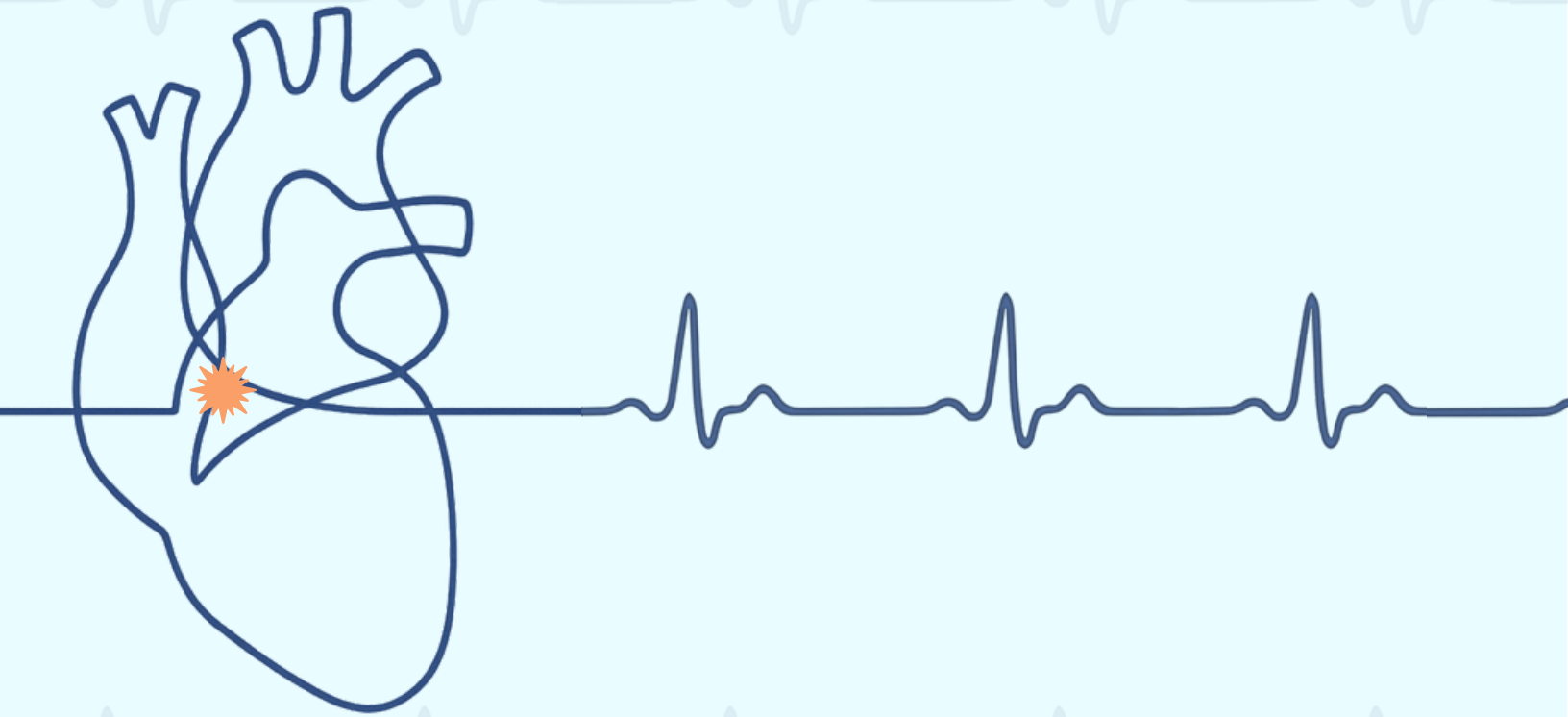
THE DETECTION OF POSTOPERATIVE JUNCTIONAL ECTOPIC TACHYCARDIA IN THE ELECTROCARDIOGRAM AT THE PAEDIATRIC INTENSIVE CARE UNIT



TECHNICAL MEDICINE
MASTER'S THESIS

JANNA MEIJER

The Detection of Postoperative Junctional Ectopic Tachycardia in the Electrocardiogram at the Paediatric Intensive Care Unit

Janna Meijer
Student number: 4842278
Date: 22 July 2025

Thesis in partial fulfilment of the requirements for the joint degree of Master of Science in

Technical Medicine

Leiden University ; Delft University of Technology ; Erasmus University Rotterdam

Master thesis project (TM30004; 35 ECTS)
Dept. of Paediatric Intensive Care Unit,
Erasmus MC Sophia Children's Hospital
24 February 2025 – 8 August 2025

Supervisors:
dr. ir. Nico Bruining
dr. Marc Sylva
Eris van Twist
Brian van Winden

Thesis committee members:
dr. ir. Nico Bruining, Erasmus MC
dr. Marc Sylva, Erasmus MC
dr. ir. Enno van der Velde, LUMC
dr. Jelle Man, Erasmus MC

Universiteit Leiden

TU Delft Delft University of Technology

Erasmus Universiteit Rotterdam

Preface

This thesis marks the end of my time as a student, a journey that spanned the past seven years and shaped me both academically and personally. From the very beginning, I enjoyed studying Technical Medicine. The bachelor programme provided us with a broad foundation of knowledge, but still left me with a level of uncertainty on how I could translate this into real-world impact. However, during the master's degree, through hands-on internships and close collaborations within hospital departments, I began to realise what our studies had been worth. Working on real projects, alongside clinicians, researchers, and fellow students, was not only inspiring but also helped me build my confidence and develop my skills. These experiences affirmed my enthusiasm for contributing to healthcare innovation.

