Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Developmentally Normalized ECG Modeling Enables Interpretable Risk Stratification for Pediatric Congenital Heart Disease

Abla Bedoui

abla.bedoui@liu.edu

Long Island University

Inés Martinez

Hospital General de Ninos "Dr. Pedro de Elizalde"

Ignacio Lugones

Long Island University

Mohammed Cherkaoui

Long Island University

Article

Keywords:

Posted Date: June 1st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8960191/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Developmentally Normalized ECG Modeling Enables Interpretable Risk Stratification for Pediatric Congenital Heart Disease

Abla Bedoui[1], Inés Martinez[2], Ignacio Lugones[1], Mohammed Cherkaoui[1]

[1]School of Engineering, Computer Science and Artificial Intelligence, Long Island University, New York, United States.
[2]Unidad de Cirugıa Cardiovascular, Hospital General de Ninos "Dr. Pedro de Elizalde", Buenos aires, Argentina.

Contributing authors: abla.bedoui@liu.edu;

Congenital heart disease (CHD) remains a leading cause of morbidity in pediatric populations, with early identification posing challenges, especially in settings lacking specialist imaging access. An interpretable electrocardiogram (ECG)-based risk model was developed, integrating developmental normalization and physiologically structured feature combination to generate calibrated probability estimates for CHD screening. A total of 14,190 pediatric ECGs (CHD prevalence 20.3%) were analyzed. Morphological intervals, voltage features, and autonomic biomarkers were extracted and normalized according to age-, heart rate-, and sex-specific expectations. Gradient boosting classifiers were trained using stratified cross-validation. The final model achieved an area under the receiver operating characteristic curve (AUROC) of 0.86 with acceptable calibration (Brier score 0.13; expected calibration error 0.13). Prespecified rule-out (sensitivity $\geq$ 95%) and rule-in (specificity $\geq$ 95%) operating points enabled three-tier clinical risk stratification. Ablation analyses indicated that developmental normalization preserved discrimination and enhanced interpretability. Independent pediatric cardiology review confirmed concordance between high-risk predictions and clinically abnormal ECG patterns. Autonomic biomarkers contributed substantially to risk prediction, indicating a physiological signal warranting further investigation. These findings support developmentally informed, calibrated ECG-based

probabilistic risk modeling as a triage-oriented risk stratification strategy within pediatric clinical workflows.

1 Introduction

The 12-lead electrocardiogram (ECG) is an inexpensive, non-invasive diagnostic tool that captures surface representations of cardiac activation and impulse conduction. Since its clinical introduction in the early twentieth century, the ECG has evolved from a primary tool for arrhythmia detection to a broadly applied modality for identifying structural heart disease and ischemic pathology. Increasing evidence suggests that ECG waveforms encode rich physiological information extending beyond traditional interval interpretation, reflecting both cardiac and systemic disease states [1–3].

Congenital heart disease (CHD) represents the most common category of congenital malformations, affecting approximately 8–10 per 1,000 live births worldwide and contributing substantially to lifelong morbidity and healthcare utilization [4]. Despite advances in surgical and interventional cardiology, early detection particularly during infancy and early childhood remains a major clinical challenge. While echocardiography is the diagnostic standard, it is resource-intensive and not universally accessible. In contrast, ECG is widely available and routinely performed in pediatric care; however, its role in structured CHD risk stratification remains limited.

A central challenge in pediatric ECG interpretation is developmental confounding. Unlike adult ECGs, pediatric waveforms evolve continuously with age, body size, heart rate, and maturation of the autonomic nervous system [5, 6]. Normal developmental processes include shortening of PR and QRS intervals, age-dependent QT dynamics, and marked variability in heart rate variability (HRV) indices during early life [6, 7]. These physiological trajectories frequently overlap with manifestations of structural heart disease, conduction abnormalities, and arrhythmogenic substrates, complicating differentiation between age-appropriate variation and pathological deviation [8, 9].

In clinical practice, pediatric cardiologists implicitly account for these maturational effects by interpreting ECG features relative to age-specific reference ranges and integrating findings across mechanistic domains, including conduction, repolarization, and autonomic regulation [10]. However, this approach requires substantial expertise and may be subject to inter-observer variability [8].

Recent advances in artificial intelligence (AI) have demonstrated that ECG signals contain diagnostic and prognostic information beyond conventional interpretation [1–3]. Most ECG-AI approaches rely on either end-to-end waveform modeling with deep neural networks or static feature vectors optimized for discrimination metrics such as the area under the receiver operating characteristic curve (AUROC). While these strategies have shown strong performance in adult populations, they often assume relatively stable electrophysiological baselines and may not generalize well to pediatric populations characterized by dynamic developmental variation [11]. Moreover, calibration and clinical operating characteristics are less frequently emphasized, despite their importance for deployment in real-world screening contexts [12]. Limited interpretability further constrains integration into clinical workflows [13].

