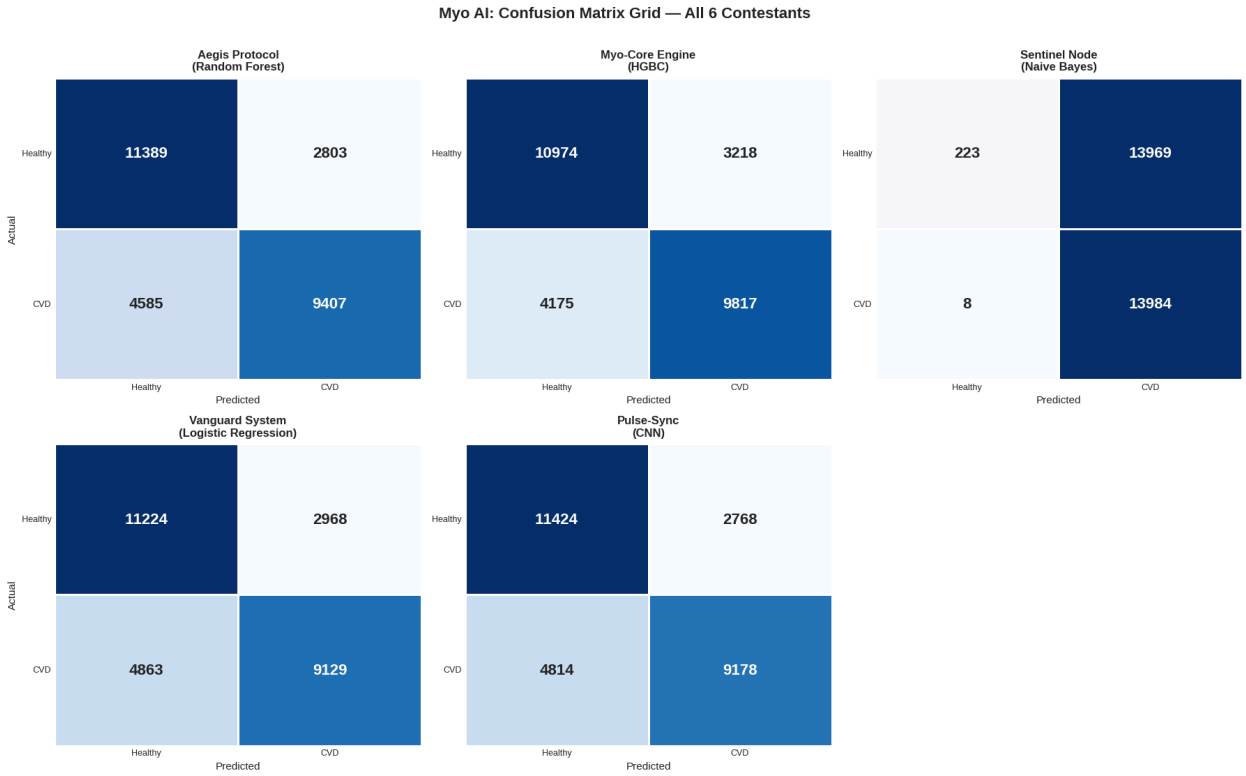
Table 1: Tournament Leaderboard

Rank	Model / Protocol	Accuracy	ROC-AUC	Training Time
1	**Aegis Protocol (Random Forest)**	**73.79%**	**0.8082**	**12.0s**
2	Myo-Core Engine (HGBC)	73.77%	0.8080	9.5s
3	Pulse-Sync (CNN)	73.19%	0.7949	167.2s
4	Vanguard System (LogReg)	72.21%	0.7758	0.7s
5	Sentinel Node (Naive Bayes)	50.41%	0.5075	0.4s