For my graduate internship, I was looking for a setting where I could apply my 'Signals & Stimulation' expertise and where I would be surrounded by ambitious, engaged supervisors. I found exactly that at the Paediatric Intensive Care Unit of the Erasmus MC Sophia Children's Hospital. The team's energy and enthusiasm were contagious from day one.

I would like to express my appreciation to my supervisors, Marc Sylva, Nico Bruining, Eris van Twist and Brian van Winden. Marc, thank you for your enthusiasm, your honest feedback, and sharing your inexhaustible level of knowledge with me. I really appreciated our meetings, which always left me energised (and well caffeinated) to move forward. Nico, thank you for finding time for me despite your full schedule, and for your critical mindset. Eris and Brian, thank you for always having the door open, your creative and critical minds, and your willingness to brainstorm and think along. I always left your office with a boost of new motivation.

To my family and friends: thank you for all your support, not just during this final chapter, but throughout my entire study period. A special thanks to my roommates for always being there to listen, share struggles, and offer distraction when needed. And to my study friends: working on our theses side-by-side, sharing doubts, questions, and plenty of coffee breaks, made this period much more fun.

All in all, I look back on this time with a smile and I am excited for what lies ahead!

Janna Meijer,
Rotterdam, July 2025

Abstract

Background and objectives: Junctional Ectopic Tachycardia (JET) is a tachyarrhythmia most commonly observed in infants and children in the postoperative setting. An automatic detection algorithm could be valuable for early identification and timely treatment of JET. However, the detection is challenging as the initial changes on the electrocardiogram (ECG) are often subtle and monitor data commonly contains substantial noise and artefacts. The objective of this study was to investigate which features contribute to accurate JET detection and to develop an automated detection model.

Methods: A retrospective study was conducted using monitor ECG data of paediatric patients admitted to the Paediatric Intensive Care Unit. The training set consisted of 17 patients, and the test set of 8 patients. ECG metrics were detected, in order to segment the signal and to derive several features. The two-dimensional vectorcardiogram (VCG) was computed for calculating features representing the beat-to-beat variability of the signal. Automatic feature selection methods were applied to identify which features most effectively differentiate JET from sinus rhythm (SR), based on balanced accuracy. Logistic regression (LR) and random forest (RF) models were finally created and performance was validated.

Results: The LR and RF models achieved balanced accuracy scores of 0,989 and 0,988, respectively, on the training dataset. The selected features included the number of P waves and the variance of the PQ interval. For the RF model, the standard deviation (SD) of the RR interval was also selected. VCG-features did not prove effective in distinguishing JET from SR. A secondary validation on the test set yielded lower scores of 0,899 and 0,892. An analysis of misclassifications revealed that they were all attributed to errors in peak detection, which occurred in cases of deviating ECG morphologies or the presence of artefacts and noise.

Conclusions: This study demonstrates that P wave-related features are most effective for distinguishing JET from SR, with simple machine learning models based on these features showing promising results for automated JET detection. Peak detection is currently the most important limiting factor for the robustness and generalisability of this method. Interpatient variability and the low quality of monitor ECG data remain important challenges. Expanding the dataset, improving the data quality and implementing signal quality assessment (SQA) methods are recommended to improve the robustness of the models.

Contents

List of Abbreviations

AUROC	Area Under the Receiver Operating Characteristic Curve
AV	Atrioventricular
BBB	Bundle Branch Block
CHD	Congenital Heart Disease
CNN	Convolutional Neural Network
DL	Deep Learning
DTW	Dynamic Time Warping
EAT	Ectopic atrial rhythm
ECG	Electrocardiogram
FNR	False Negative Rate
FPR	False Positive Rate
HLHS	Hypoplastic Left Heart Syndrome
HR	Heart Rate
IQR	Interquartile Range
JET	Junctional Ectopic Tachycardia
LIME	Local Interpretable Model-Agnostic Explanations
LR	Logistic Regression
MI	Mutual Information
PICU	Paediatric Intensive Care Unit
PVC	Premature Ventricular Contraction
Q1	First quartile (25th percentile)
Q3	Third quartile (75th percentile)
RF	Random Forest
RFE	Recursive Feature Elimination
SD	Standard Deviation
SQA	Signal Quality Assessment
SR	Sinus Rhythm
VA	Ventriculoatrial
VCG	Vectorcardiography

1 Introduction

Congenital heart disease (CHD) represents the largest subgroup of congenital anomalies, with a global incidence of 8 per 1,000 live births [1]. The CHD group includes a wide spectrum of diseases ranging from minor defects to severe cardiac malformations with heart failure and/or severe cyanosis immediately after birth [2]. Approximately 25% of children born with CHD require life-saving interventions such as heart surgery in the first year of life [3]. Arrhythmias after paediatric cardiac surgery are common and can be life-threatening [4]. They occur intraoperatively or may appear shortly after surgery.

One such arrhythmia is Junctional Ectopic Tachycardia (JET). Postoperative JET, occurring within 72 hours after surgery, is observed in 1-15% of children following correction of congenital cardiac defects [5]. A congenital form of JET also exists, typically occurring in the first six months after birth, but this form is rare [6].

JET originates in the region of the atrioventricular (AV) node or proximal His bundle [7]. Postoperative JET has been attributed to ischemia, stretching, and direct injury to the AV conduction tissue of the heart during surgical repair of these defects [6]. JET usually exhibits a gradual onset ('warm-up' phase) [8]. It is often a narrow complex tachycardia but may present as a wide complex tachycardia as a result of aberrant conduction. It typically involves AV dissociation, where the atrial rate is slower than the ventricular rate. In some cases, it may also present with 1:1 retrograde ventriculoatrial (VA) conduction [9].

