HACK4HEALTH:
BYTE2BEAT 2026

HEART GUARD

Prepared By :
Mostafa Abayazead

github.com/MostafaAY43/Byte2Beat

mostafaabayazead@gmail.com
linkedin.com/in/mostafa-abayazead-2bbb1821b/

HeartGuard: A Multi-Modal Deep Learning System for ECG-Based Cardiovascular Risk Prediction and Clinical Intervention Planning

Mostafa Abayazead

Hack4Health: Byte2Beat Hackathon 2026

Abstract

Cardiovascular disease (CVD) remains the leading global cause of mortality, claiming approximately 17.9 million lives annually. Although electrocardiography is common in medical practice, ECG data is rarely used for automatic, personalized risk assessment and treatment planning. I present HeartGuard, a system that processes raw 12-lead ECG signals to predict cardiovascular risk, explain its predictions to clinicians, and generate personalized intervention recommendations. Using the PTB-XL dataset (21,799 ECGs, 18,869 patients), we compare four models: XGBoost, 1D CNN, CNN-LSTM, and a multi-modal fusion architecture. Our best model achieves an AUC of 0.8345 on human-validated test data. The system produces three categories of clinical output: patient-level risk reports with explainable predictions, population-level risk stratification analytics, and prioritized intervention plans with evidence-based risk reduction estimates. HeartGuard demonstrates how ECG-based machine learning can be extended beyond diagnosis into actionable clinical decision support.

1. Introduction

Cardiovascular disease is the leading cause of death worldwide, responsible for an estimated 17.9 million deaths annually. Early identification of at-risk individuals remains a cornerstone of preventive cardiology, yet traditional risk scoring tools such as the Framingham Risk Score rely primarily on demographic and laboratory data, often overlooking the diagnostic information embedded in electrocardiographic recordings. While recent deep learning research has demonstrated strong performance on ECG classification tasks, most systems stop at diagnosis and do not address the downstream clinical question: what should be done for this patient?

HeartGuard addresses this gap by building a complete pipeline from raw ECG signal to clinical action. The system (1) predicts patient-level cardiovascular risk from 12-lead ECG recordings; (2) explains which signal features and ECG regions drive each prediction; (3) generates individualized intervention recommendations with quantified risk reduction estimates; and (4) produces population-level analytics for healthcare resource planning. We evaluate on the PTB-XL dataset, a benchmark comprising 21,799 clinical-grade ECGs with expert annotations, and demonstrate that the system produces clinically coherent outputs across all four stages.

2. Methods

2.1 Dataset and Risk Labeling

We use the PTB-XL dataset: 21,799 clinical 12-lead ECGs (500 Hz, 10 seconds) from 18,869 patients, with expert annotations across 71 SCP-ECG diagnostic statements in five superclasses: Normal (9,514), Myocardial Infarction (5,469), ST/T Change (5,235), Conduction Disturbance (4,898), and Hypertrophy (2,649). We use patient-stratified folds 1-8 for training, fold 9 for validation, and fold 10 (human-validated) for testing. A composite risk score is engineered by mapping diagnostic codes to three severity tiers (high: acute MI, atrial fibrillation, complete heart block; moderate: hypertrophy, bundle branch blocks; low: sinus rate variants), combined with age and sex adjustments, and discretized into four risk classes.

2.2 Feature Extraction and Model Architectures

We extract 165 features per ECG: HRV time-domain metrics (SDNN, RMSSD, pNN50) from R-peak detection in Lead II, per-lead statistical features (mean, std, skewness, kurtosis, percentiles, energy), and morphological features (R-wave amplitudes, QRS duration). Four models are compared: (1) XGBoost on tabular features (200 trees, depth 6); (2) 1D CNN with four convolutional blocks (32/64/128/256 filters) operating on raw signals; (3) CNN-LSTM adding bidirectional LSTMs for

temporal modeling; (4) Multi-Modal Fusion concatenating a CNN branch with a tabular dense branch. All deep learning models use PyTorch with class-weighted cross-entropy loss, Adam optimizer, early stopping, and learning rate scheduling.