To address these challenges, we developed CHD-SMART, a developmentally normalized ECG-based framework for probabilistic CHD risk estimation in pediatric populations. Rather than operating directly on raw ECG measurements, we model expected age-, heart rate-, and sex-dependent trajectories of ECG biomarkers using regression-based normalization approaches analogous to growth-curve modeling [14, 15]. Standardized residuals quantify deviations from age-appropriate physiological expectations, isolating candidate pathological signal from maturational effects.

Normalized features are subsequently aggregated into mechanistically structured axes representing conduction, repolarization, and autonomic regulation, reflecting established domains of cardiac electrophysiology [10]. Gradient-boosted classifiers are trained on these structured representations to generate calibrated CHD probability estimates. Model interpretability is assessed using Shapley additive explanations (SHAP) to quantify feature-level and axis-level contributions to risk predictions [16].

Using 14,190 pediatric ECG recordings, we evaluate discrimination, calibration, subgroup robustness, and clinically motivated operating characteristics. We demonstrate that developmental normalization reduces age-related confounding while preserving discrimination and enabling physiologically structured risk representation. These findings support a decision-oriented approach to pediatric ECG-based CHD screening grounded in developmental physiology and calibrated probabilistic modeling.

2 Methods

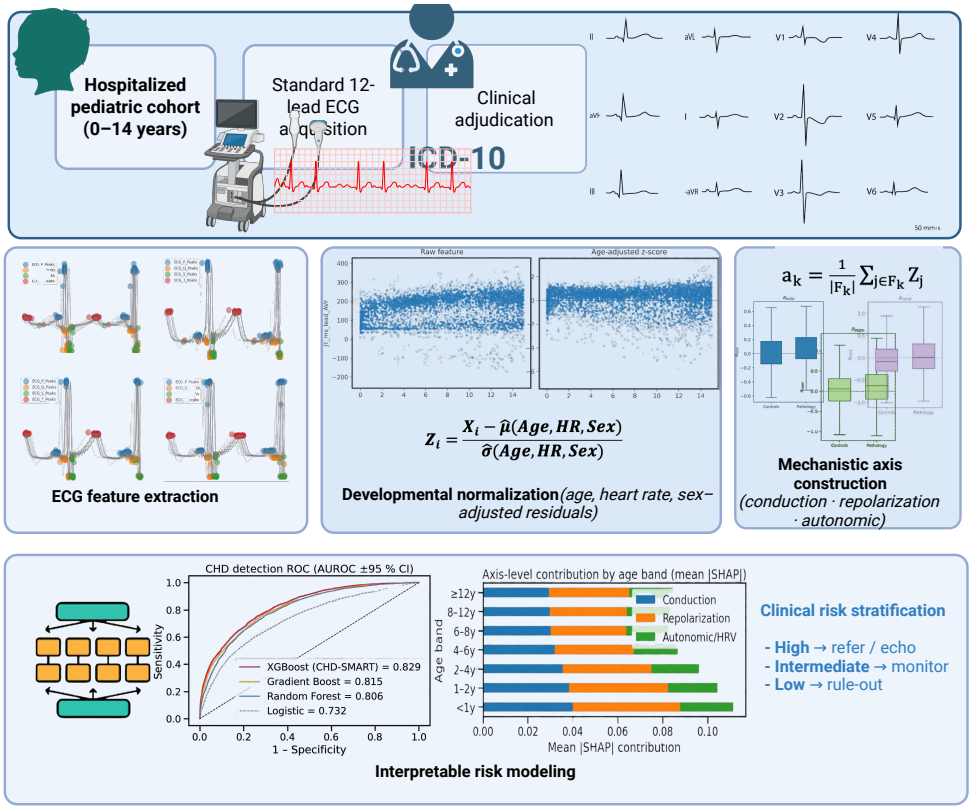


Fig. 1: **Overview of the CHD-SMART pipeline.** Hospital-acquired pediatric ECGs undergo feature extraction followed by developmental normalization, where ECG biomarkers are transformed into age-, heart rate–, and sex-adjusted residuals. Normalized features are aggregated into mechanistic axes capturing conduction, repolarization, and autonomic regulation. A gradient-boosted model integrates these representations to generate calibrated and explainable CHD risk estimates, supporting clinically interpretable risk stratification for screening and triage.

2.1 Study Design and Overview

We developed CHD-SMART, a machine-learning pipeline for estimating congenital heart disease (CHD) risk from standard 12-lead electrocardiograms (ECGs). The central methodological component is developmental normalization of ECG biomarkers to remove age-, heart rate–, and sex-dependent physiological variation prior to model training. The pipeline comprises four stages: (i) ECG signal processing and feature

extraction, (ii) developmental normalization using regression-based modeling and z-score transformation, (iii) aggregation of normalized features into physiologic axes, and (iv) supervised probabilistic CHD risk modeling with calibration and interpretability analysis.

