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.

Machine learning for classification of advanced rheumatic heart disease using electrocardiogram in cardiology ward

Amsalu Tomas Chuma

amsaluthomas.chuma@kuleuven.be

KU Leuven

Melkamu Hunegnaw Asmare

KU Leuven

Carolina Varon

KU Leuven

Desalew Mekonnen Kassie

Addis Ababa University

Chunzhuo Wang

KU Leuven

Bart Vanrumste

KU Leuven

Research Article

Keywords: Rheumatic Heart Disease, ECG, Machine Learning, Convolutional Neural Network, Relative Wavelet Energy, Global Health

Posted Date: June 17th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-6674556/v1

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

Additional Declarations: No competing interests reported.

Machine learning for classification of advanced rheumatic heart disease using electrocardiogram in cardiology ward

Amsalu Tomas Chuma[1,3*], Melkamu Hunegnaw Asmare[1,4], Carolina Varon[2], Desalew Mekonnen Kassie[5], Chunzhuo Wang[1], Bart Vanrumste[1]

[1*]Department of Electrical Engineering (ESAT), eMedia Research Lab/STADIUS, KU Leuven, Andreas Vesaliusstraat 13, Leuven, 3000, Belgium.
[2]Department of Electrical Engineering (ESAT), STADIUS, Leuven, 3001, Belgium.
[3]College of Engineering, Addis Ababa Science and Technology University, Addis Ababa, 16470, Ethiopia.
[4]Addis Ababa Institute of Technology, Center of Biomedical Engineering, Addis Ababa University, Addis Ababa, 385, Ethiopia. [5]Addis Ababa University, College of Health Sciences, Addis Ababa University, Addis Ababa, 5657, Ethiopia.
*Corresponding author(s). Email(s): amsaluthomas.chuma@kuleuven.be;

Abstract

Background: Rheumatic heart disease (RHD) arises from untreated streptococcal throat infections caused by beta-hemolytic group A streptococci, leading to progressive damage to cardiac valves. While echocardiography is the gold standard for RHD diagnosis, its use in low-income countries is limited due to scarce resources and a lack of trained professionals. Automated RHD detection via echocardiography and phonocardiography data has shown promising, but the effectiveness of electrocardiogram (ECG) for detecting RHD in endemic regions at cardiac wards with limited resource remains uncertain. This study explores the viability of ECG as a cost-effective tool for RHD detection in cardiac wards.

Methods: The study utilizes a dataset comprising single-lead ECG recordings from 124 confirmed RHD patients and 46 healthy controls collected at a major referral hospital in Ethiopia. Additionally, an extended-RHDECG dataset, which consists of age-matched ECGs from the Physikalisch-Technische Bundesanstalt (PTB-XL) dataset and RHD ECGs, was utilized. The single lead ECG segments of 10-second duration were resampled at 250Hz. Temporal and relative wavelet energy (RWE) features combined with Convolutional Neural Network (CNN) model features were employed for classification of prevalent cardiovascular diseases in the context of the Global South.

Results: A 5-folds cross-validation on RHDECG dataset using CNN model showed an accuracy of 88.6±0.2% in detecting RHD from healthy controls. This result was improved with combined features of CNN, temporal and RWE to an accuracy of 94.6±0.1%. Similar evaluation on extended-RHDECG yielded an accuracy of 51.2±0.3%, which reached to 69.1±0.5% considering the top-2 predictions of the six classes. The RWE and mainly the PR interval were the most relevant features for classification. Moreover, the RWE concentrations at higher frequency bands were significantly greater in the RHD group than in the healthy control group.

Conclusion: The findings suggest that single-lead ECG in combination with machine learning could serve as a potential tool for automatically classifying patients with advanced RHD. ECG time–frequency features effectively capture abnormalities associated with RHD, enhancing the performance and generalizability of deep CNN models. This approach offers a cost-effective tool that might be useful in addressing the RHD burden at resource-constrained cardiac clinics in alignment with global health development goals.

Keywords: Rheumatic Heart Disease, ECG, Machine Learning, Convolutional Neural Network, Relative Wavelet Energy, Global Health

1 Background

Rheumatic heart disease (RHD) is a preventable condition that remains a significant cause of morbidity and mortality in children and young adults, particularly in low- and middle-income countries [1, 2]. The disease is a consequence of recurrent, untreated group A streptococcal (GAS) infections, which progress to acute rheumatic fever (ARF), an autoimmune inflammatory response. This inflammatory reaction primarily affects the heart's valves, leading to structural damage, such as stiffness, fibrosis, or tissue loosening [1–4]. The cumulative mechanical stress on the valves results in severe and irreversible damage, often resulting in fatal outcomes. The mitral valve is most frequently affected, although RHD can impact all layers of the heart, including the pericardium, myocardium, and endocardium [1, 2, 5]. Early detection and management through benzathine penicillin G (BPG) injections can significantly reduce disease progression [5]. The disease's repercussions for public health are enormous and appear to persist in low-income countries [5–7].

The global burden of RHD is profound, with approximately 40 million people affected and over 310,000 deaths reported in 2019 [8]. Sub-Saharan Africa (SSA) accounts for approximately 70% of these cases, with the highest incidence among women aged 25 - 29 years [8, 9]. Epidemiological surveys in disease-prevalent countries of the global south reported a high prevalence of RHD [8]. Studies among school children aged 6 -18 years reported RHD prevalence rates of 19 per 1000 [10] rising to 56.7 per 1000 in rural areas of Ethiopia [11]. The disease disproportionately affects young working-age adults, leading to premature deaths, with an average age of death of approximately 25 years [3, 8].

Moreover, the global burden of disease studies (1990–2019) and other reports [12–15] show an increasing cardiovascular disease (CVD) burden in SSA. In individuals aged younger than 40 years, the leading causes of CVD-related deaths are RHD (14%), ischemic heart disease (12%), stroke (10%), and hypertensive heart disease (7%). For individuals under 30 years of age, RHD alone accounts for 15% of CVD mortality, which exceeds the combined contribution of the other three leading CVDs (14%). Challenges

in implementing diagnostic criteria, scarcity of healthcare resources, low awareness of ARF, and inadequate intervention policies exacerbate the RHD burden in high-risk populations [3, 7, 9]. These challenges led to a revision of the echocardiographic diagnostic criteria for RHD [7], suggesting an integration of primary prevention and screening, taking into account the higher susceptibility of younger adults to RHD than to other CVDs.

Echocardiographic-based diagnosis criteria are considered the gold standard for detecting RHD, offering better disease detection performance compared to cardiac auscultation or electrocardiogram (ECG) [7]. The research by [16] demonstrated the feasibility of automating the detection of advanced rheumatic heart disease cases from healthy individuals based on echocardiograph images and clips of mitral valve regurgitation features. The study achieved an area under the curve (AUC) of 93% and F1 score of 87%. However, this method requires the use of comprehensive echocardiograph recordings, which is not practical in RHD-prone areas. In contrast, the use of low-cost phonocardiogram (PCG) or ECG signals as optional alternatives for RHD detection has not been explored in depth.

In this context, this study aims to develop robust feature extraction methods and a classification model for detecting RHD cases in a resource-limited cardiac ward setting. This study's contributions are twofold; first, it provides an affordable RHD detection mechanism for high-risk populations in resource-limited settings; second, it enhances primary healthcare systems by reducing the diagnostic burden at cardiac referral hospitals. By combining time- and frequency-domain features with Convolutional Neural Networks (CNN) features, this study seeks to improve the accuracy of RHD detection, ultimately leading to better patient outcomes and more efficient use of medical resources. The rest of this paper is as follows: Section 2 presents related work, Section 3 discusses the datasets and methodology used for feature extraction and provides details on classification model architecture, where the results and comparisons of classification performance are discussed in Section 4. Finally, Section 5 and Section 6 discuss the results obtained, and conclusions are provided.