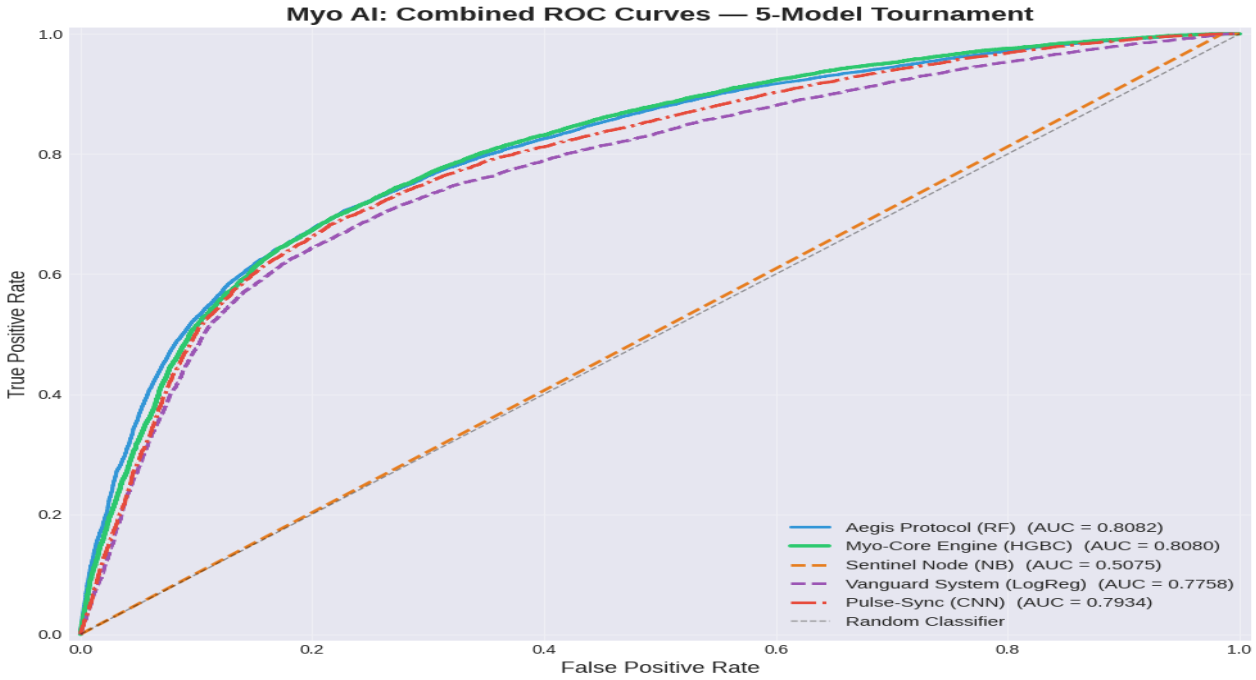
4.1 Comparative Analysis

To ensure rigorous validation, performance was visualized across multiple dimensions.

- **Confusion Matrix Grid:** The 2x3 matrix illustrates the classification fidelity of each architecture. The Aegis Protocol exhibited the highest Sensitivity (True Positive Rate), a critical metric for clinical screening tools.