Dataset Description

The primary dataset comprised 14,190 pediatric 12-lead ECG recordings collected from hospitalized children aged 0–14 years [17]. ECGs were sampled at 500 Hz with recording durations ranging from 5 to 120 seconds and included 9–12 standard leads. Only one ECG per patient was included in the analysis. ECGs were excluded if they exhibited excessive noise, incomplete demographic metadata, or physiologically implausible measurements. Clinical diagnoses were assigned according to International Classification of Diseases, Tenth Revision (ICD-10) codes. Children with confirmed congenital heart disease (CHD) were labeled as CHD-positive. Recordings without structural or electrical abnormalities served as controls. All ECGs were independently reviewed by two pediatric cardiologists to ensure diagnostic consistency.

2.2 Clinical Expert Review of Representative ECG Cases

To assess the clinical plausibility of model outputs, representative ECGs across risk strata (true positive, true negative, false positive, false negative, and intermediate-risk cases) were independently reviewed by a pediatric cardiologist blinded to model development. Full 12-lead ECGs were evaluated according to standard pediatric criteria, including rhythm assessment, QRS axis determination, voltage criteria for chamber enlargement, and repolarization analysis. This qualitative review was used to contextualize model predictions and assess concordance between predicted risk categories and established clinical interpretation.

2.3 ECG Signal Processing and Developmental Normalization

Digital 12-lead ECGs were processed using a standardized Python pipeline (WFDB, NeuroKit2, and custom routines). Signals were bandpass filtered to reduce baseline wander and high-frequency noise while preserving waveform morphology. Recordings were segmented into cardiac cycles, and R peaks were detected using the Pan–Tompkins algorithm. Automated delineation identified P-wave onset and offset, QRS boundaries, and T-wave peaks.

From delineated waveforms, quantitative biomarkers were derived, including interval measurements (PR, QRS, QT, JT, ST, Tp–Te), lead-specific voltage amplitudes, and heart rate variability (HRV) metrics. HRV features included time-domain measures (SDNN, RMSSD, pNN50) and nonlinear descriptors (sample entropy, detrended fluctuation analysis). Additional spectral and waveform features were obtained via discrete wavelet decomposition (Daubechies-4) and slope-based descriptors. All features were assembled in tabular format indexed by subject.

Pediatric ECG morphology varies systematically with age and physiological maturation [18, 19]. To account for these developmental effects, each feature was normalized

relative to its expected value using a polynomial regression model fitted on control subjects, conceptually analogous to growth-curve normalization methods [20],

$$\hat{\mu}_k(\text{Age}, \text{HR}, \text{Sex}) = \mathbf{x}^\top \hat{\boldsymbol{\beta}}_k, \tag{1}$$

where $\mathbf{x}$ comprised polynomial terms of age and HR (including interaction terms) and sex as a covariate. Standardized residuals were then computed as,

$$z_k = \frac{f_k - \hat{\mu}_k(\text{Age}, \text{HR}, \text{Sex})}{\hat{\sigma}_k}, \tag{2}$$

with $\hat{\sigma}_k$ denoting the residual standard deviation estimated from the control population. These z-scores quantify deviations from age-appropriate physiologic norms independent of maturational effects and serve as developmentally adjusted biomarkers for downstream modeling. Normalization models were fitted exclusively within training folds during cross-validation to prevent information leakage into validation data. To enhance physiological interpretability, developmentally normalized features were aggregated into mechanistically motivated axes reflecting established electrophysiological domains: atrioventricular and intraventricular conduction (PR and QRS), ventricular repolarization (QT, JT, TpTe), and autonomic regulation (HRV-derived measures) [21, 22]. The conduction axis was defined as,

$$a_{\text{cond}} = \frac{1}{|F_{\text{cond}}|} \sum_{k \in F_{\text{cond}}} z_k, \tag{3}$$

where F_{cond} includes interval features related to atrioventricular and intraventricular conduction. The repolarization axis was defined as

$$a_{\text{repol}} = \frac{1}{|F_{\text{repol}}|} \sum_{k \in F_{\text{repol}}} z_k, \tag{4}$$

capturing ventricular recovery abnormalities. The autonomic axis was computed as

$$a_{\text{auto}} = -\frac{1}{|F_{\text{auto}}|} \sum_{k \in F_{\text{auto}}} z_k, \tag{5}$$

where F_{auto} comprises HRV and entropy-based features. The autonomic axis was sign-inverted such that reduced variability corresponds to higher axis values, enabling consistent directional interpretation across axes.

2.4 Machine Learning Risk Modeling

Supervised models were developed to estimate patient-level CHD risk from developmentally normalized features. Performance was evaluated using 5-fold stratified cross-validation at the patient level. Missing values were imputed using median imputation within each training fold to prevent data leakage. We evaluated regularized logistic regression, random forest, and gradient-boosted decision trees. Gradient boosting (XGBoost) was selected as the primary model based on discrimination and probability stability. Hyperparameters were optimized using randomized search within cross-validation. Model performance was summarized using out-of-fold predicted probabilities.