2 Related Work

The applications of echocardiograph images and auscultations for RHD detection have been explored in few existing studies. Evidence-based criteria, [17], rely on pathological and morphological findings of mitral and aortic valve regurgitation and stenosis. Recent studies emphasize that automated extraction of these features is crucial for RHD diagnosis. For example, the work in [16] used a dense 3D-CNN to extract morphological features of mitral regurgitation, which subsequently employed support vector machines (SVM). This model was evaluated on 528 subjects (43.8% were RHD positives and 50.2% were negative for RHD), and yielded an improved F1-score of 87%, compared to an F1-score of 80% for an end-to-end classification from echocardiographic clips using deep learning. Similarly, texture-based features from echocardiograph videos trained using SVM achieved 99.5% accuracy for the estimation of mitral regurgitation severity [18]. Another study in [19] employed a 3D-CNN for pediatric MR detection and reported F1-score ranging from 82% to 100% across different views, with parasternal long-axis views demonstrating the highest performance. Furthermore, the use of handheld ultrasound devices with a video-based 3D-CNN for visual feature extraction and supervised aggregation of meta features trained using Random Forest (RF) classifier was presented in [20]. The RF classifier achieved an accuracy of 73% when evaluated on 511 subjects (55.2% were RHD positive and 44.8% were negative for RHD), which outperformed the standard 2DCNN methods that leverage spatiotemporal features. Despite differences in echo views and data types, image quality and view specificity remain common challenges.

The PCG signal, a digitized recording of heart sounds during auscultation, is also valuable for the automatic detection of heart murmurs due to regurgitation and stenosis of the MV and AV using machine learning. A study in [21] proposed an SVM-based method to detect murmurs caused by RHD utilizing undecimated wavelet transforms with Daubechies (db5) wavelet and Mel frequency cepstral coefficient (MFCC) features. The model was evaluated on 40 RHD patients and 69 healthy subjects and achieved an accuracy of 74.5% and an F1 score of 70%. In contrast, [6] presented an RHD classifier

using 2D-CNN trained on log-Mel spectrograms and demonstrated an accuracy of 96.1%. However, the data distribution (RHD=124 vs. non-RHD=46) represents a cardiology ward setting that does not reflect population prevalence. While the work of [21] prioritized noise robustness but struggled with data quality, the work of [6] excelled with a deep learning-based approach in a controlled setting of RHD-burdened cardiac wards.

The ECG findings are not specific to RHD because of their overlap with other cardiac conditions [22, 23]. Common ECG findings in advanced RHD might include atrial fibrillation (AF), atrial enlargement, arrhythmias, PR and QT interval prolongation, P-wave changes, QRS axis deviations, ST-T changes, T-wave inversion, ST-segment elevation, and sinus tachycardia or bradycardia [22–27]. The presence of these findings mostly reflects myocardial involvement rather than direct markers of valvular damage. Automated detection of such ECG abnormalities has been reported in various studies. Studies [28, 29] used hybrid CNN-based and CNN-attention based models trained on the 12-lead ECG of the PTB-XL dataset (18869 patients, 5 superdiagnostic classes with 44 diagnostic subclasses) [30] and achieved an AUC score of 85%. This performance result was improved to an AUC of 89% by revisiting the classification task as an anomaly detection problem in [31]. Analogously, wide residual networks with squeeze-excite blocks employed on PhysioNet/CinC Challenge 2021 12-lead ECG dataset (43,344 patients with 27 diagnoses) [32] achieved the best challenge performance score of 0.57 [33].

Furthermore, the work of [34] on six common arrhythmia classification achieved 93% accuracy on a large-scale Chinese ECG benchmark dataset from 272,753 using the ECGNet model, which consists of mixed blocks of CNNs, residual, and BiLSTM with attention layers. Specifically, on a single-lead ECG, the differentiation of AF from normal and other arrhythmias by extracting continuous wavelet transform images demonstrated an accuracy of 91% [35]. Moreover, the study in [36] revealed that time-domain heart rate variability (HRV), parameters were significantly lower in ARF patients than in controls, which indicates reduced parasympathetic activity in ARF patients. As shown in [37], combining ECG diagnostic features such as temporal parameters, HRV,

frequency-domain and entropy characteristics [38, 39] and metadata improves model generalizability. Currently available large ECG databases lack recordings from RHD patients, and prior research has not explored automated RHD classification using ECG. Nevertheless, detection of the aforementioned RHD-related ECG abnormalities from other CVDs has demonstrated encouraging results.

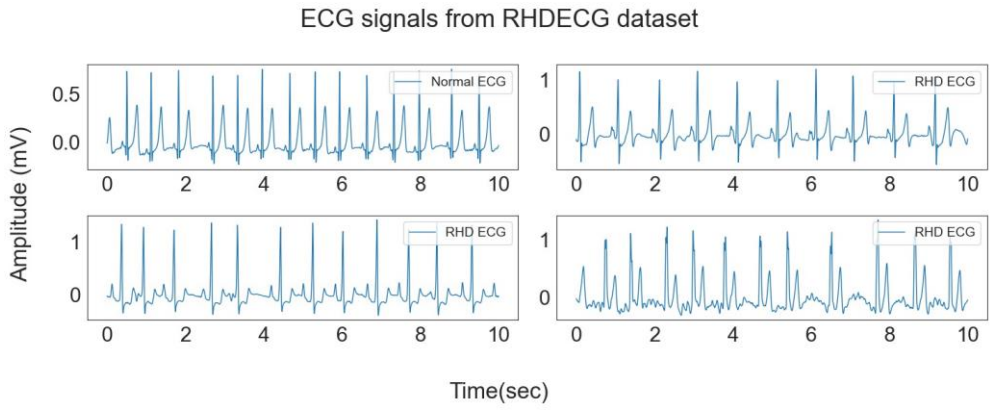
3 Methods

This section outlines the data and methods used to distinguish subjects with RHD from those with normal sinus rhythm or other common CVDs, namely, ischemic STT changes (STTC), hypertrophies (HYP), conduction disturbance (CD) and myocardial infarction (MI), which were obtained from [30].

3.1 Datasets

3.1.1 RHDECG dataset

The RHDECG dataset used for the experimental evaluation is described in [27]. The study was conducted at Tikur Anbessa referral teaching hospital (TASH), Ethiopia, with an ethical file reference 014/2018. The dataset comprised single-lead ECG recordings from 170 individuals, 124 of whom were diagnosed with RHD (PwRHD hereafter), and the remaining individuals had normal healthy diagnosis outcomes (healthy controls hereafter). Example ECG waveforms from both the RHD and healthy groups are shown in Figure 1. The recordings were obtained with a Beat2Phone™ ECG sensor (VitalSignum Oy, Helsinki, Finland) [40] for durations ranging from 1 to 3 minutes. The sampling frequency was 500Hz. Diagnostic decisions were made on the basis of echocardiograph assessment of the 2012 WHF criteria [17]. The mean and standard deviation (SD) of the age range of the PwRHD patients were 22.9±8.9 years, whereas those of the healthy controls were 14.4±10.4 years.



191

192 **Figure 1**. Example ECGs from PwRHD and healthy subjects

193 **Table 1**: Summary of the age and sex distributions of the datasets

	RHDECGdataset(Mean ± SD)		AddedfromPTB-XLdataset(Mean ± SD)				
Variables	PwRHD(n=124)	Normals(n=46)	Normals(n=1575)	STTC(n=79)	HYP(n=40)	CD(n=103)	MI(n=34)
Age	22.9 ± 8.9	14.1 ± 10.4	25.6 ± 6.1	27.3 ± 6.0	28.5 ± 5.8	25.7 ± 6.6	29.9 ± 4.2
Gender							
Female	74	15	648	32	36	77	26
Male	50	31	927	47	4	26	8

194

195 ### 3.1.2 PTB-XL dataset

196 The PTB-XL public dataset contains 21799 12-lead ECG 10-second recordings from
197 18869 patients. The recordings were also sampled at 500Hz. This dataset has five super
198 diagnostic classes (Normal, STTC, HYP, CD, and MI), which are split into 44 diagnostic
199 subclasses. Class labels are non-mutually exclusive; hence, some of the recordings are
200 categorized into multi-labeled classes. In this study, five diagnostic super classes, are
201 considered, as shown in Table 1. The patients' ages ranged between 0 and 95 years, with
202 a median of 62 years. Considering the CVD burden in cardiac clinics within RHD-endemic
203 regions and the age distribution of the RHDECG dataset cohort, age-matched 10-second
204 length signals from Lead-II were merged with the RHDECG dataset. An age range
205 between 5 and 35 years was considered. This combined dataset, referred to as the
206 extended-RHDECG hereafter, was used for further experimental evaluation.
207 Furthermore, multi-labeled records were excluded, and during training, a single record
208 per subject was used. A summary of the age and sex distributions per class in both
209 datasets is shown in Table 1.