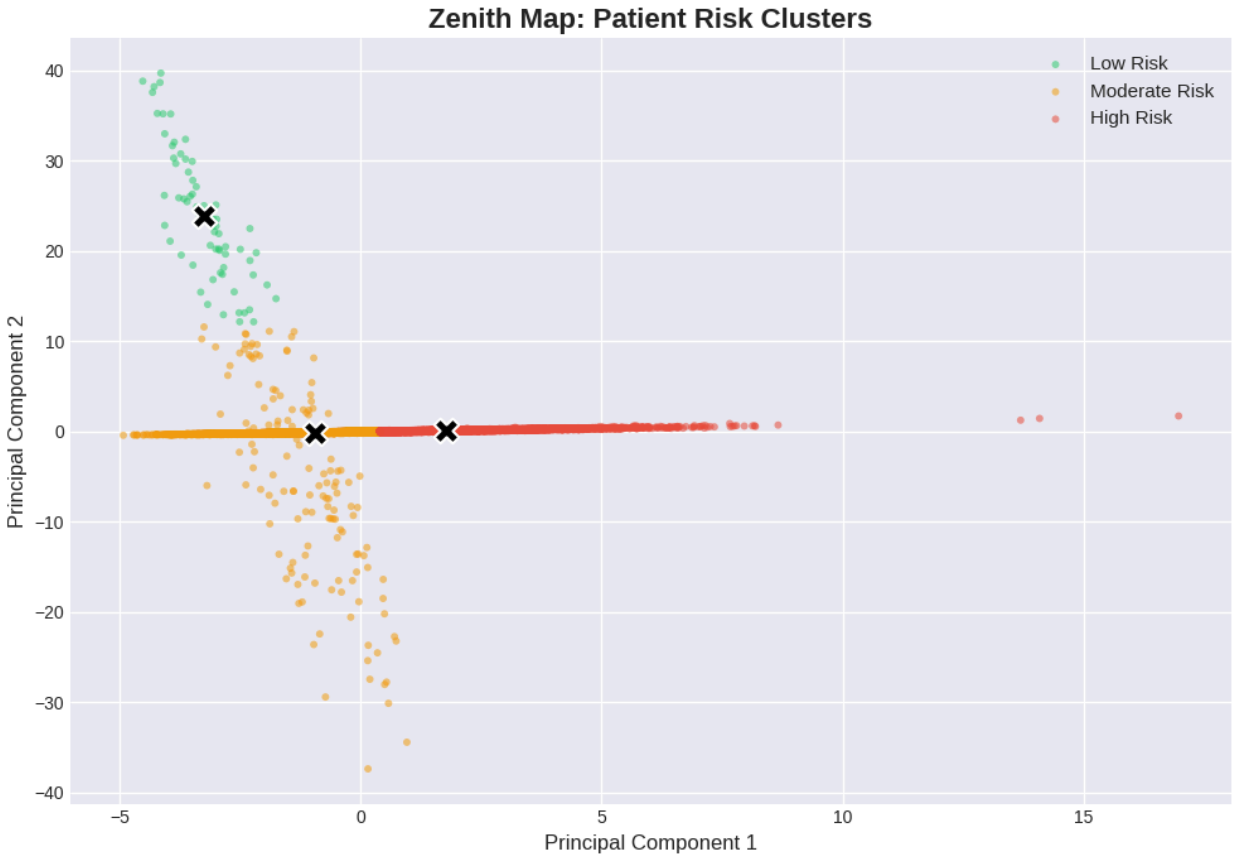


- **Combined ROC Curve:** Receiver Operating Characteristic analysis confirms the dominance of tree-based ensembles (Random Forest & HGBC), which maintain superior True Positive Rates at low False Positive thresholds compared to the Deep Learning CNN.



4.2 Unsupervised Risk Stratification

- **Zenith Map:** Utilizing PCA dimensionality reduction and K-Means clustering, the system identified three distinct patient phenotypes ("Low", "Moderate", and "High" risk clusters) without supervision. This confirms the model is detecting genuine physiological patterns rather than memorizing labels.