2.5 Statistical Analysis

Discrimination was assessed using the area under the receiver operating characteristic curve (AUROC). Ninety-five percent confidence intervals were estimated via non-parametric bootstrap resampling (2,000 iterations). Moreover, calibration was evaluated using calibration curves, calibration intercept and slope derived from logistic recalibration, Brier score, and expected calibration error (ECE). Clinically motivated thresholds were defined using (i) the Youden index and (ii) predefined rule-out (sensitivity $\geq 95\%$) and rule-in (specificity $\geq 95\%$) criteria derived from cross-validated predictions. Decision curve analysis (DCA) was performed to quantify net benefit across threshold probabilities:

$$\text{Net Benefit} = \frac{TP}{N} - \frac{FP}{N} \cdot \frac{p_t}{1 - p_t}, \tag{6}$$

where p_t denotes the decision threshold probability.

2.6 Model Interpretability

Model interpretability was assessed using SHapley Additive exPlanations (SHAP). SHAP values quantify feature contributions to individual predictions in log-odds units. Global importance was summarized by mean absolute SHAP values. Contributions were stratified by age bands to assess developmental variation in feature importance across pediatric subgroups.

3 Results

3.1 Developmental normalization and mechanistic axis construction

Raw ECG interval features demonstrated strong monotonic associations with age, reflecting expected maturational changes in pediatric electrophysiology. Across 143 conduction and repolarization biomarkers, Spearman correlations with age were moderate to strong (median $|\rho| = 0.35$–0.45; all $p < 10^{-6}$), confirming substantial developmental confounding. After applying regression-based normalization conditioned on age, heart rate, and sex, correlations between age and feature z-scores were markedly attenuated ($|\rho| < 0.05$ for the majority of biomarkers), indicating effective removal of maturational trends while preserving inter-individual variability (Supplementary Fig. S2). Developmentally normalized features were subsequently aggregated into three physiologic axes: conduction (a_{cond}), repolarization (a_{repol}), and autonomic regulation (a_{auto}). In controls, axis distributions were symmetric and centered near zero (Supplementary Fig. S3). In contrast, CHD recordings exhibited population-level shifts along one or more axes (Fig. 2), with elevated a_{cond} in conduction-related abnormalities and shifts in a_{repol} and a_{auto} corresponding to repolarization and autonomic perturbations.

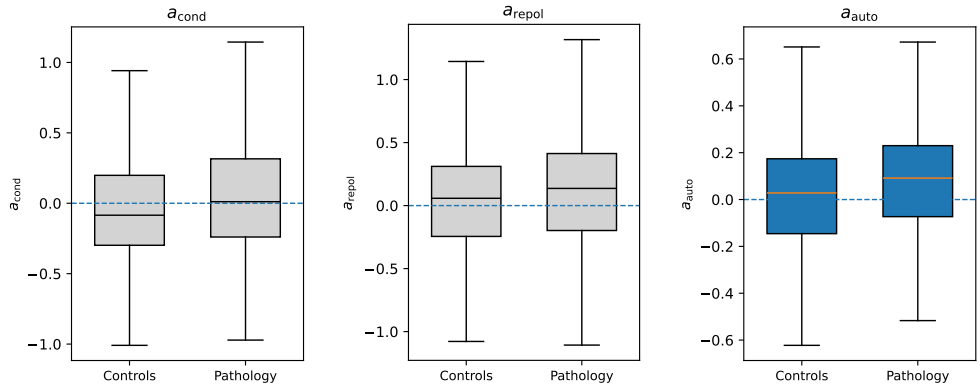


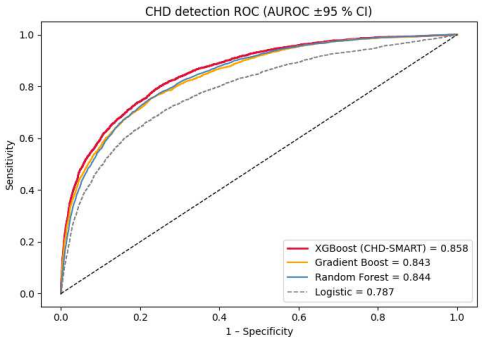
Fig. 2: **Mechanistic axes separate controls from pathology-associated recordings.** Distributions of (A) conduction (a_{cond}), (B) repolarization (a_{repol}), and (C) autonomic (a_{auto}) axes in control and CHD recordings. Axes are computed from developmentally normalized (z-scored) biomarkers and summarize physiologic deviations from age-appropriate norms. Population-level shifts indicate structured electrophysiologic perturbations in CHD.