3.2 Classification workflow

The classification process was started by recording the ECG signals using a mobile phone connected via Bluetooth to the Beat2Phone sensor and then transferring them to a computer for analysis. After preprocessing the ECG signals, 10-second segments were extracted, and time- and frequency-domain features were computed. The model was evaluated for classification in two experimental settings: one using the RHDECG dataset from patients with advanced RHD and the other adding agematched recordings from the PTB-XL dataset [30]. The extracted time- and frequency-domain features were combined with the features from the CNN model trained on the raw ECG data. Five folds stratified cross-validation was employed, and model explainability was assessed with SHapley Additive exPlanations (SHAP) [41] for time-frequency features and Gradient-weighted Class Activation Mapping (Grad-CAM) [42] for CNN features.

3.3 Preprocessing

Preprocessing of each ECG recording is essential due to the presence of noise. The RHDECG dataset was purged manually, wherein any unreadable segments were discarded from each recording, ensuring that a minimum of 10-seconds of clean segments per subject were retained. This resulted in discarding three subjects from 124 PwRHD, and one subject from 46 healthy groups. No subject was discarded from single labeled PTB-XL classes, except for a random selection of 100 records for the normal class to balance the class distribution with other classes. Each recording then underwent a denoising process consisting of two stages. First, a zero-phase bandpass fourth-order Butterworth with a low cutoff frequency of 0.5Hz to attenuate baseline wander noise, and a high cutoff frequency of 100Hz was applied to preserve diagnostic information within the recordings, as recommended in [43]. Subsequently, power line interference at 50Hz was removed using an IIR notch filter. After the denoising process, the signals were resampled at 250Hz. Since the inherent manifestation of abnormal ECGs in advanced RHD was consistently observed over time in the recordings, a window segment of 10-seconds was selected from a random position. The time-frequency

features were then computed from this segment for each subject. To train the CNN model, the segments were normalized to zero mean and unit variance.

3.4 Frequency and Temporal Features

The frequency domain and time domain diagnostic features are extracted from the preprocessed ECG segment. The frequency domain features were computed by performing maximal overlap discrete wavelet transform (MODWT) multiresolution analysis [44]. The time domain features related to the HRV and temporal parameters such as segments, intervals and wave peak amplitudes were examined. In total, eight wavelet energy features and 25 temporal features were extracted.

3.4.1 Relative wavelet energy features

The waves present in a normal sinus ECG signal often have overlapping frequency bands. Given the typical frequency ranges of fundamental ECG waves and their diagnostic significance [45–47], analyzing the signal's energy accentuation information in various bands while taking into account the overlapping boundaries is beneficial. For each ECG signal, **x(t)**, a wavelet energy is computed using MODWT with db4 at eight decomposition levels. The MODWT is an energy-preserving transform; thus, the total signal energy is distributed over the levels. In contrast to the dyadic grid arrangement of the wavelet and scale filters in DWT, MODWT does not downsample the coefficients, thereby maintaining the original signal length at each level. The wavelet coefficients ($W_{j,t}$) at decomposition levels are computed using the equations $\sum_{k=0}^{L-1} \tilde{h}_{j,k} x_{t-k}$, and the scaling coefficients at these levels ($V_{j,t}$) are computed as $\sum_{k=0}^{L-1} \tilde{g}_{j,k} x_{t-k}$, where $W_{j,t}$ are the detail coefficients at level j, $V_{j,t}$ are the scaling coefficients at level j, $\tilde{h}_{j,k} \in \{0...,L\}$ and $\tilde{g}_{j,k} \in \{0...,L\}$ are the MODWT wavelet and scaling filters, respectively.

The approximate boundary frequency range at each level, f_j, for a sampling frequency of f_s is given as $\left[\frac{f_s}{2^{j+1}}, \frac{f_s}{2^j}\right]$. Therefore, in the first decomposition level, there are no energies since the highest cutoff frequency of the signal is 100Hz. Since the recorded ECG is a finite discrete-time signal [48], its energy can be computed using eq. (1). The energy of the signal, E_j, for detail coefficients at decomposition level j is given by:

$$E_j = \frac{1}{N}\sum_{k=0}^{N-1}|D_j(k)|^2 \quad , \tag{1}$$

where N is the length of the coefficients at D_j, and D_j is the detail coefficient of the signal that corresponds to level j. Hence, the total energy, E_{tot}, over all the detail coefficients at decomposition levels is computed as:

$$E_{tot} = \frac{1}{N}\sum_{j}\sum_{k}^{N-1}|D_j(k)|^2 \quad . \tag{2}$$

The Relative Wavelet Energy (RWE) is calculated to extract the amount of wavelet energy that each band represents relative to the total energy of the input ECG signal, E_{tot}, using eq. (3). Thus, the RWE of the j^{th} decomposition level is the proportion contained in E_j out of E_{tot} and is written as:

$$RWE_j = \frac{E_j}{E_{tot.}} \tag{3}$$

Since RWE_j is the percentage distribution of energies of detail coefficients only, the output feature vector has a dimensionality of eight, where each dimension belongs to the frequency bands available in the input ECG.

3.4.2 Time-domain parameters

The detection of particular wave peaks and the delineation of their boundary fiducial points is a key step in parameters extraction and computation in an ECG signal. These fiducial points define the onset, offset and peak of each wave, as shown in Figure 2. Based on these delineated fiducial points, different time-domain parameters, such as intervals, segments and peak amplitudes, as shown in Figure 3, can be computed from each input ECG. These parameters were subsequently used as time-domain features. The list of time-domain features computed include the wave durations of the P-wave ($P_{duration}$) from [P_{onset},P_{offset}], T-wave ($T_{duration}$) from [T_{onset},T_{offset}], QRS complex ($QRS_{duration}$) from [QRS_{onset},QRS_{offset}], and intervals such as the RR interval [$R_{peak(i)}$,$R_{peak(i+1)}$], PR[P_{onset},QRS_{onset}], QT[QRS_{onset},T_{offset}], and STT[QRS_{offset},T_{offset}], where [a,b] indicates the duration between the incidence a and b. In addition to the aforementioned temporal parameters, the corrected QT interval (QTc) in eq. (4), the duration from the T-wave peak

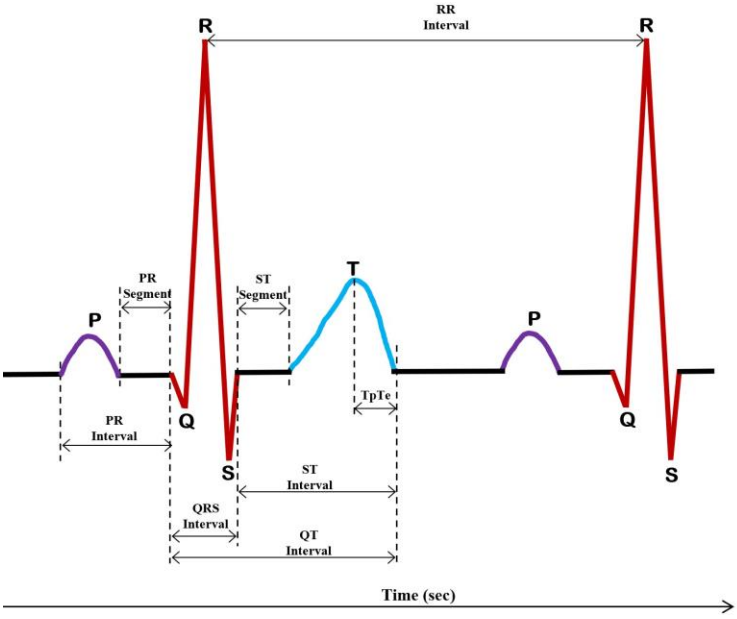


Figure 2. ECG waves, intervals and segments in a single heartbeat

to the end of the T-wave (TpTe), and the index of cardio electrophysiological balance (iCEB) in eq. (5) are calculated at every RR interval. The TpTe parameter reflects transmural dispersion of ventricular repolarization, whereas the QTc indicates overall activity of the ventricles in a single cardiac cycle. The QTc parameter is calculated according to Bazett's formula [49] as:

$$QTc = \frac{QT}{\sqrt{RR}} \tag{4}$$

Furthermore, the rate of dispersion of ventricular activities in each RR interval is also an important marker for the diagnosis of cardiac issues [50]. Thus, the ratio of TpTe over QTc (TpTe/QTc) is included in our parameters. The final parameter is iCEBc, which is cardiac wavelength and indicates the product of effective refractory period and conduction velocity [51], which correlates QTc with $QRS_{duration}$ in the given RR interval and is thereby computed as:

$$iCEBc = \frac{QTc}{QRS} \tag{5}$$

where for each RR interval, the QTc interval in seconds is divided by the QRS complex duration in seconds. This indicates the time taken for depolarization and repolarization of cardiac myocytes in a single cycle. To calculate iCEB, the numerator in eq. (5) is