3. Results

3.1 Model Performance

Table 1 presents test set performance (fold 10, n=2,198). The 1D CNN achieves the highest AUC (0.8345), while Multi-Modal fusion achieves the highest accuracy (70.3%). All deep learning models outperform the XGBoost baseline in AUC, validating the value of learning directly from raw ECG waveforms. The CNN-LSTM does not improve upon the simpler CNN, suggesting local morphological features are more discriminative than long-range temporal patterns for 10-second risk stratification.

Model	Accuracy	F1 (Macro)	F1 (Weighted)	AUC
XGBoost	0.6715	0.5536	0.6648	0.8030
1D CNN	0.6984	**0.5765***	0.6955	**0.8345***
CNN-LSTM	0.6683	0.5354	0.6807	0.8266
Multi-Modal	**0.7029***	0.5702	**0.7027***	0.8278

*Table 1. Test set performance. Best values marked with * and bolded.*

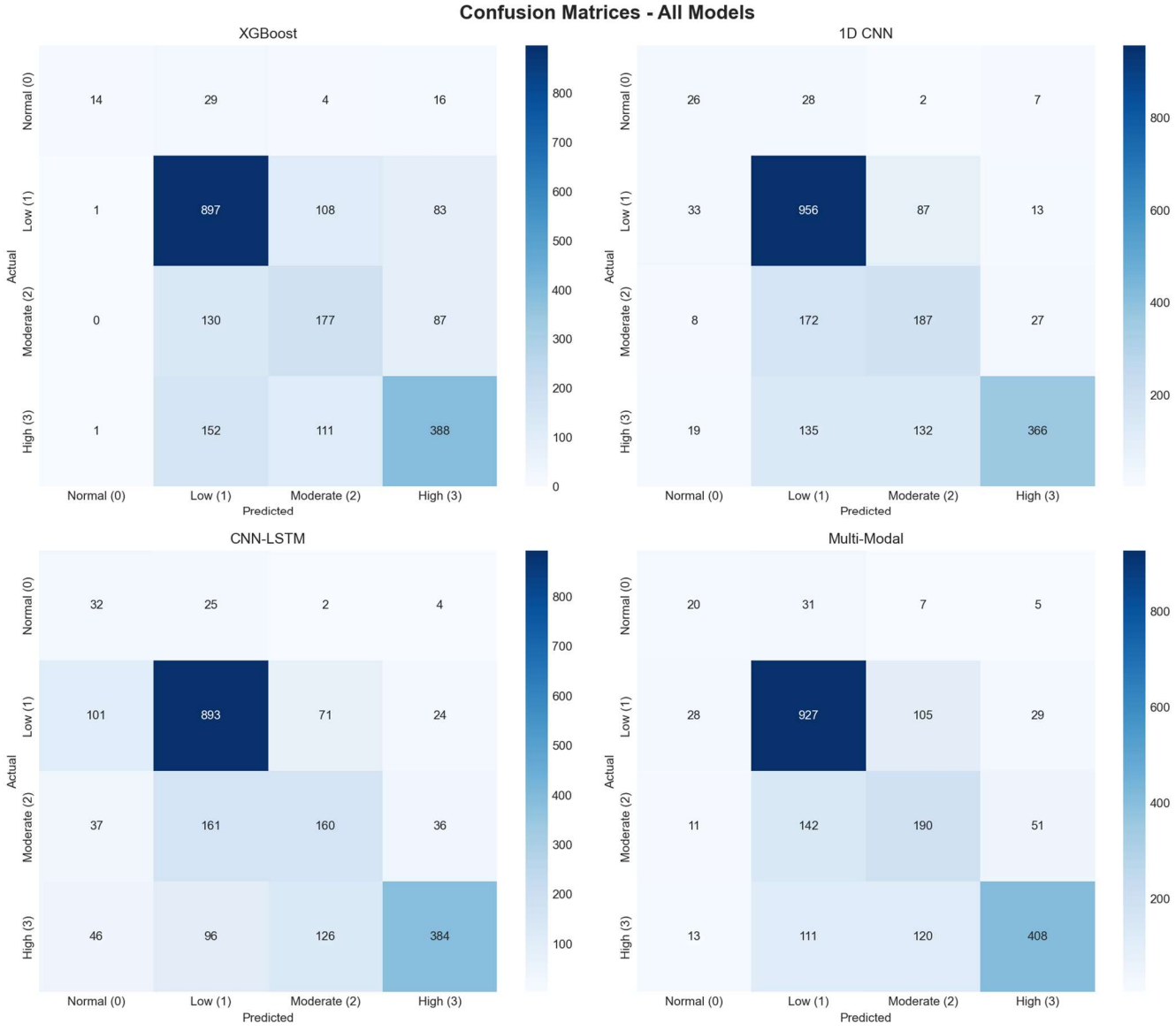

Figure 1. Confusion matrices for all four models on the test set, showing predicted vs. actual risk class distributions.

3.2 Model Explainability

SHAP TreeExplainer applied to XGBoost (Figure 2) reveals that HRV metrics, particularly SDNN, RMSSD, and mean heart rate, are dominant predictors for the high-risk class, consistent with clinical evidence linking reduced HRV to adverse outcomes. Grad-CAM adapted for 1D CNNs (Figure 3) visualizes temporal attention on ECG traces, showing the model focuses on QRS complexes (conduction assessment) and ST segments (ischemia detection), confirming clinically coherent behavior.

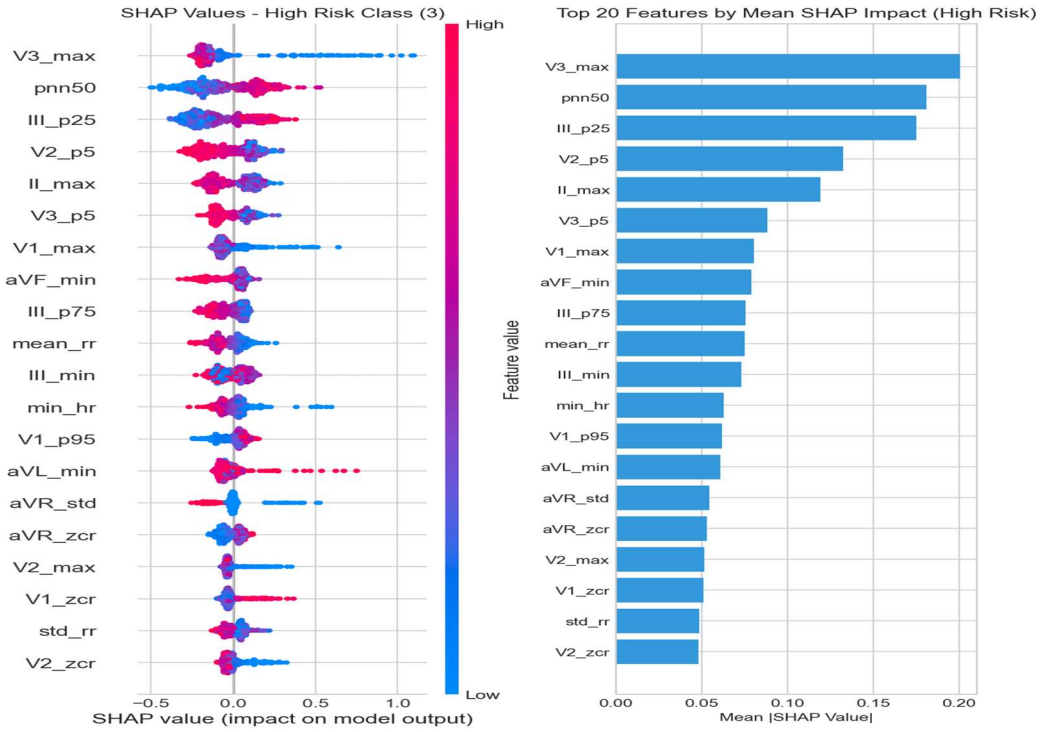

Figure 2. Top 20 features by mean |SHAP value| for the high-risk class (XGBoost), showing feature importance for risk prediction.