Postoperative JET is a self-limiting disorder that usually resolves within one week [10]. However, due to the rapid heart rate (HR) that impairs adequate ventricular filling and the loss of AV synchrony, JET can lead to an acute impairment of cardiac output and haemodynamic instability [11]. JET is associated with extended intensive care stay and increased risk of morbidity and mortality, with a reported 30-day mortality rate nearly 9 times higher among paediatric intensive care unit (PICU) patients who develop JET [12, 13]. Several strategies, including cooling, sedation, electrolyte correction, reducing catecholamine doses, and administering antiarrhythmic medications, can be used to treat JET and lower the ventricular rate [7, 11]. Early treatment has been shown to improve patient outcomes, underscoring the importance of early detection [11]. Furthermore, the ability to retrospectively identify periods of JET would be valuable for research into the effectiveness of various treatment strategies. However, early recognition of JET remains challenging because the initial changes on the electrocardiogram (ECG) are often subtle, and continuous monitoring of the ECG by healthcare staff of the PICU is not feasible. Automatic detection of JET using ECG signals could offer a solution.

The ECG may reveal certain features like a high ventricular rate and the disappearance of the P-wave or the appearance of retrograde P-waves, which could be used in a detection model. However, reliable identification is challenging due to several factors: the high frequency of tachycardia, noisy monitoring data, low P-wave amplitudes, and cases where the P-wave is obscured by the QRS complex. In two studies on JET detection, the models were mainly focused on P-wave features [14, 15]. It was shown that particularly high frequency tachycardias, interpatient variability, P waves with low prominences and abnormal morphology caused classification errors.

For detecting arrhythmias, the temporal evolution of the ECG signal is also relevant. Incorporating not just isolated beats but also the preceding rhythm could offer valuable diagnostic information. Patterns such as signal consistency, gradual transitions, or sudden changes over time may all hold predictive value.

Additionally, leveraging information from multiple leads instead of only using one ECG lead could enhance the robustness of detection models. Combining the data of multiple leads can provide two-dimensional (2D) or three-dimensional (3D) representations of the heart's electrical activity which may provide additional insights. Vectorcardiography (VCG) is a technique in which three orthogonal leads are constructed, representing the right-left axis (Vx), head-to-feat axis (Vy), and anteroposterior axis (Vz), after which the vector magnitude in several planes can be computed according to the theorem of Pythagoras [16]. VCG loops can be visualised in the frontal (xy), sagittal (yz) and transverse (xz) plane. Furthermore, the 3D loop can be constructed. A literature review has been conducted to look into the different applications of VCG on the detection, classification and characterization of cardiac arrhythmias [17]. It revealed that there are many parameters available for VCG analysis of cardiac arrhythmias. For a real-time JET detection model, different parameters could be considered to quantify loop orientation and beat-to-beat variability.

Previous research conducted at the PICU of Sophia Children's Hospital explored automatic JET detection using a manually constructed decision tree based on ECG and 2D VCG features [18]. The model achieved high accuracy (87,8%) but relied on accurate P wave detection, an element that proved unreliable, and required a sinus rhythm (SR) reference fragment for comparison. This dependency limits clinical applicability, especially in cases where no post-operative SR is available. Moreover, the model was only evaluated on the small development dataset, raising concerns about overfitting and generalisability.

1.1 Goals and objectives

Several significant challenges remain in achieving accurate JET detection. These include high-frequency tachycardias, low P wave prominences, interpatient variability leading to diverse ECG morphologies, and noisy monitoring data. In addition, JET can manifest in multiple forms, with varying types of atrioventricular conduction patterns, which further complicates the development of a robust detection model.

The goal of this thesis is to identify ECG and VCG features that reliably distinguish JET from SR, and to develop an accurate, automated detection model based on monitor ECG data.

After computing different ECG and VCG features and exploring their potential for distinguishing JET from SR, feature-based classification models will be developed and trained on a labelled dataset. These models will then be validated on longer, more heterogeneous datasets, which may include varying morphologies, artefacts, noise and mixed rhythms, to evaluate their applicability in real-world scenarios. This evaluation may reveal limitations or highlight cases where the model underperforms, guiding future refinements and iterations.

2 Methods

This single-centre retrospective study was conducted at the Erasmus MC Sophia Children's Hospital, utilising monitor ECG data from children with CHD admitted to the PICU between 2018 and 2025. A waiver for ethical approval was obtained for data collection using standard of care bedside monitoring (MEC-2021-0937). Figure 1 provides an overview of the workflow of this project, including the different analysis steps, which will be elaborated on in the following subsections. Data analysis was performed using Python version 3.12.

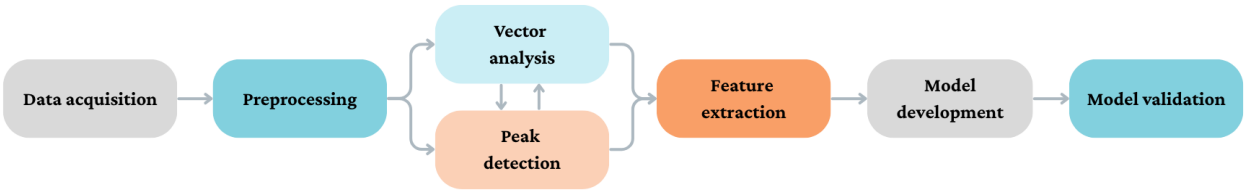


Figure 1: Overview of project workflow.