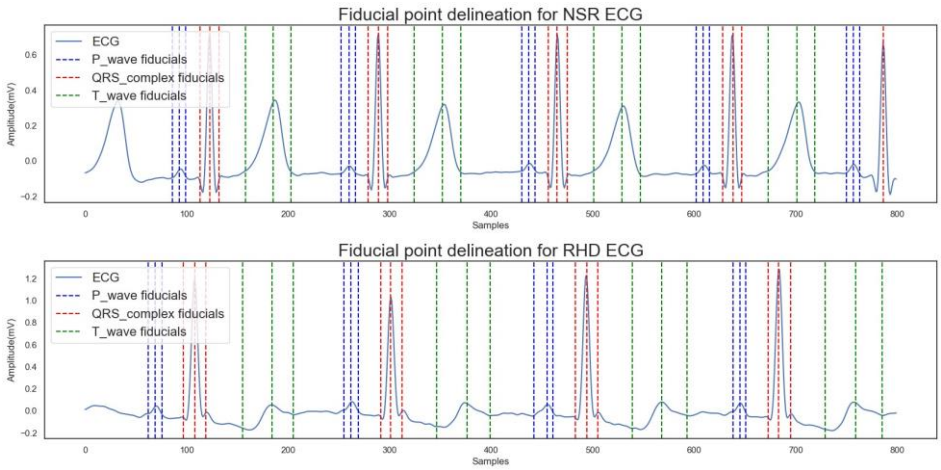


Figure 3. Delineated ECG fiducial points

changed to the QT interval rather than the QTc interval. All of these temporal parameters were computed on the basis of the detected fiducial peaks.

A benchmark comparative analysis of various ECG peak detection algorithms performed by [52] on a multitask ECG dataset reported NeuroKit as the best performing algorithm [52]. This algorithm was employed to find peak fiducial points and intervals related analysis for each record. The delineation result of a sample ECG record is shown in Figure 3. As indicated in [38], RRi values differing by more than 20% from the mean of the last 9 RRi values were considered ectopic beats. Analogous to [53], the records that contain ectopic beats above 15% of the total number of beats were not considered in the parameter computations. This condition was observed in ECG records from three subjects in the PwRHD class, where incorrect delineation of T-waves and P-waves led to prolonged parameter ranges beyond acceptable limits for their age groups [54–56]. To address such delineation deviations, statistical values from multiple ECG recordings per subject were considered. Beats with missing fiducial points were excluded by assigning Not-a-Number (NAN) values, and were thus not included in the temporal feature calculations. Then, features related to HRV were extracted from the remaining set of ECG recordings. The computed HRV-related features are shown in Appendix A.

The median values were calculated for each temporal feature, resulting in a total of 46 dimensions per ECG segment. Since the RWE features were normalized by the total energy, only the temporal features were converted to milliseconds.

Table 2: List of features extracted.

Frequency domain features	Temporal features	
	HRV related	Intervals and segments
RWE	HRV features (see Appendix A)	P duration, QRS duration T-duration, PR interval QT interval, STT interval QTc, TpTe, iCEBc, TpTe/QTc

To reduce dimensionality, recursive feature elimination with cross validation (RFECV) was applied, eliminating collinear features and selecting the best subset for input into classification models. The key features extracted are summarized in Table 2.

Statistical comparisons of distributions between subjects who had RHD and normal ECGs were assessed using a two-tailed Welsh's t-test or the Mann–Whitney U test, where appropriate. The former t-test is used for normally distributed dependent variables, and the latter test is used for nonnormally distributed dependent variables. The null hypothesis was defined as no significant mean differences between the wavelet energy features for the PwRHD and healthy groups having normal ECG. The cutoff alpha value was set to 0.05.

3.5 Model setup

The deep neural network model was an adaptation of InceptionTime [57], which outperformed other deep models in ECG classification [28]. The original model was modified by adjusting the depth to 6 and using inception kernels with sizes of 3, 5, 9, and 11 at each depth, as shown in Figure 4. Then, the time–frequency features were concatenated with the output feature map of the global average pooling layer (GAP) of the CNN model. The combined features were subsequently passed down to fully connected (FC) layers. During the training process, batch normalization (BN) helps mitigate weight spikes at various layers of the model. An Adam optimizer was utilized, starting with an initial learning rate of 0.001, which decayed exponentially every 30 epochs over time with a mini-batch size of 32. Additionally, a dropout rate of 0.2 was applied at the FC layer to prevent complex co-adaptations that could lead to overfitting.

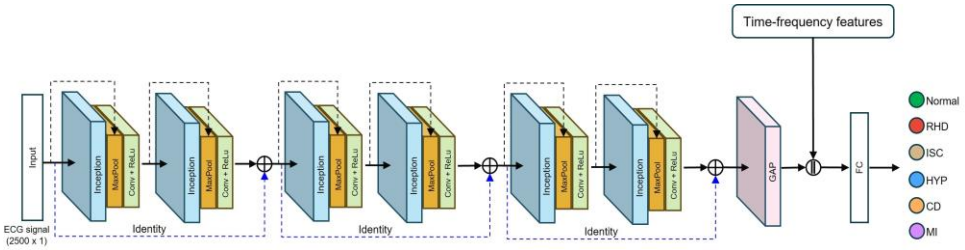


Figure 4. The inceptionTime CNN architecture, adapted from [57]

The model output layer was configured as a binary classifier for RHDECG evaluation, whereas a six-class multiclass classifier was employed for extended-RHDECG analysis.

3.5.1 Model evaluation metrics

Model performance was evaluated using accuracy score metric, which is defined in Equation (6).

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (6)$$

where TP represents true positive, TN represents true negative, FP represents false positive, and FN represents false negative predictions of the model. Model performance was evaluated through five-folds group stratified cross-validation. In this approach, the dataset was split into five equal sets while ensuring that the subjects included in the training set were mutually exclusive from those in the test set; therefore, no subject data leakage into the test set occurred. Hence, the model prediction was at the subject level.

4 Results

4.1 Statistical evaluation of time–frequency features

A statistical comparison of the time–frequency features of the RHD and healthy groups is shown in Table 3 and Table 4. Table 3 presents the t-test results at each decomposition level for the RWE percentage distributions for the RHDECG dataset. The t-test results revealed a significant mean difference between PwRHD and normal ECGs at higher frequency bands above 4Hz, and at 0.9Hz - 1.9Hz. However, RWE at bands 1.9Hz - 3.9Hz (t(Healthy = 45, PwRHD=121) = 0.14, p=0.747) and 0.5Hz - 0.97Hz (t(Healthy = 45, PwRHD = 121) = 0.89, p=0.394) is not significant for the cutoff value of 0.05. The energy concentrations over all the bands in PwRHD showed greater deviations from the

Table 3: Statistical t-test on relative wavelet energy percentage distributions between PwRHD vs Healthy controls in the RHDECG dataset

RWE Frequency Bands	PwRHD(mean±SD)	Normals(mean±SD)	p value
31.25Hz – 62.5Hz	0.8±0.57	1.3±0.5	<**0.001**
15.6Hz – 31.25Hz	12.4±6.0	15.1±3.4	<**0.001**
7.8Hz – 15.6Hz	29.9±9.1	27.6±7.2	**0.004**
3.9Hz – 7.8Hz	26.2±6.3	23.2±3.9	<**0.001**
1.9Hz – 3.9Hz	16.2±6.5	16.1±4.7	0.747
0.9Hz – 1.9Hz	7.9±4.9	10.7±5.4	<**0.001**
0.5Hz – 0.97Hz	4.5±5.5	4.1±4.6	0.394

The bold text shows statistically significant (p value <0.05) mean differences for both classes.