3.2 CHD discrimination and clinically actionable risk stratification

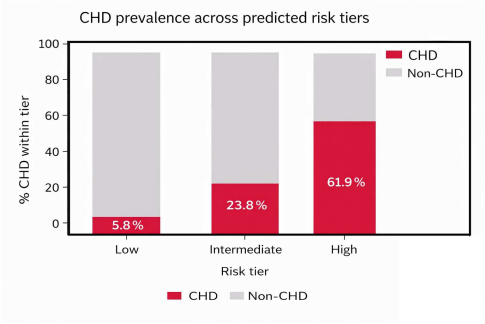
We evaluated CHD discrimination within a hospital-based pediatric cohort comparing CHD cases to normal controls. Using 5-fold stratified cross-validation at the patient level, the XGBoost model achieved an AUROC of 0.8583 (Fig. 3a). To assess clinical utility, predicted probabilities were stratified into low-, intermediate-, and high-risk tiers using predefined rule-out and rule-in criteria. Observed CHD prevalence increased monotonically across tiers in cross-validated predictions (low: 5.8%, intermediate: 23.8%, high: 61.9%), indicating effective risk enrichment with increasing predicted probability (Fig. 3b). Ablation analysis quantified the contribution of alternative feature representations (Table 1). Developmentally normalized features demonstrated comparable discrimination to raw-feature counterparts (AUROC 0.8583 vs 0.8584), indicating that normalization preserves predictive signal while reducing maturational confounding. Inclusion of demographic covariates modestly improved performance (AUROC 0.8716 vs 0.8707). Models restricted to the three mechanistic axes exhibited substantially lower discrimination (AUROC 0.61), suggesting that the full feature representation retains additional predictive information beyond axis-level summaries.

3.3 Model Explainability and Feature Contributions

We used SHAP (SHapley Additive exPlanations) values to measure how much each ECG-derived feature contributed to the predicted probability of congenital heart disease (CHD) [23]. Figure 4 shows the overall importance and direction of the most influential features. Heart rate variability (HRV) indices emerged as the top predictors, with HRV_SDNN_z and HRV_MeanNN_z exhibiting the highest mean absolute



(a) Training set ROC (CHD vs non-CHD).



(b) Observed CHD prevalence across risk tiers.

Fig. 3: **Risk stratification based on predicted CHD probability.** (**a**) Discrimination in cross-validation using ROC analysis. (**b**) CHD prevalence increases monotonically across tiers, supporting a triage-oriented use of CHD-SMART (rule-out / monitor / refer).

Table 1: **Ablation study across input representations.** Performance of the same gradient-boosted model evaluated with 5-fold out-of-fold probabilities. Raw (matched) denotes the raw counterparts of the same features used in the developmentally normalized (z) set.

Input representation	Dim	AUROC	Brier	ECE
Raw + Age/HR/Sex	155	0.8716	0.1275	0.1321
DevNorm (z) + Age/HR/Sex	146	0.8707	0.1279	0.1323
Raw (matched)	152	0.8584	0.1344	0.1389
DevNorm (z)	143	0.8583	0.1342	0.1384
Axes only	3	0.6121	0.2317	0.2599

SHAP values, indicating the greatest contribution to the model. Repolarization and conduction features, such as QT interval (lead III), PR interval (lead V4), and QRS duration (lead V1), also exerted substantial influence. The SHAP beeswarm plot revealed distinct patterns: higher HRV-related feature values were generally associated with increased predicted CHD risk, whereas certain conduction and repolarization markers showed bidirectional effects across leads.

3.4 Clinical Operating Characteristics and Subgroup Robustness

We tested clinical usability by checking how well the model worked at set probability thresholds for rule-out ($p \leq 0.33$) and rule-in ($p \geq 0.67$) decisions. We also analyzed subgroups by age, heart rate, sex, QRS duration, and tachycardia status (Figure 5). Across all subgroups, rule-out sensitivity was always above 0.80, and some pediatric