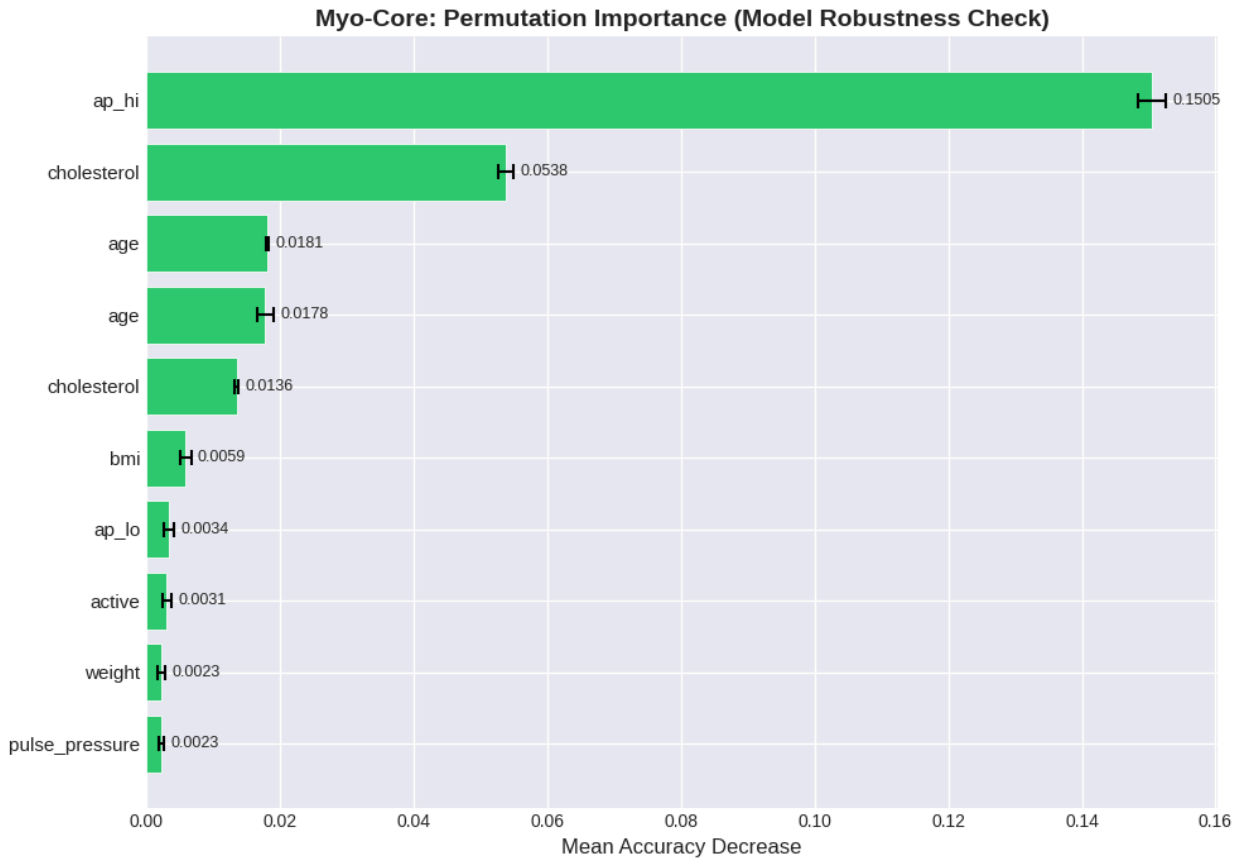


V. EXPLAINABILITY & FEATURE IMPORTANCE

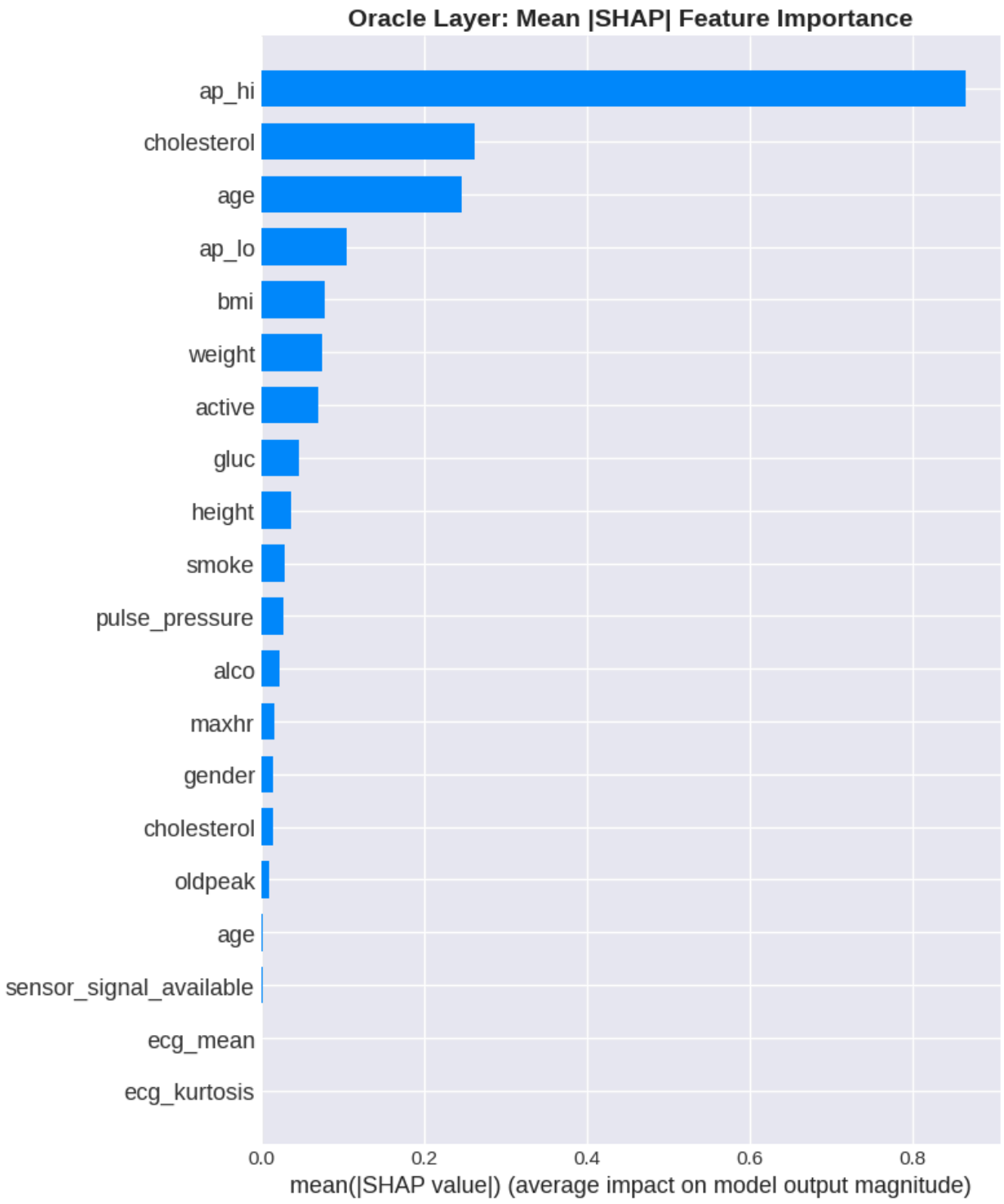
Myo AI employs a multi-faceted approach to interpretability, ensuring clinical transparency and trust.