2.1 Data acquisition

ECG was continuously recorded with at least 3 electrodes (ECG, 3M St. Paul, Minnesota, United States), sampled at 200 Hz and stored on a digital server (Dräger, Lübeck, Germany). For the initial training dataset, 60-seconds fragments were selected from 17 patients, and labelled as either SR or JET by a paediatric intensivist. For each patient, at least one SR and one JET fragment was selected. In addition to the primary rhythm label, all JET fragments received an additional subtype label based on the observed atrioventricular relationship: AV dissociation, visible retrograde P wave, or unclear AV relation. Data fragments were only selected when at least lead I, II and III were saved. Data fragments with paced rhythms, and data containing movement artifacts or arrhythmias other than JET were excluded from the analysis.

For a secondary validation, additional data was collected from eight patients, four of whom had experienced JET and four of whom had not. For each patient, two data fragments of one hour were included. Only data containing at least leads I, II and III and from periods of time when patients' rhythms were not paced are included.

2.2 Preprocessing

Initially, all signal analysis steps are performed using 60-seconds windows. Additionally, shorter window sizes of 30, 20, and 10 seconds are evaluated by dividing the original fragments into smaller segments.

The data processing and feature derivation steps from raw signal to final features are schematically presented in the flowchart in Figure 3. First, leads aVL, aVR and aVF are derived from leads I, II and III using the following equations:

$$aVL = \frac{1}{2}\,(I - III) \tag{1}$$

$$aVR = -\frac{1}{2}(I + II) \tag{2}$$

$$aVF = \frac{1}{2}(II + III) \tag{3}$$

The directions of these leads are visualised in figure 2. To remove baseline wander resulting from low frequency signals, such as respiration or body movement, and to filter out high frequency noise, a third order Butterworth bandpass filter with cut-off frequencies of 0.8 and 80 Hz is applied [19].

For VCG analysis, a vector is created by combining multiple leads. As leads V1 to V6 are not available in our dataset, vector analysis is limited to 2D representations of the heart's electrical activity, in the frontal

plane. Leads I and aVF are used to create this vector over time. The length $|\vec{v}|$ and angle α of this vector are calculated with the following equations:

$$|\vec{v}| = \sqrt{v_{\mathrm{I}}^2 + v_{\mathrm{aVF}}^2} \tag{4}$$

$$\alpha = \deg\left(\arctan 2(v_{\mathrm{aVF}}, v_{\mathrm{I}})\right) \tag{5}$$

where v_{I} and v_{aVF} are the components of the vector along the horizontal (lead I) and vertical (lead aVF) axes respectively. α gives the angle in degrees, where the direction of lead I is 0 degrees, as illustrated in figure 2.

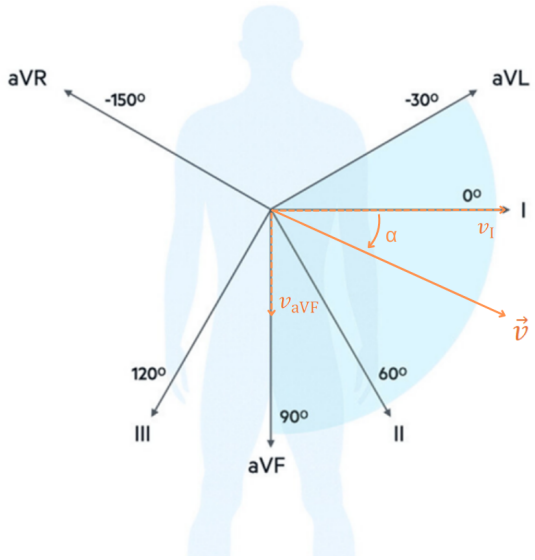


Figure 2: Direction of the ECG leads I, II, III, aVL, aVR and aVF. In orange, the vector created by combining leads I and aVF is illustrated.

2.3 Peak detection

After preprocessing, peak detection is performed to enable both the direct calculation of several time-based features and the segmentation of the signal into the QRS complex and the intervals between, needed for VCG features. Altogether, peak detection includes the R peak, Q peak, S peak, the start and end of the QRS complex, the T wave and the P wave. Existing methods from previous work, of which a summary is provided in Appendix A.1, act as a starting point and were refined iteratively based on observed detection errors [18].

R peaks

R peaks are detected first, as these typically have the highest amplitude. Since the direction of ventricular depolarization may not align perfectly with any of the standard ECG leads, the vector (Eq. 4) is used for optimal accuracy. By identifying the time point at which the vector reaches its maximum value, a new, fictive lead is created in this direction, by projecting the vector signal onto that direction, using the following equation:

$$comp_a v = \frac{a \cdot v}{|a|} \tag{6}$$

where a is the vector at the point of maximum magnitude, v is the vector at each time point, and $|a|$ is the length of the maximum vector. If the highest amplitude of this fictive lead is in the negative direction, the signal is inverted to ensure that R peaks are positive. Peaks are then detected in this fictive lead. The characteristic steep signal slope of R peaks is used to distinguish them from T and P waves. The mean of the smoothed derivative of the signal surrounding each peak (25 ms before and 25 ms after) is computed