Table 4: Statistical test of temporal variables between RHD patients and healthy subjects for the RHDECG and extended-RHDECG datasets

Independent variable	p value	
	RHDECG	extended-RHDECG
P duration	< **0.001**	**0.002**
QRS duration	< **0.001**	< **0.001**
T duration	0.236	**0.016**
PR	< **0.001**	< **0.001**
QT	0.855	0.191
STT	**0.024**	< **0.001**
RR	0.339	0.788
QTc	0.164	0.794
TpTe	0.702	0.436
TpTe/QTc	0.179	0.356
iCEBc	< **0.001**	< **.001**
MeanNN	**0.014**	0.104
MadNN	**0.032**	0.065
MCVNN	**0.029**	0.061
IQRNN	**0.009**	0.051
HTI	**0.002**	**0.002**
SDNN	0.850	0.265
RMSSD	0.952	0.095
SDSD	0.937	0.092
CVNN	0.402	0.270
CVSD	0.476	0.084
MedianNN	0.063	0.063
Prc20NN	0.320	0.773
pNN50	0.932	**0.023**
pNN20	0.883	**0.030**

The bold text shows statistically significant (p value <0.05) mean differences for both classes.

mean than those in the Normal ECGs. While none of the healthy controls had ectopic beats occurring above 15% of the number of beats, this was observed in approximately 4% (n=3) of the PwRHD patients. There were significant differences in the temporal parameters between the two groups: STT (p<0.024), iCEBc (p<0.001), P-duration (p<0.001), QRS duration (p<0.001), and PR interval (p<0.001). The HRV parameters MeanNN (p<0.014), MadNN (p<0.032), IQRNN (p<0.009), MeanNN (p<0.029), and HTI (p<0.002) differed significantly between the two groups (see Table 4).

Table 5: Classification performance evaluation of both RHDECG and extended-RHDECG dataset with six classes

Dataset	Approach	Accuracy (%)
RHDECG	InceptionTime (w/o time-frequency)	88.3±0.2
	InceptionTime (time-frequency)	**94.6±0.1**
extended-RHDECG	InceptionTime (w/o time-frequency)	36.1±0.2 (58.7±0.5)
	InceptionTime (time-frequency)	**51.2±0.3 (69.1±0.5)**

The values inside parentheses represent the model results of the top-2 classes.

4.2 Experiments

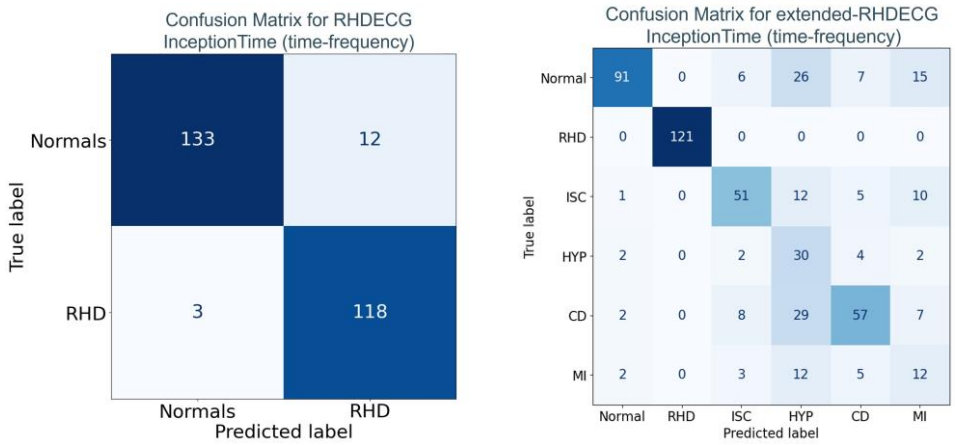
4.2.1 Experiment on the RHDECG dataset

This experiment aimed to classify between RHD patients and healthy subjects via the RHDECG dataset. The performance results of the model using the InceptionTime [57] CNN model alone without fusing time–frequency features and when time–frequency features are combined with CNN features are summarized in Table 5. The confusion matrix results across the five folds are presented in Figure 5a. On a balanced RHDECG with additional randomly selected age-matched normal ECGs from the PTB-XL dataset, the model achieved an accuracy of 88.3±0.2% without fusing time–frequency features. This result improved to an accuracy of 94.6±0.1% when the time–frequency features were fused with the CNN features.

The importance of the time–frequency features evaluated using eXtreme gradient boosting (XGBoost) algorithm is shown in Figure 6. Temporal parameters, including PR, iCEBc, QTc, and TpTe/QTc, and the RWE at all energy bands except above the 62.5Hz band were the most significant features.

4.2.2 Experiment on the extended-RHDECG dataset

In this experiment, the aim was to evaluate the model on an extended-RHDECG dataset that consisted of six classes. The model performance results are shown in Table 5. The results from the combined CNN features and time–frequency features achieved an accuracy of 51.2±0.3%. However, the top-2 accuracy was 69.1±0.5%. In particular, the



(a) (b)

Figure 5. Confusion matrix: (a) Evaluation of the model on RHDECG to classify between RHD and Normal ECGs, (b) Evaluation of the model on extended-RHDECG with six classes

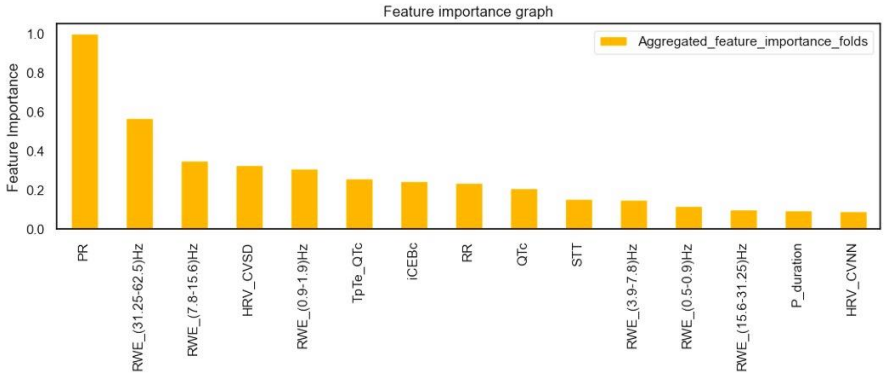


Figure 6. Feature importance graph of the top 15 features evaluated on RHDECG dataset

detection accuracy for PwRHD was above 98% in the top-1 RHD cases but missing none of the RHD cases at top-2 as shown in Figure 5b. However, the lowest performance score was for the MI class (24%), which had the smallest number of patients. When time-frequency features were not combined, the model exhibited a reduced performance, with an accuracy of 36.1±0.2%.

The class activation maps extracted from the last convolutional layer of the model for test recordings from different classes are shown in Figure 7. The colors represent amount of gradient weights used for prediction. On the first top row of the ECG waveform the model focused more on waves and their onsets and offsets but less on the isoelectric line or changes where no significant amplitude changes. On the bottom last ECG waveform, more attention is given to segments where there is an abnormal abrupt change. Thus, the convolutional model captures ECG waveform morphologies and pays

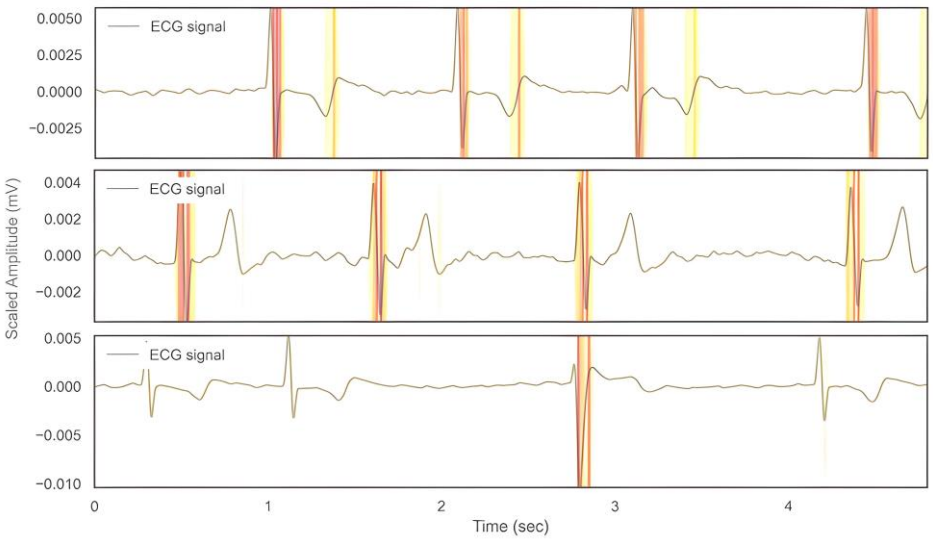


Figure 7. Class activation heatmaps of example test ECGs

The heat map highlights key regions of the ECG where arrhythmic patterns, such as inconsistent beat intervals and waveforms, are present. The intensity of the color indicates the gradient, where the higher the intensity, the greater the importance of the region to the model decision, and vice versa.

greater attention to rhythm changes, fundamental waves and their patterns.