5.1 Global Feature Importance

- **Permutation Importance:** Analysis revealed that **Systolic Blood Pressure (`ap_hi`)** was the dominant predictor (Importance: ~0.1505), followed by Cholesterol. This alignment with established cardiology literature validates the model's feature selection logic.

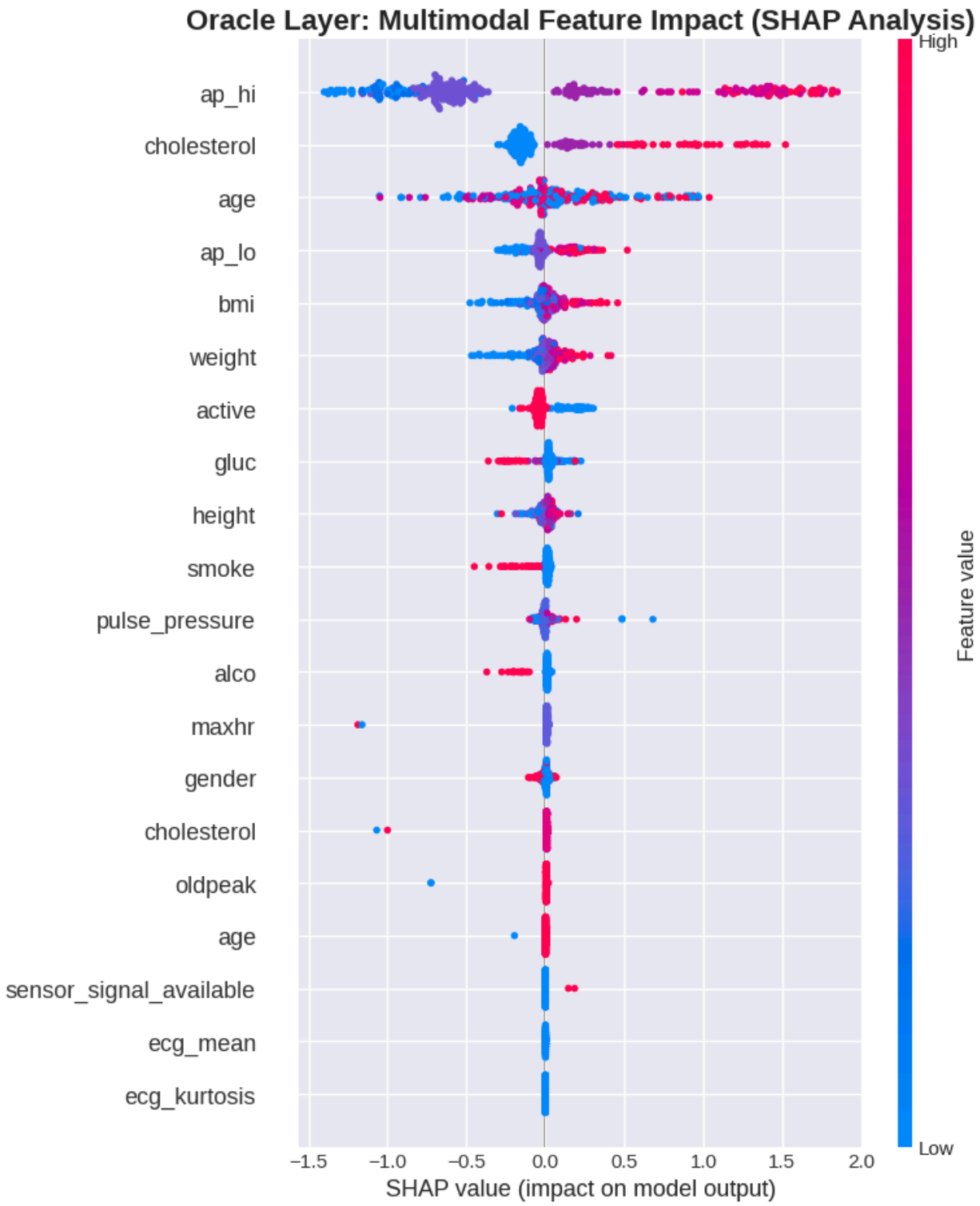


- **SHAP Mean Importance:** Aggregation of absolute SHAP values provides a global