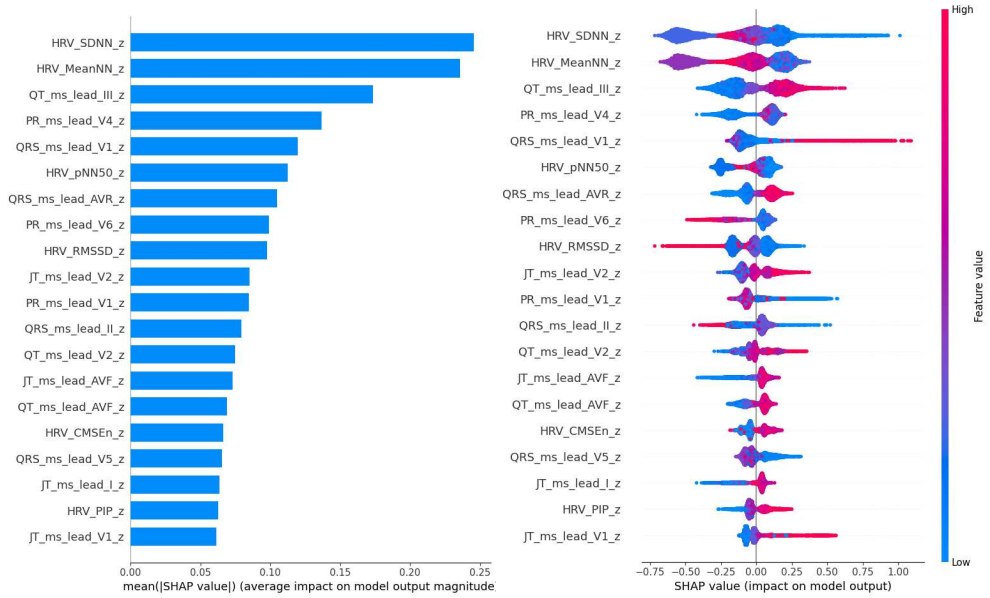


Fig. 4: Global model explainability using SHAP values. (A) Mean absolute SHAP values showing the most influential ECG-derived features. (B) SHAP beeswarm plot illustrating the direction and magnitude of feature effects on predicted CHD risk. HRV features dominate overall importance, with additional contributions from conduction and repolarization markers.

age groups and non-tachycardic patients had sensitivities over 0.90. Rule-in specificity was also high, often exceeding 0.90, indicating strong discrimination even across mixed clinical groups. No subgroup demonstrated a statistically significant reduction in sensitivity or specificity.

3.5 Calibration and Decision Curve Analysis

Calibration curves demonstrated appropriate risk dispersion with mild overestimation at higher probabilities, indicating appropriate risk dispersion with slight overestimation at the population level (Fig. 6). The Brier score and expected calibration error (ECE) were consistent with adequate probabilistic accuracy. Decision curve analysis showed positive net benefit across clinically relevant threshold probabilities compared with treat-all and treat-none strategies.

4 Discussion

Early identification of congenital heart disease (CHD) remains a persistent clinical challenge, particularly in environments where access to advanced imaging and pediatric cardiology expertise is limited. In this study, we demonstrate that routinely

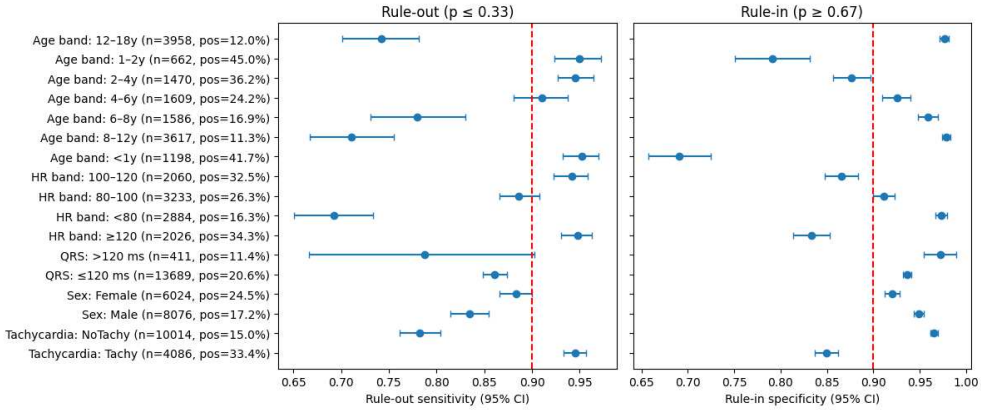


Fig. 5: Clinical operating attributes across predefined subgroups. Left panel: rule-out sensitivity at $p \leq 0.33$. Right panel: rule-in specificity at $p \geq 0.67$. Error bars indicate 95% confidence intervals. Performance remains stable across age, heart rate, sex, and conduction-based subgroups.

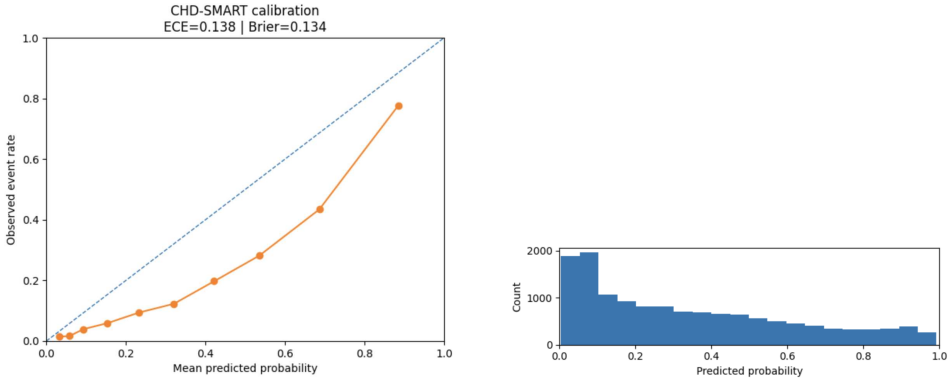


Fig. 6: **Calibration and clinical utility of CHD-SMART.** (**Left**) Calibration curve with probability distribution; dashed line indicates perfect calibration. We report ECE and Brier score to summarize probabilistic accuracy. (**Right**) Decision curve analysis showing the net benefit of CHD-SMART relative to the treat-all and treat-none strategies across threshold probabilities.

collected 12-lead ECG data can be transformed into calibrated probabilistic risk estimates that meaningfully stratify CHD likelihood within a hospital-based pediatric population.