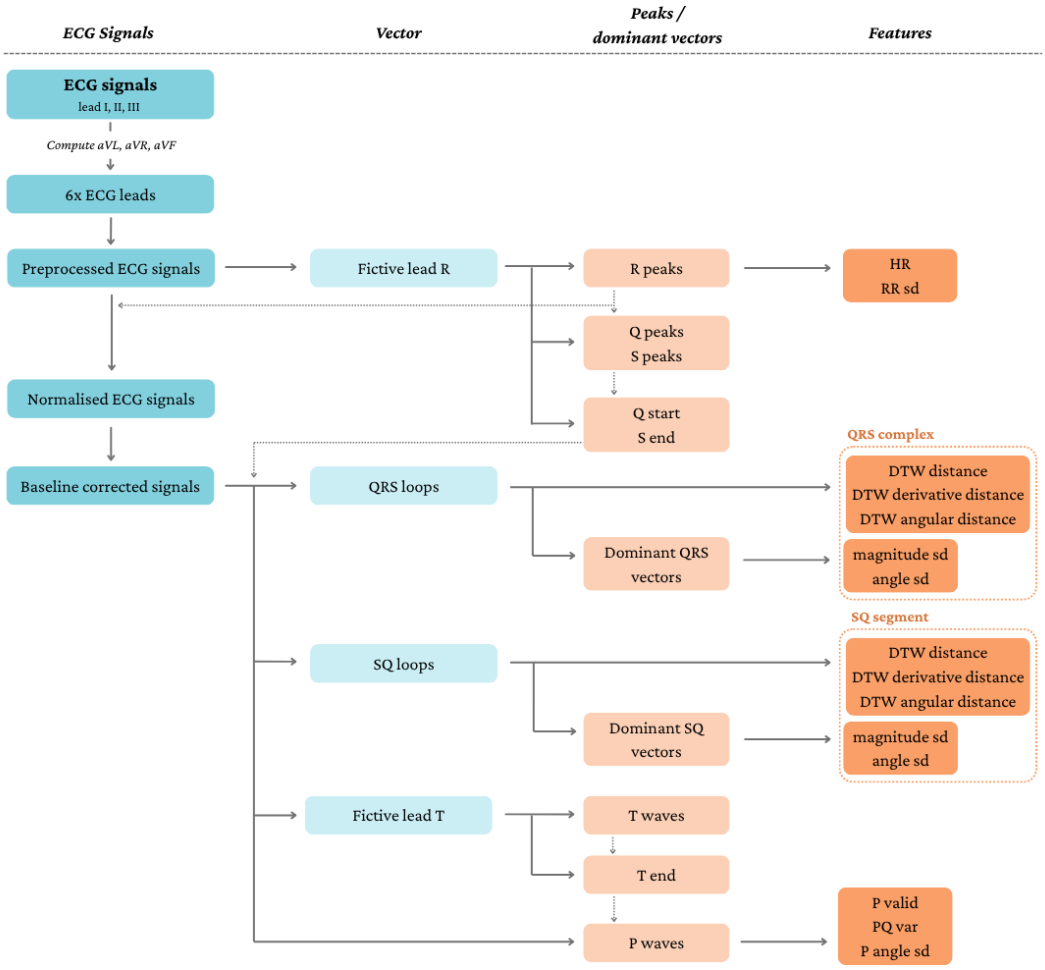


Figure 3: Flowchart of the data processing steps. The columns from left to right show the ECG signals, the derived vector signals, the detected peaks and the extracted features.

and visualised in a histogram. Assuming a bimodal distribution, with one mode representing R peaks and the other non-R peaks, the two highest histogram peaks are identified. The local minimum between them is chosen as the slope threshold. Final peak detection is performed on the fictive lead using a minimum distance of 150 ms and a minimum height of the 90th percentile of the signal. Only peaks with surrounding derivative segments exceeding the slope threshold are retained.

After the detection of R peaks, all ECG signals, including the fictive lead, are normalised by dividing them by the mean amplitude of the R peaks.

Q and S peaks

For each R peak, the corresponding Q and S peaks are searched within a time window of 100 ms before and after the R peak, respectively. While these peaks typically occur close to the R peak, conduction abnormalities, such as bundle branch blocks (BBB), require a broader search window. As the Q and S peak are expected to be negative relative to the R peak, peaks are searched for in the inverted fictive lead. Similar to R peaks, the Q and S points are characterised by steep signal slopes. Q candidates must be followed by a positive derivative segment (25 ms) with a mean >0.1, and S candidates must be preceded by a similarly steep segment. If multiple candidates meet these criteria, the lowest point is selected. If no valid peaks are found, such as in signals lacking a clear negative Q or S deflection, fallback rules apply: the Q peak is assigned to the most negative point within the search window, and the S peak is defined at the first point after the R peak where the derivative drops below 0.1, to avoid misclassifying inverted T waves.

The onset (Q start) and offset (S end) of the QRS complex are also determined. This involves computing the absolute derivative of the fictive lead, smoothed with a 7-sample (35 ms) moving average filter. Q start is defined as the point before the Q peak where the derivative reaches a minimum value of 0.05, and S end is defined as the first point after the S peak where it first drops below 0.02. These thresholds were chosen empirically.

T waves

Before detecting T waves, baseline correction is applied using a baseline fitting method (snip) from the pybaselines library, in order to centre the baseline around zero [20]. The QRS complexes are then deleted by setting the signal to zero between the previously identified Q start and S end points, to isolate the remaining parts of the signal, primarily containing T and P waves. From this modified signal, a new fictive lead is created (using Eq. 6), which theoretically aligns with the maximum amplitude of the T wave. This lead is then smoothed using a Gaussian filter (with σ=2) to reduce noise. Next, peaks are detected in the intervals between consecutive QRS complexes. All peaks are initially identified, the two peaks with the highest prominence are selected, and the first one (in time) is labelled as T wave.