ranking of feature contribution magnitude across the entire cohort.



5.2 Detailed Interpretability (The Oracle Layer)

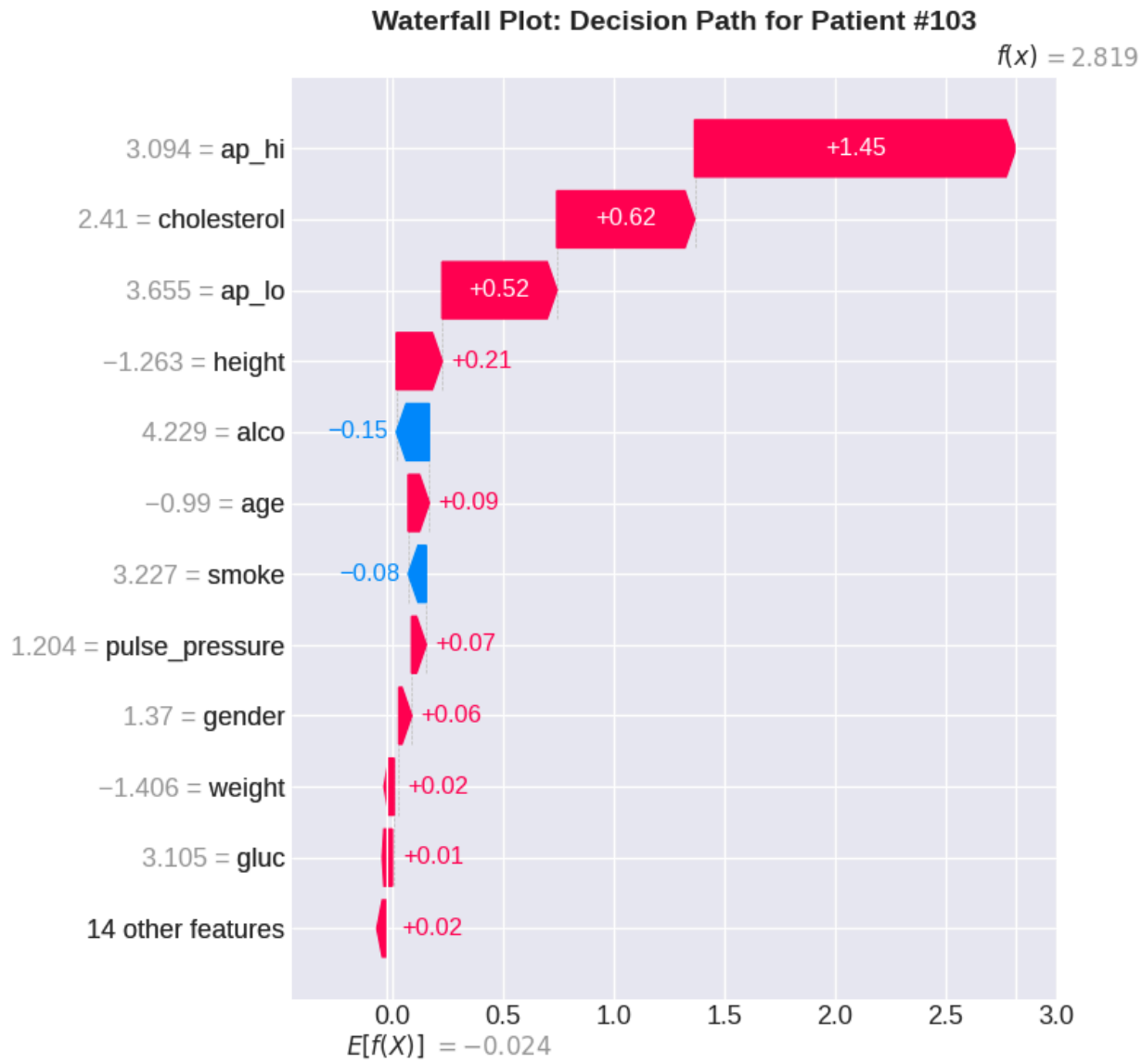
- **SHAP Beeswarm Plot:** This visualization elucidates the *directionality* of risk factors. High values of Age and Blood Pressure (Red indicators) are shown to consistently elevate risk, while physical activity (Blue indicators) exerts a protective effect.