5 Discussion

ECG intervals and rhythm-related characteristics have been utilized as diagnostic markers for assessing the risk of ARF and RHD [24, 36, 58]. In alignment with prior findings, including those reported in our earlier work [27], the present study further substantiates that HRV parameters and wavelet energy features derived from ECG signals exhibit significant differences between individuals diagnosed with RHD and those with normal ECG profiles. Furthermore, our analysis demonstrated that raw 10-second single-lead ECG recordings, when combined with the extracted time–frequency features, can effectively contribute to the classification of patients with progressed RHD. These findings highlight the diagnostic relevance of short-duration ECG recordings and underscore the importance of advanced signal processing techniques, such as wavelet transform-based energy quantification, in enhancing the sensitivity of automated classification systems. This approach may offer a scalable and noninvasive tool for early detection and disease monitoring in resource-limited settings where RHD remains prevalent.

5.1 Multiresolution wavelet energy in RHD ECG

The performance of the wavelet energy features provides additional support for the classifier model. Specifically, when examining the percentage of multiresolution relative wavelet energy contained in higher frequency bands above 15.6Hz, it becomes evident that healthy group tends to have higher values than the PwRHD group. This is because of the fact that majority of the spectral content for normal P-waves and normal QRS complexes lies within this range. However, prolonged intervals, prolonged or no P waves, wide QRS complexes, and bradycardia, which are more common in individuals with RHD than in those with normal sinuses [25–27], also contributed to this difference. Similarly, in the case of the extended-RHDECG dataset, RWE features for diseased patient ECGs (RHD, STTC, HYP, CD, and MI) compared with normal ECGs were significantly different for higher-band wavelet energy features above 7.8Hz as shown in Table 4.

The bandwidth between 2Hz to 4Hz primarily captures the spectral characteristics of the PQRST wave sequence, partially including the normal T-wave spectrum. This band also overlaps with abnormal P-wave bands [45]. A relatively similar wavelet energy in this range for both classes relates to the presence of spectral contents of normal T-waves and prolonged P-waves in PwRHD. This was supported by the presence of a significant mean difference between the two groups for P-duration but no significant mean difference for T- duration, as mentioned in Table 4. Analogously, there was no statistically significant difference between the healthy and PwRHD groups within the 0.5Hz to 1.01Hz band. This might suggest the influence of rhythmic sympathetic activity overlapping in this range along with baseline frequency components [59]. Notably, the highest frequency range, [62.5-100]Hz, contributed smaller compared to the lower frequency bands below 62.5Hz, as illustrated in Figure 6. This is primarily because most diagnostic information is found below 70Hz [45, 59].

5.2 Effect of temporal parameters

About a half of the temporal parameters have statistically significant differences in their means for both classes, primarily MeanNN, IQRNN, and HTI. This is due to the fact that rhythms in PwRHD were mostly normal sinus, with abnormal ECGs observed in only

493 26.5% of this group [27]. As reported in [27], the absence of skipped QRS beats
494 associated with advanced conduction blocks from the AV to bundle branches in PwRHD
495 suggests preserved ventricular conduction, which might contribute to reduced mean
496 differences in HRV parameters between the two groups.

497 In general, the normal range cutoff points for variability in RRi decrease with age and
498 differ by sex [54]. However, findings from large-scale studies on HRV-related parameters
499 [54, 56, 60] in adolescents and young individuals (aged 10 to 29 years) indicated that
500 sex had a greater influence than age did, particularly on the rMSSD and pNN50 measures
501 [54]. As shown in Table 4, the temporal parameters related to its standard deviation,
502 absolute deviation or range of the middle half of RRi as independent variables for
503 statistical comparisons between the normal and diseased patient groups did not show
504 strong evidence except for HTI, pNN20, and pNN50. These parameters reflect the high-
505 frequency variance components of the heart rate that are estimated from short time
506 course measurements [59].

507 Furthermore, prolongation of PR, QTc, P-duration, STT, and elevated iCEBc, which are
508 often observed in subjects with active ARF [25, 26], were also encountered in PwRHD
509 ECGs (see Table 4). The statistically significant differences in these parameters were
510 consistent in both the RHDECG and extended-RHDECG datasets for healthy and diseased
511 patients. The effects of these parameters were crucial during classification, as illustrated
512 in Table 5 and Figure 6.

513 Model performance becomes more robust when frequency and temporal features are
514 integrated, although variabilities in temporal patterns depend on factors such as disease
515 stage and myocardial damage [27]. Additionally, the obtained performance at top-2 was
516 competitive with the results when phonocardiogram signals were used to differentiate
517 PwRHD patients from healthy subjects [6]. This highlights the potential of ECGs as
518 valuable tool for automated RHD detection, particularly in resource-constrained cardiac
519 wards. Overall, the experimental results show the potential benefit of utilizing ECG
520 spectral information and temporal parameters with machine learning for detection of
521 advanced RHD and other prevalent CVDs from normal cases in a simulated RHD-prone
522 cardiac wards settings.

6 Limitations

This study also has some limitations. First, it is worth noting here that the differences in recording devices and disease stages may have different consequences. The RHDECG was acquired using a single-lead wearable sensor, whereas a 12-lead Schiller AG ECG device was used for the PTB-XL dataset. The use of only a single lead from 12-lead recordings of the PTB-XL dataset combined with the RHDECG dataset might be attributed to loose diagnostic information. Nevertheless, this study demonstrated the feasibility of using single lead ECGs in resource-limited cardiac clinics for detecting prevalent CVDs, which may reduce diagnosis burden.

Second, our study cohort may not represent the general population or those with milder forms of the disease. This study was a single-center study obtained from symptomatic cohort who were referred to a major referral hospital. Moreover, it may also lack diversity in terms of ethnicity, socioeconomic status, and environmental factors, which may limit the applicability of the results to broader populations. Nevertheless, the majority of the cases in the population screening among at-risk young age groups may perhaps be at an early stage and asymptomatic. Future work should focus on expanding datasets, improving model features, and incorporating multimodal data to improve model generalizability in asymptomatic cases.

7 Conclusion

This study presented an automated method to distinguish advanced RHD patients from healthy individuals using single-lead ECG signals in resource-limited cardiology ward settings. The CNN model, in combination with the time and frequency features across multiple resolution bands, demonstrated high classification performance. These time–frequency features effectively capture abnormalities associated with RHD, thereby enhancing the generalization capability of the classifier. Furthermore, the interpretability of key predictive features was consistent with that of clinically recognized diagnostic indicators. Therefore, the proposed technique could serve as a potential cost-effective solution for RHD detection, especially in cardiac care environments with constrained resources.

Abbreviations

RHD: Rheumatic Heart Disease

CNN: Convolutional Neural Network

RWE : Relative Wavelet Energy

MeanNN: The mean of the RR intervals

HRV: Heart Rate Variability

CVD: Cardiovascular Diseases

SDNN: The standard deviation of the RR intervals.

RMSSD: The square root of the mean of the squared successive differences between adjacent RR intervals.

SDSD: The standard deviation of the successive differences between RR intervals.

CVNN: The standard deviation of the RR intervals (SDNN) divided by the mean of the RR intervals (MeanNN).

CVSD: The root mean square of successive differences (RMSSD) divided by the mean of the RR intervals (MeanNN).

MedianNN: The median of the RR intervals.

MadNN: The median absolute deviation of the RR intervals.

MCVNN: The median absolute deviation of the RR intervals (MadNN) divided by the median of the RR intervals (MedianNN).

IQRNN: The interquartile range (IQR) of the RR intervals.

Prc20NN: The 20th percentile of the RR intervals. pNN50: The percentage of absolute differences in successive RR intervals greater than 50ms.

pNN20: The percentage of absolute differences in successive RR intervals greater than 20ms.

HTI: The HRV triangular index, measuring the total number of RR intervals divided by the height of the RR intervals histogram.

Declarations

Ethics approval and consent to participate

The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committee of the Department of Internal Medicine, College of Health Sciences, Addis Ababa University (No: 014/2018). Clinical trial number is not applicable. Informed consent was obtained from all subjects involved in the study.

Consent for publication

Written informed consent was obtained from the patient(s) to publish the study results.

Availability of data and materials

The PTB-XL data used for this research is publicly available at https://physionet.org/content/ptb-xl/1.0.3/. The RHDECG dataset analyzed in this manuscript is not publicly available. However, it can be provided upon reasonable request to the corresponding author. The code used for the experiments can be accessed from https://github.com/amsto2/dnnRHD.

Competing interests

The authors have no competing interests to declare.