P waves

First, a search window is determined for P wave detection. After each T wave, the first zero crossing of the second derivative (i.e., the inflection point) is identified in the fictive T wave lead, which is first smoothed with a Gaussian filter (with σ=1). The resulting point is used as an estimate for the T wave end. P waves are detected between this point and the subsequent Q start.

While lead II typically provides the clearest P wave morphology, inter-patient variability, especially in patients with congenital cardiac defects, necessitates a more flexible approach. Therefore, P waves are detected separately in leads I, II and III. Peaks are identified with a minimum prominence of 0.02 and the peak with the highest prominence is selected. If no peak is found, a NaN value is added for that beat. To automatically determine which lead provides the best P wave detection, two metrics are computed per lead: the ratio of valid P waves (relative to the number of Q peaks), and the standard deviation (SD) of P wave angles. In case of correct P wave detection, the variation in P wave angles within this patient group is expected to be low, reflecting a consistent direction of atrial depolarisation. A combined score is computed for each lead as follows:

$$p\_score = p\_valid - 0.1 * p\_angle\_sd \tag{7}$$

with p_valid the fraction of detected P waves (from 0 to 1), and p_angle_sd the SD of the P wave angles (Eq. 5). The lead with the highest p_score is selected, and its corresponding P peaks are used for further analysis.

2.4 Feature extraction

After peak detection, the signal is segmented to extract vector loops, and features are calculated. Table 6 in Appendix A.3 provides an overview of all features with brief descriptions.

2.4.1 Time domain features

Using the detected R peaks, HR is calculated using the following equation:

$$HR = \frac{n_{r\_peaks}}{t_{r_n} - t_{r_0}} * 60 \tag{8}$$

where n_{r_peaks} is the number of R peaks and t_{r_0} and t_{r_n} the time points of the first and last R peak, respectively. From all consecutive RR intervals, the standard deviation (RR_sd) is then calculated as a measure of variability.

Additionally, features are defined to characterise the presence and timing of P waves relative to the QRS complex. The proportion of QRS complexes that are preceded by a detected P wave is given by:

$$p\_valid = \frac{n_{p\_waves}}{n_{q\_peaks}} \tag{9}$$

with n_{p_waves} and n_{q_peaks} the number of P waves and Q peaks, respectively. A value of 1 indicates each QRS complex is preceded by a P wave within the defined search area. For all detected P waves, the interval between the P and Q peaks is calculated, and the variance of these intervals (pq_var) is used as an additional feature to capture irregularity in atrioventricular timing. Finally, the SD of the angles (Eq. 5) of all P waves is computed and saved as a feature (p_ang_sd).

2.4.2 VCG features

Several VCG-based features are calculated with the aim to analyse the beat-to-beat variability. The previously identified Q start and S end points are used to segment the signal into two parts: the QRS complexes (QRS loops) and the signals in between (SQ loops).

The vector lengths and angles are calculated as previously defined (Eqs. 4 and 5). For each QRS loop and for each SQ loop, the point in time where the vector has the highest magnitude is identified, and both the corresponding magnitude and the angle are recorded. Subsequently, the SD of these values across all beats is computed, yielding two features: $magnitude\_sd$ and $angle\_sd$. These reflect the variability in magnitude and orientation of the dominant vector of each segment. In the case of QRS loops, this dominant vector typically corresponds to the R peak and represents the cardiac axis. For SQ loops, the dominant vector is expected to reflect the T wave. However, in cases of repolarization abnormalities, the P wave may become dominant instead. Additionally, if the P wave is superimposed on the T wave, this overlap may also influence the characteristics of the dominant vector.

Next to analysing the dominant vector, a more detailed analysis of the morphology of the beats is done to capture more subtle differences. For both QRS loops and SQ loops, beats across time are compared to each other using multiple methods. Variability can be quantified by computing distance or similarity metrics between loops, typically done on a sample-by-sample basis, necessitating uniform sampling across beats [17]. However, beat length variability, which is expected in JET but also to a certain extent in SR, may complicate this, as not all beats will contain the same number of samples. Resampling could distort the original loop shape, which could result in misleading results of distance and similarity measures. To address this, Dynamic Time Warping (DTW) is used. DTW allows for the alignment of similar shapes that may be out of phase in the time axis, by applying non-linear temporal warping [21]. Figure 4 provides an illustration of DTW. This technique is applied to compare the morphology of both QRS loops and SQ loops across time.

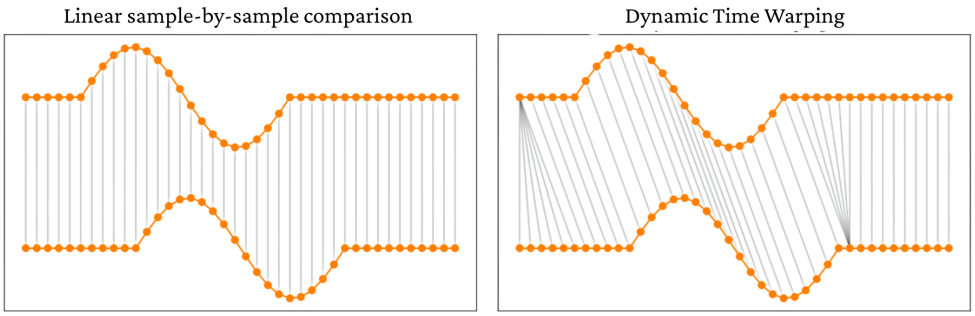


Figure 4: Illustration of linear, sample-by-sample comparison of two signals on the left, and Dynamic Time Warping on the right [22].