A central contribution of this work is the explicit modeling of developmental confounding. Pediatric ECG morphology evolves dynamically with age, heart rate, and physiologic maturation. By normalizing ECG biomarkers relative to age-, heart rate–, and sex-specific expectations, the proposed framework isolates deviations from expected physiology rather than learning maturational patterns. Ablation analysis

demonstrated that developmental normalization preserves discrimination while reducing age-related confounding, supporting its role as a principled preprocessing strategy for pediatric ECG modeling.

Unlike many prior ECG-AI studies that prioritize discrimination alone, this work emphasizes probability calibration and operating characteristics. Although AUROC provides useful discrimination assessment, real-world clinical deployment depends critically on the reliability of predicted probabilities. The observed near-unity calibration slope, coupled with acceptable Brier score and ECE, suggests that risk dispersion is appropriate. The mildly negative intercept indicates modest population-level overestimation of absolute risk; therefore, recalibration would be recommended prior to deployment in populations with differing prevalence or case-mix. Given preservation of slope, simple intercept adjustment may suffice without structural model retraining.

Mechanistic axis aggregation provided additional interpretive structure. While models restricted to three physiologic axes exhibited lower discrimination than the full feature representation, axis-level summaries enabled physiologically coherent interpretation of risk shifts across conduction, repolarization, and autonomic domains. SHAP analysis further demonstrated that autonomic biomarkers contributed substantially to CHD risk prediction, suggesting that subtle autonomic signatures may accompany structural abnormalities. This finding warrants further investigation in future mechanistic and prospective studies.

Qualitative expert review provided an additional layer of clinical validation. High-risk predictions were generally concordant with abnormal voltage patterns, axis deviations, or rhythm disturbances identified according to pediatric cardiology standards. Although this review was not performed as a formal adjudication study, it supports alignment between model-derived risk estimates and established clinical interpretation.

This study has limitations. Analyses were conducted in a single tertiary pediatric cohort, and performance may differ in lower-prevalence screening settings. Although patient-level cross-validation and calibration analyses were implemented to mitigate overfitting, external validation in independent cohorts will be important before clinical deployment.

In summary, this work advances pediatric ECG-based CHD assessment from binary classification toward developmentally informed, calibrated risk modeling. By explicitly addressing maturational confounding and emphasizing probabilistic interpretation, the proposed framework supports triage-oriented risk stratification within pediatric clinical workflows. Future research should focus on multi-center validation, subtype characterization, and prospective evaluation of clinical impact.

Data Availability

The ECG dataset analyzed in this study is publicly available through the Figshare repository: Tan et al., *A pediatric ECG database with disease diagnosis covering 11,643 children* (https://doi.org/10.6084/m9.figshare.27078763).

All preprocessing, developmental normalization, and modeling code used to generate the results in this study are publicly available at: https://github.com/ablabedoui1/CHD_SMART.

Funding

The authors received no specific funding for this work.

Author Contributions

A.B. conceived and designed the study. A.B. developed the methodological framework, performed data processing, feature extraction, statistical analyses, and machine learning modeling. A.B. generated all figures and wrote the initial draft of the manuscript. I.L. and I.M. provided clinical expertise, contributed to interpretation of electrocardiographic findings, and performed qualitative expert review of representative ECG cases. M.C. supervised the project, contributed to study design and critical revision of the manuscript. All authors reviewed, edited, and approved the final manuscript.

Ethics Approval

The ECG dataset analyzed in this study was obtained from a publicly available repository. Ethical approval and patient consent were managed by the original data contributors as described in [17].

References

[1] Attia, Z.I., Noseworthy, P.A., Lopez-Jimenez, F., Asirvatham, S.J., Deshmukh, A.J., Gersh, B.J., Carter, R.E., Yao, X., Rabinstein, A.A., Erickson, B.J., Kapa, S.: An artificial intelligence–enabled ecg algorithm for the identification of patients with atrial fibrillation during sinus rhythm. Nature Medicine **25**(1), 70–77 (2019) https://doi.org/10.1038/s41591-018-0268-3

[2] Hannun, A.Y., Rajpurkar, P., Haghpanahi, M., Tison, G.H., Bourn, C., Turakhia, M.P., Ng, A.Y.: Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nature Medicine **25**(1), 65–69 (2019) https://doi.org/10.1038/s41591-018-0268-3

[3] Ribeiro, A.H., Ribeiro, M.H., Paixão, G.M.M., Oliveira, D.M., Gomes, P.R., Canazart, J.A., Ferreira, M.P.S., Andersson, C., Macfarlane, P.W., Meira, W., Schoeninger-Filho, P.: Automatic diagnosis of the 12-lead ecg using a deep neural network. Nature Communications **11**(1), 1760 (2020) https://doi.org/10.1038/s41467-020-15432-4