Authors' contributions

Amsalu Tomas: Performed computer programming, carried out experiments and prepared the first draft manuscript. Melkamu Hunegnaw: Conceptualized the study, collected the dataset, analyze the experimental settings and results, and reviewed the manuscript. Carolina Varon: Designed the methodology, analyzed the results and reviewed the manuscript. Desalew Mekonnen: Frame dataset acquisition, analyzed results, reviewed the draft. Chunzhou Wang: Reviewed the experiments and results, and reviewed the manuscript. Bart Vanrumste: Conceptualized the study, designed methodology and reviewed the manuscript.

Funding

This research was funded by KU Leuven with reference number B/22/032, and Group-T 5E fund with reference number REF23123123 under Leuven center for affordable health technology.

Acknowledgement

The authors are grateful to collaborators of the RHDECG dataset[6] for carrying out the experiments in this research. We also extend our gratitude to the Flanders AI program for supporting this study.

References

[1] Passos, L.S., Nunes, M.C.P., Aikawa, E.: Rheumatic heart valve disease pathophysiology and underlying mechanisms. Frontiers in Cardiovascular Medicine 7, 612716 (2021)

[2] Dougherty, S., Khorsandi, M., Herbst, P.: Rheumatic heart disease screening: current concepts and challenges. Annals of pediatric cardiology 10(1), 39 (2017)

[3] Ou, Z., Yu, D., Liang, Y., Wu, J., He, H., Li, Y., He, W., Gao, Y., Wu, F., Chen, Q.: Global burden of rheumatic heart disease: trends from 1990 to 2019. Arthritis Research & Therapy 24(1), 138 (2022)

[4] Samadzadeh, S., *et al.*: The unfinished agenda of communicable diseases among children and adolescents before the covid-19 pandemic, 1990-2019: A systematic analysis of the global burden of disease study 2019. Lancet 402(10398), 313–335 (2023)

[5] Al-Jazairi, A., Al-Jaser, R., Al-Halees, Z., Shahid, M., Al-Jufan, M., Al-Mayouf, S., Al-Rajhi, A., Al-Hajjar, S.: Guidelines for the secondary prevention of rheumatic heart disease: endorsed by saudi pediatric infectious diseases society (spids). International Journal of Pediatrics and Adolescent Medicine 4(1), 47–50 (2017)

[6] Asmare, M.H., Woldehanna, F., Janssens, L., Vanrumste, B.: Rheumatic heart disease detection using deep learning from spectro-temporal representation of

unsegmented heart sounds. In: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 168–171 (2020)

[7] Rwebembera, J., Marangou, J., Mwita, J.C., Mocumbi, A.O., Mota, C., Okello, E., Nascimento, B., Thorup, L., Beaton, A., Kado, J., *et al.*: 2023 world heart federation guidelines for the echocardiographic diagnosis of rheumatic heart disease. Nature Reviews Cardiology, 1–14 (2023)

[8] Ruan, R., Liu, X., Zhang, Y., Tang, M., He, B., Zhang, Q.-W., Shu, T.: Global, regional, and national advances toward the management of rheumatic heart disease based on the global burden of disease study 2019. Journal of the American Heart Association 12(13), 028921 (2023)

[9] Aliyu, I.A., Bala, J.A., Yusuf, I., Amole, T.G., Musa, B.M., Yahaya, G., Quashie, P.K., Binyet, M., Soon-Shiong, P., Foley, K., *et al.*: Rheumatic heart disease burden in africa and the need to build robust infrastructure. JACC: Advances 3(12 Part 2), 101347 (2024)

[10] Yadeta, D., Hailu, A., Haileamlak, A., Gedlu, E., Guteta, S., Tefera, E., Tigabu, Z., Tesfaye, H., Daniel, W., Mekonnen, D., *et al.*: Prevalence of rheumatic heart disease among school children in ethiopia: A multisite echocardiography-based screening. International journal of cardiology 221, 260–263 (2016)

[11] Gemechu, T., Mahmoud, H., Parry, E.H., Phillips, D.I., Yacoub, M.H.: Community-based prevalence study of rheumatic heart disease in rural Ethiopia. European journal of preventive cardiology 24(7), 717–723 (2017)

[12] Yuyun, M.F., Sliwa, K., Kengne, A.P., Mocumbi, A.O., Bukhman, G.: Cardiovascular diseases in subsaharan Africa compared to high-income countries: an epidemiological perspective. Global heart 15(1), 15 (2020)

[13] Alhuneafat, L., Al Ta'ani, O., Tarawneh, T., ElHamdani, A., Al-Adayleh, R., AlAjlouni, Y., Naser, A., Al-Abdouh, A., Amoetang, R., Taffe, K., *et al.*: Burden of cardiovascular disease in subsaharan africa, 1990-2019: an analysis of the global burden of disease study. Current problems in cardiology, 102557 (2024)

[14] Yadeta, D., Walelgne, W., Fourie, J.M., Scholtz, W., Scarlatescu, O., Nel, G., Gebremichael, M.: Cardiovascular topics Ethiopia country report. CARDIOVASCULAR JOURNAL OF AFRICA 32(1), 37–46 (2021)

[15] Minja, N.W., Nakagaayi, D., Aliku, T., Zhang, W., Ssinabulya, I., Nabaale, J., Amutuhaire, W., Loizaga, S.R., Ndagire, E., Rwebembera, J., *et al.*: Cardiovascular diseases in Africa in the twenty-first century: gaps and priorities going forward. Frontiers in Cardiovascular Medicine 9, 1008335 (2022)

[16] Brown, K., Roshanitabrizi, P., Rwebembera, J., Okello, E., Beaton, A., Linguraru, M.G., Sable, C.A.: Using artificial intelligence for rheumatic heart disease detection by echocardiography: Focus on mitral regurgitation. Journal of the American Heart Association 13(2), 031257 (2024)

[17] Rem´enyi, B., Wilson, N., Steer, A., Ferreira, B., Kado, J., Kumar, K., Lawrenson, J., Maguire, G., Marijon, E., Mirabel, M., *et al.*: World heart federation criteria for echocardiographic diagnosis of rheumatic heart disease—an evidence-based guideline. Nature reviews cardiology 9(5), 297–309 (2012)

[18] Moghaddasi, H., Nourian, S.: Automatic assessment of mitral regurgitation severity based on extensive textural features on 2d echocardiography videos. Computers in biology and medicine 73, 47–55 (2016)

[19] Edwards, L.A., Feng, F., Iqbal, M., Fu, Y., Sanyahumbi, A., Hao, S., McElhinney, D.B., Ling, X.B., Sable, C., Luo, J.: Machine learning for pediatric echocardiographic mitral regurgitation detection. Journal of the American Society of Echocardiography 36(1), 96–104 (2023)

[20] Martins, J.F.B., Nascimento, E.R., Nascimento, B.R., Sable, C.A., Beaton, A.Z., Ribeiro, A.L., Meira Jr, W., Pappa, G.L.: Toward automatic diagnosis of rheumatic heart disease on echocardiographic exams through video-based deep learning. Journal of the American Medical Informatics Association 28(9), 1834–1842 (2021)

[21] Springer, D.: Mobile phone-based rheumatic heart disease detection. PhD thesis, University of Oxford (2015)

[22] WHO: Rheumatic Fever and Rheumatic Heart Disease: Report of a WHO Expert Consultation, Geneva, 29 October-1 November, 2001 vol. 923, p. . World Health Organization, (2004)

[23] Kumar, R.K., Antunes, M.J., Beaton, A., Mirabel, M., Nkomo, V.T., Okello, E., Regmi, P.R., Rem´enyi, B., Sliwa-H¨ahnle, K., Zu¨hlke, L.J., *et al.*: Contemporary diagnosis and

management of rheumatic heart disease: implications for closing the gap: a scientific statement from the american heart association. Circulation 142(20), 337–357 (2020)

[24] Sokolow, M.: Significance of electrocardiographic changes in rheumatic fever. The American Journal of Medicine 5(3), 365–378 (1948)

[25] Alp, H., Baysal, T., Altın, H., Karata¸s, Z., Karaarslan, S.: Qt and p-wave dispersions in rheumatic heart disease: Prospective long-term follow up. Pediatrics International 56(5), 681–688 (2014)

[26] Kohn, L.: Changes in the t wave of the electrocardiogram in acute rheumatic fever. Proceedings of the Society for Experimental Biology and Medicine 31(2), 184–187 (1933)