As changes during JET can occur gradually, especially during the warm-up phase of this arrhythmia, comparing only consecutive loops might not be sufficient to capture these changes. Instead, a set of loops is selected at regular intervals within each time window. All pairwise combinations of these loops are compared, and the results are averaged to obtain a final distance score. The number of selected loops per window is adapted to the window size: 30 loops for 60 seconds, 15 for 30 seconds, 10 for 20 seconds, and 5 for 10 seconds.

Using DTW, each pair of loops is aligned by minimising the total cumulative distance between points. The time difference between the two loops is used as input parameter to define the maximum allowed time shift for DTW. In order to perform DTW, a distance metric must be defined to quantify the similarity between two samples (i.e., time points) from different loops. In this study, three variations of distance measures are

used: Euclidean distance, derivative-based Euclidean distance, and angular distance. The Euclidean distance is calculated between the 2D coordinates of two vector points, from the frontal plane representation of the VCG, at a given time point:

$$d_{eucl} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{10}$$

where (x_1, y_1) and (x_2, y_2) represent the spatial positions of the two samples being compared. This distance reflects the spatial difference between two loops at each time point, capturing differences in loop shape and position.

Secondly, to emphasise local signal dynamics, a second approach, referred to as the derivative DTW, is applied. In this variation, each loop is first transformed into a sequence of 2D slope vectors by calculating the difference between consecutive points. The Euclidean distance formula is then applied to these derivative vectors. This method captures the beat-to-beat changes in trajectory and direction within the VCG loop.

A third distance measure, the angular distance, quantifies the directional difference between two vectors $\vec{v}_1$ and $\vec{v}_2$ by computing the angle between them:

$$\theta = \arccos\left(\frac{\vec{v}_1 \cdot \vec{v}_2}{|\vec{v}_1| \cdot |\vec{v}_2|}\right) \tag{11}$$

This captures how similarly oriented the vectors are at each time point, independent of their magnitude.

For each pairwise loop comparison, the total distance after alignment is normalised by the alignment path length (the number of links between samples, similar to the number of grey connection lines in Figure 4). The mean distance across all loop combinations is then saved as a feature. Additionally, the top 10% of local (pointwise) distances, representing the most dissimilar regions, are identified and averaged, and saved as separate features to highlight high-variance segments within the loops.

2.5 Feature analysis

An initial exploratory analysis is conducted to examine feature distributions and relationships, using the labelled JET subtypes to explore how patterns vary across different presentations of the arrhythmia.

Mutual information (MI) scores are used to assess how much each feature reduces uncertainty about the classification label (JET or SR). In parallel, a Pearson correlation matrix is computed to identify features that are strongly correlated with one another. If any pair of features exhibit a high correlation (Pearson correlation coefficient $>$0.7), the feature with the lower MI score is excluded to reduce redundancy.

2.6 Model development

Two different machine learning models are used: Logistic Regression (LR), a simple approach for benchmarking, and Random Forest (RF), which is better suited to handle non-linear relationships and capture more complex patterns in the data. Models are created for different window sizes of 60, 30, 20 and 10 seconds.

First, feature selection is performed with use of nested cross-validation, with five outer loops and three inner loops (see Figure 5). To prevent data leakage and maintain class balance, folds were stratified and grouped at the patient level. For LR, recursive feature elimination (RFE) is applied to identify the optimal subset of features by recursively removing the least important ones based on model coefficients. For the RF model, feature selection is performed using $SelectFromModel$, retaining features with the highest impurity-based importance scores as determined during tree construction. Within each outer fold, hyperparameter tuning, including the number of selected features, is performed using a grid search. The best model from the inner loop is then evaluated on the outer fold. Tables 7 and 8 in Appendix A.6 list the tuned parameter values. To assess feature stability, the selected features are tracked across outer folds. Only features selected in at least four out of five folds, i.e. 'stable features', are retained in the final model.

Using only these stable features, a final grid search was conducted on the entire training data using 5-fold stratified group cross-validation to determine the optimal hyperparameters. The final LR and RF models were then trained on the full training set using the selected features and hyperparameters.

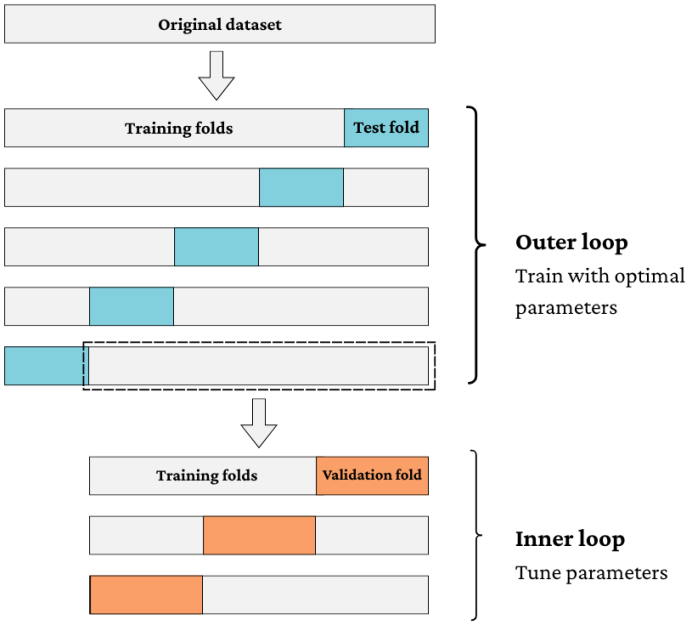


Figure 5: Nested cross-validation with a 5-fold split in the outer loop and 3 folds in the inner loop.