[4] Linde, D., Konings, E.E.M., Slager, M.A., Witsenburg, M., Helbing, W.A., Takkenberg, J.J.M., Roos-Hesselink, J.W.: Birth prevalence of congenital heart disease worldwide: a systematic review and meta-analysis. Journal of the American College of Cardiology **58**(21), 2241–2247 (2011) https://doi.org/10.1016/j.jacc.2011.08.025

[5] Davignon, A., Rautaharju, P., Boisselle, E., Soumis, F., Métras, D., Choquette, A.: Normal ecg standards for infants and children. Circulation **59**(2), 390–398 (1979) https://doi.org/10.1161/01.CIR.59.2.390

[6] Macfarlane, P.W., Coleman, E.N.: The normal ecg in childhood and adolescence. Heart **103**(12), 1004–1012 (2017) https://doi.org/10.1136/heartjnl-2016-310783

[7] Rijnbeek, P.R., Herpen, G., Bots, M.L., Man, S., Verweij, N., Hofman, A., Kors, J.A.: Normal values of the electrocardiogram for ages 0–16 years. European Heart Journal **35**(40), 2730–2739 (2014) https://doi.org/10.1093/eurheartj/eht472

[8] Hancock, E.W., Deal, B.J., Mirvis, D.M., Okin, P., Kligfield, P., Gettes, L.S.: Aha/accf/hrs recommendations for the standardization and interpretation of the electrocardiogram. Circulation **119**(10), 235–240 (2009) https://doi.org/10.1161/CIRCULATIONAHA.109.192675

[9] Park, M.K., Guntheroth, W.G.: How to read pediatric ecgs. Pediatric Cardiology **35**(6), 937–945 (2014) https://doi.org/10.1007/s00246-013-0858-7

[10] Kligfield, P., Gettes, L.S., Bailey, J.J., Childers, R., Deal, B.J., Hancock, E.W., Herpen, G., Kors, J.A., Macfarlane, P., Mirvis, D.M., Okin, P.: Recommendations for the standardization and interpretation of the electrocardiogram. Circulation **115**(10), 1306–1324 (2007) https://doi.org/10.1161/CIRCULATIONAHA.106.180200

[11] Liu, C., Li, Q., Clifford, G.D.: Deep learning for pediatric electrocardiogram analysis: challenges and opportunities. Frontiers in Cardiovascular Medicine **9**, 846329 (2022) https://doi.org/10.3389/fcvm.2022.846329

[12] Van Calster, B., McLernon, D.J., Smeden, M., Wynants, L., Steyerberg, E.W., tests, S., models": Calibration: the achilles heel of predictive analytics. BMC Medicine **17**(1), 230 (2019) https://doi.org/10.1186/s12916-019-1466-7

[13] Doshi-Velez, F., Kim, B.: Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608 (2017)

[14] Cole, T.J.: The lms method for constructing normalized growth standards. Statistics in Medicine **9**(1–2), 45–60 (1990) https://doi.org/10.1002/sim.4780090601

[15] Wright, E.M., Royston, P.: Age-specific reference intervals: modelling with fractional polynomials. Statistics in Medicine **16**(24), 2785–2803 (1997) https://doi.org/10.1002/(SICI)1097-0258(19971230)16:24⟨2785::AID-SIM725⟩3.0.CO;2-F

[16] Lundberg, S.M., Lee, S.-I.: A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems, vol. 30 (2017)

[17] Tan, J., Fan, H., Luo, J., *et al.*: A pediatric ecg database with disease diagnosis

covering 11,643 children. Scientific Data **12**, 867 (2025) https://doi.org/10.1038/s41597-025-05225-z

[18] Rijnbeek, P.R., Witsenburg, M., Schrama, E., Hess, J., Kors, J.A.: New normal limits for the paediatric electrocardiogram. European Heart Journal **22**(8), 702–711 (2001) https://doi.org/10.1053/euhj.2000.2399

[19] Davignon, A., Rautaharju, P., Boisselle, E., Soumis, F., Méthot, M., Choquette, A.: Normal ecg standards for infants and children. Pediatric Cardiology **1**(2), 123–131 (1979) https://doi.org/10.1007/BF02083173

[20] Cole, T.J., Green, P.J.: Smoothing reference centile curves: The lms method and penalized likelihood. Statistics in Medicine **11**(10), 1305–1319 (1992) https://doi.org/10.1002/sim.4780111005

[21] Antzelevitch, C., Yan, G.-X.: Cellular basis for the electrocardiographic j wave. Circulation **115**(16), 2099–2112 (2007) https://doi.org/10.1161/CIRCULATIONAHA.106.677625

[22] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology: Heart rate variability: standards of measurement, physiological interpretation and clinical use. Circulation **93**(5), 1043–1065 (1996) https://doi.org/10.1161/01.CIR.93.5.1043

[23] Lundberg, S.M., Lee, S.-I.: A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems, vol. 30, pp. 4765–4774 (2017)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterial.pdf