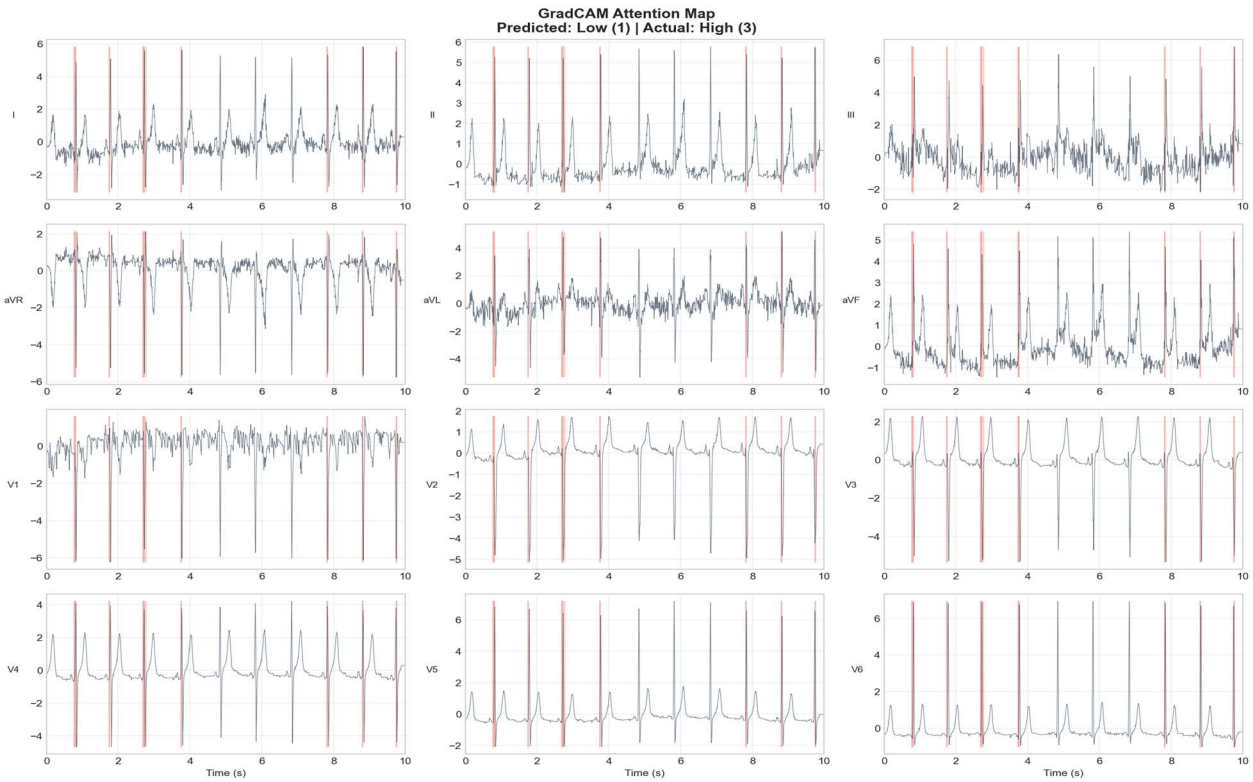

Figure 3. Grad-CAM attention heatmaps for high-risk ECG samples. Warm regions indicate where the CNN focuses its prediction.

4. System Outputs and Clinical Application

4.1 Individualized Patient Risk Reports

For each patient, HeartGuard generates a comprehensive risk report containing: the predicted risk category (Low, Moderate, or High) with a numerical composite score; detected ECG diagnoses and their clinical significance; key physiological metrics including heart rate and HRV indicators; and a visual explanation showing which ECG regions the model attended to via Grad-CAM. Sample reports were generated for representative patients across all risk tiers. A low-risk patient (age 40, normal sinus rhythm, composite score 0.26) received routine monitoring guidance. A high-risk patient (age 30, anteroseptal MI, composite score 0.68) was flagged for urgent cardiology evaluation with a critical-urgency alert.