[27] Asmare, M.H., Chuma, A.T., Varon, C., Woldehanna, F., Janssens, L., Vanrumste, B.: Characterization of rheumatic heart disease from electrocardiogram recordings. Physiological Measurement 44(2), 025002 (2023)

[28] Strodthoff, N., Wagner, P., Schaeffter, T., Samek, W.: Deep learning for ecg analysis: Benchmarks and insights from ptb-xl. IEEE Journal of Biomedical and Health Informatics 25(5), 1519–1528 (2020)

[29] Wang, J., Qiao, X., Liu, C., Wang, X., Liu, Y., Yao, L., Zhang, H.: Automated ecg classification using a nonnon-local convolutional block attention module. Computer methods and programs in biomedicine 203, 106006 (2021)

[30] Wagner, P., Strodthoff, N., Bousseljot, R.-D., Kreiseler, D., Lunze, F.I., Samek, W., Schaeffter, T.: Ptb-xl, a large publicly available electrocardiography dataset. Scientific data 7(1), 154 (2020)

[31] Jiang, A., Huang, C., Cao, Q., Wu, S., Zeng, Z., Chen, K., Zhang, Y., Wang, Y.: Multi-scale cross-restoration framework for electrocardiogram anomaly detection. In: International Conference on Medical Image Computing and ComputerAssisted Intervention, pp. 87–97 (2023)

[32] Reyna, M.A., Sadr, N., Alday, E.A.P., Gu, A., Shah, A.J., Robichaux, C., Rad, A.B., Elola, A., Seyedi, S., Ansari, S., *et al.*: Will two do? varying dimensions in electrocardiography:

the physionet/computing in cardiology challenge 2021. In: 2021 Computing in Cardiology (CinC), vol. 48, pp. 1–4 (2021)

[33] Han, H., Park, S., Min, S., Choi, H.-S., Kim, E., Kim, H., Park, S., Kim, J., Park, J., An, J., *et al.*: Towards high generalization performance on electrocardiogram classification. In: 2021 Computing in Cardiology (CinC), vol. 48, pp. 1–4 (2021)

[34] Jin, Y., Li, Z., Wang, M., Liu, J., Tian, Y., Liu, Y., Wei, X., Zhao, L., Liu, C.: Cardiologist-level interpretable knowledge-fused deep neural network for automatic arrhythmia diagnosis. Communications Medicine 4(1), 31 (2024)

[35] Sun, L.-C., Lee, C.-C., Ke, H.-Y., Wei, C.-Y., Lin, K.-F., Lin, S.-S., Hsiu, H., Chen, P.-N.: Deep learning for the classification of atrial fibrillation using wavelet transform-based visual images. BMC Medical Informatics and Decision Making 22(Suppl 5), 349 (2025)

[36] Karacan, M., Ceviz, N., Olgun, H.: Heart rate variability in children with acute rheumatic fever. Cardiology in the Young 22(3), 285–292 (2012)

[37] S, L.D. Pal czyn´ski K, *et al.*: Ecg signal classification using deep learning techniques based on the ptb-xl dataset. Entropy 23(9), 1121 (2021)

[38] Acar, B., Savelieva, I., Hemingway, H., Malik, M.: Automatic ectopic beat elimination in short-term heart rate variability measurement. Computer methods and programs in biomedicine 63(2), 123–131 (2000)

[39] Kautzner, J., John Camm, A.: Clinical relevance of heart rate variability. Clinical cardiology 20(2), 162–168 (1997)

[40] Beat2Phone ECG sensor, Mobile Application and Cloud Service. https://www.vitalsignum.com/en/beat2phone/. Accessed: 2025-02-28

[41] Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I.: From local explanations to global understanding with explainable ai for trees. Nature machine intelligence 2(1), 56–67 (2020)

[42] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: visual explanations from deep networks via gradient-based localization. International journal of computer vision 128, 336–359 (2020)

[43] Bailey, J.J., Berson, A.S., Garson Jr, A., Horan, L.G., Macfarlane, P.W., Mortara, D.W., Zywietz, C.: 'recommendations for standardization and specifications in automated electrocardiography: bandwidth and digital signal processing. a report for health professionals by an ad hoc writing group of the committee on electrocardiography and cardiac electrophysiology of the council on clinical cardiology, american heart association. Circulation 81(2), 730–739 (1990)

[44] Percival, D.B., Walden, A.T.: Wavelet Methods for Time Series Analysis vol. 4. Cambridge university press, (2000)

[45] Tereshchenko, L.G., Josephson, M.E.: Frequency content and characteristics of ventricular conduction. Journal of electrocardiology 48(6), 933–937 (2015)

[46] Xu, X., Liu, Y.: Adaptive threshold for qrs complex detection based on wavelet transform. In: 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, pp. 7281–7284 (2006)

[47] Tereshchenko, L.G., Waks, J.W., Kabir, M., Ghafoori, E., Shvilkin, A., Josephson, M.E.: Analysis of speed, curvature, planarity and frequency characteristics of heart vector movement to evaluate the electrophysiological substrate associated with ventricular tachycardia. Computers in biology and medicine 65, 150–160 (2015)

[48] P´erez-Riera, A.R., Barbosa-Barros, R., Daminello-Raimundo, R., Abreu, L.C.: Main artifacts in electrocardiography. Annals of Noninvasive Electrocardiology 23(2), 12494 (2018)

[49] Bazett, H.: An analysis of the time-relations of electrocardiograms. Heart 7, 353 (1920)

[50] Lazzerini, P.E., Laghi-Pasini, F., Capecchi, P.L., Boutjdir, M.: Emerging risk factors for qt interval prolongation and torsades de pointes. In: Torsades de Pointes, pp. 113–156. Elsevier, (2022)

[51] Robyns, T., Lu, H.R., Gallacher, D.J., Garweg, C., Ector, J., Willems, R., Janssens, S., Nuyens, D.: Evaluation of index of cardio-electrophysiological balance (iceb) as a new biomarker for the identification of patients at increased arrhythmic risk. Annals of Noninvasive Electrocardiology 21(3), 294–304 (2016)

[52] Makowski, D., Pham, T., Lau, Z.J., Brammer, J.C., Lespinasse, F., Pham, H., Sch¨olzel, C., Chen, S.A.: Neurokit2: A python toolbox for neurophysiological signal processing. Behavior research methods, 1–8 (2021)

[53] Kleiger, R.E., Stein, P.K., Bosner, M.S., Rottman, J.N.: Time domain measurements of heart rate variability. Cardiology clinics 10(3), 487–498 (1992)

[54] Umetani, K., Singer, D.H., McCraty, R., Atkinson, M.: Twenty-four hour time domain heart rate variability and heart rate: relations to age and gender over nine decades. Journal of the American College of Cardiology 31(3), 593–601 (1998)

[55] Park, M.K., Salamat, M.: Park's Pediatric Cardiology for Practitioners: South Asia Edition-E-Book vol. 7. Elsevier Health Sciences, (2020)

[56] Palhares, D.M., Marcolino, M.S., Santos, T.M., Silva, J.L., Gomes, P.R., Ribeiro, L.B., Macfarlane, P.W., Ribeiro, A.L.: Normal limits of the electrocardiogram derived from a large database of brazilian primary care patients. BMC Cardiovascular Disorders 17, 1–23 (2017)

[57] Ismail Fawaz, H., Lucas, B., Forestier, G., Pelletier, C., Schmidt, D.F., Weber, J., Webb, G.I., Idoumghar, L., Muller, P.-A., Petitjean, F.: Inceptiontime: Finding alexnet for time series classification. Data Mining and Knowledge Discovery 34(6), 1936–1962 (2020)

[58] Nishimura, R.A., Otto, C.M., Bonow, R.O., Carabello, B.A., Erwin III, J.P., Guyton, R.A., O'Gara, P.T., Ruiz, C.E., Skubas, N.J., Sorajja, P., *et al.*: 2014 aha/acc guideline for the management of patients with valvular heart disease: a report of the american college of cardiology/american heart association task force on practice guidelines. Circulation 129(23), 521–643 (2014)

[59] Electrophysiology, T.F.o.t.E.S.o.C.t.N.A.S.o.P.: Heart rate variability: standards of measurement, physiological interpretation, and clinical use. Circulation 93(5), 1043–1065 (1996)

Jensen-Urstad, K., Storck, N., Bouvier, F., Ericson, M., Lindbland, L., Jensen-Urstad, M.: Heart rate variability in healthy subjects is related to age and gender. Acta Physiologica Scandinavica 160(3), 235–241 (1997)