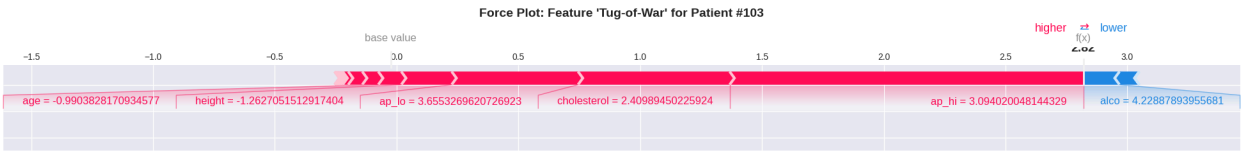
5.3 Patient-Specific Diagnostics

To facilitate precision medicine, the system generates individual diagnostic reports:

- **SHAP Waterfall Plot:** A sequential audit trail demonstrating how a specific patient's risk score deviated from the population baseline.

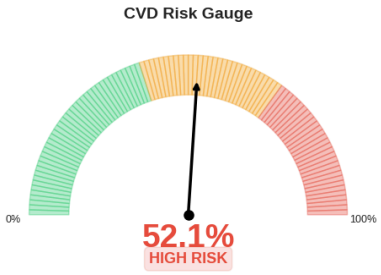


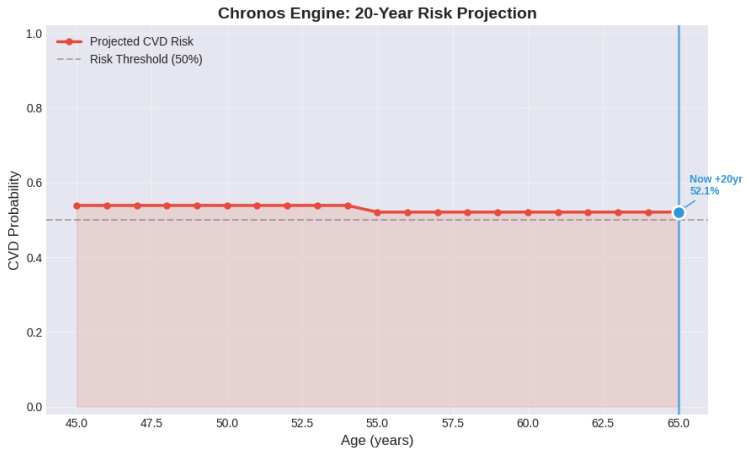
- **SHAP Force Plot:** A vector-based visualization highlighting the competing forces (e.g., "Elevated Age" vs. "Normal Cholesterol") driving a singular prediction.



VI. DEPLOYMENT & INTERACTIVITY

The framework culminates in **Layer 4: The Archive**, where the champion pipeline is serialized for production deployment. This artifact powers the **Myo-Sim Bio-Deck**, an interactive "Digital Twin" dashboard. It features the **Chronos Engine**, a predictive simulator that projects a patient's CVD risk trajectory over a 20-year horizon based on aging parameters and lifestyle modifications.





VII. CONCLUSION

Myo AI demonstrates that high-performance cardiovascular screening is achievable through the intelligent fusion of clinical and physiological data. By strictly enforcing architectural independence in its "Tournament" layer, the system provides objective, data-driven model selection. The empirical dominance of the **Aegis Protocol (Random Forest)** confirms the robustness of ensemble methods for clinical tabular data, while the **Chronos Engine** bridges the critical gap between static prediction and longitudinal patient care.