4.2 Evidence-Based Intervention Recommendations

The system maps each patient's findings to categorized interventions across five domains: immediate actions, medication recommendations, lifestyle modifications, monitoring schedules, and specialist referrals. Each recommendation is assigned an urgency level (critical, high, medium, low) and, where supported by clinical trial data, a quantified expected risk reduction. For example, statin therapy carries a 25-30% reduction in major adverse cardiac events, smoking cessation a 50% CVD risk reduction, and blood pressure control a 20-25% reduction per 10 mmHg systolic. Risk reduction is projected using a multiplicative model capped at 70% to maintain clinical plausibility. The high-risk MI patient described above received six prioritized recommendations, including dual antiplatelet therapy, high-intensity statin, and cardiac rehabilitation, with a projected overall risk reduction of 39%.

4.3 Population-Level Risk Stratification

Applied to the full test cohort (2,198 patients), HeartGuard produces population-level analytics for healthcare resource planning (Figure 4). The system stratifies the population into risk tiers, identifies the highest-risk individuals requiring immediate attention, and quantifies intervention coverage across the cohort. This analysis reveals which intervention types (medication adjustments, lifestyle counseling, specialist referrals) are needed most frequently, enabling hospitals to allocate cardiology resources and preventive care programs based on predicted demand. The population view complements individual reports by providing administrators with an aggregate picture of cardiovascular risk burden.

Figure 4. Population-level risk stratification across the test cohort, showing risk distribution by demographics and category.

5. Discussion and Conclusion

HeartGuard demonstrates that ECG-based machine learning can extend well beyond diagnostic classification into a complete clinical decision support pipeline. The 1D CNN achieves the best discriminative performance (AUC 0.8345) with a relatively simple architecture, while the Multi-Modal fusion model achieves the highest accuracy (70.3%), confirming that engineered HRV features complement raw signal representations. Critically, the system's value lies not only in prediction accuracy but in its downstream outputs: explainable risk reports that clinicians can audit, intervention plans grounded in published trial evidence, and population analytics that support healthcare resource allocation.

The explainability pipeline validates clinical coherence at two levels. SHAP analysis confirms that the model relies on established cardiac risk markers (HRV, precordial lead morphology), while Grad-CAM attention maps focus on diagnostically relevant ECG segments such as QRS complexes and ST intervals. Together, these provide the interpretive transparency required for physician trust in automated recommendations.

Limitations

Risk labels are derived from ECG diagnoses rather than true clinical outcomes such as mortality or major adverse cardiovascular events. Clinical features including blood pressure, cholesterol, and comorbidity flags are synthetically generated based on diagnostic correlations. The PTB-XL dataset originates from a single center, and generalization to diverse populations remains unvalidated. The normal-risk class is underrepresented in the test set, affecting per-class F1 for that category. Prospective clinical validation is required before deployment.

Future Directions

Future work should target validation with true outcome data, prospective clinical studies, electronic health record integration for real patient data, real-time monitoring capabilities, and multi-center generalization testing.

6. References

[1] Wagner, P., et al. (2020). "PTB-XL, a large publicly available electrocardiography dataset." Scientific Data, 7(1), 154.

[2] Lundberg, S. M. & Lee, S. I. (2017). "A unified approach to interpreting model predictions." NeurIPS, 30.

[3] Selvaraju, R. R., et al. (2017). "Grad-CAM: Visual explanations from deep networks via gradient-based localization." ICCV, 618-626.

[4] Ribeiro, A. H., et al. (2020). "Automatic diagnosis of the 12-lead ECG using a deep neural network." Nature Communications, 11(1), 1760.

[5] Chen, T. & Guestrin, C. (2016). "XGBoost: A scalable tree boosting system." ACM SIGKDD, 785-794.

[6] Hannun, A. Y., et al. (2019). "Cardiologist-level arrhythmia detection using a deep neural network." Nature Medicine, 25(1), 65-69.

[7] Task Force of the ESC. (1996). "Heart rate variability: Standards of measurement and clinical use." European Heart Journal, 17(3), 354-381.

[8] Strodthoff, N., et al. (2021). "Deep learning for ECG analysis: Benchmarks and insights from PTB-XL." IEEE JBHI, 25(5), 1519-1528.