2.7 Model Validation

Following model development and internal evaluation on the training set, the final LR and RF models are further validated using an external validation set comprising new patients not involved in model development. This secondary dataset contains two one-hour ECG recordings from eight additional patients, and may include signals with different morphologies, mixed rhythms, or artefacts, providing a more realistic representation of clinical conditions. Model predictions are validated by reviewing one-minute data fragments (each corresponding to a single prediction) sampled at five-minute intervals, which is performed by a paediatric intensivist. Given the limited dataset size, the resulting metrics are considered indicative rather than fully representative. Misclassified cases are further examined to understand the rationale behind the model's predictions. This secondary validation is particularly aimed at identifying areas in need of refinement.

2.8 Statistics and evaluation metrics

Continuous variables are reported as median values with first and third quartiles (Q1–Q3). Categorical variables are summarised as counts with corresponding percentages.

Model evaluation is based on the following performance metrics: balanced accuracy, area under the receiver operating characteristic curve (AUROC), false positive rate (FPR) and false negative rate (FNR). Balanced accuracy is used as the primary optimisation metric, as it accounts for class imbalance in the dataset. For the nested cross-validation, the mean and SD of the balanced accuracy scores are reported. For the evaluation of the final models with only the stable features on the training set, the mean performance metrics across the cross-validation folds are reported.

3 Results

3.1 Research population

A total of 17 patients were included in the initial training dataset, and 8 patients were included in the secondary, validation dataset. Table 1 summarises these patients' characteristics. The training dataset comprises 53 fragments of JET and 30 fragments of SR. From the validation dataset, 137 were labelled as SR, 47 as JET and 8 as other or mixed rhythms.

Table 1: Characteristics of included research population.

Characteristic	Training dataset (N=17)	Validation dataset (N=8)
Gender, N (%)		
Male	15 (88)	5 (63)
Female	2 (12)	3 (37)
Age (days), median (Q1 - Q3)	119 (44 - 163)	83 (34 - 131)
Diagnosis, N (%)		
ToF	6 (35)	1 (13)
TGA	2 (12)	1 (13)
VSD	6 (35)	1 (13)
HLHS	1 (6)	0 (0)
Borderline left heart	1 (6)	0 (0)
Truncus arteriosus	0 (0)	1 (13)
Aortic valve stenosis	0 (0)	2 (25)
Double aortic arch	0 (0)	1 (13)
Ebstein anomaly	0 (0)	1 (13)

Q1 = first quartile, Q3 = third quartile, ToF = Tetralogy of Fallot, TGA = Transposition of the Great Arteries, VSD = Ventricular Septal Defect, HLHS = Hypoplastic Left Heart Syndrome.

3.2 Peak detection methods

Figures 6, 7 and 8 present examples of the fictive lead R, fictive lead T, and the selected lead for P wave detection, respectively, with the corresponding detected peaks annotated. Additional examples, including detection errors from earlier iterations, are provided in Appendix A.2.

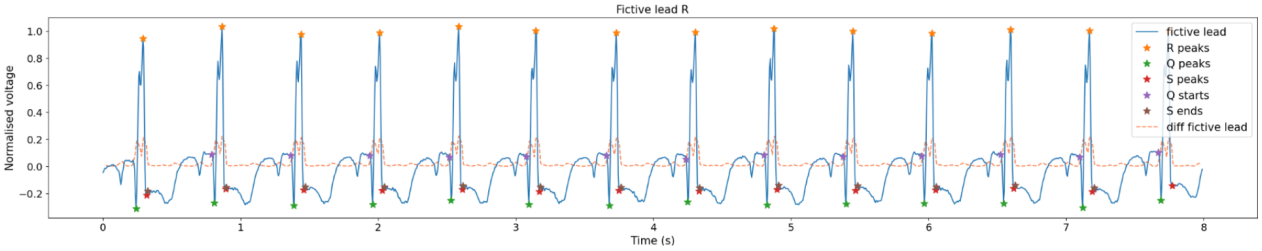


Figure 6: Fictive lead R with detected Q, R, and S peaks and QRS start and end points.

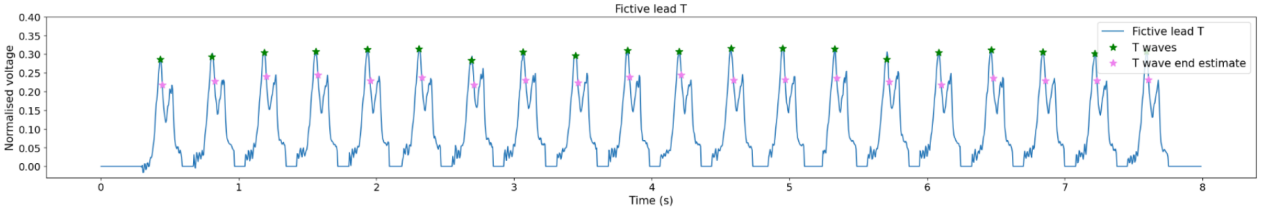


Figure 7: Fictive lead T with detected T waves and T wave end estimates.

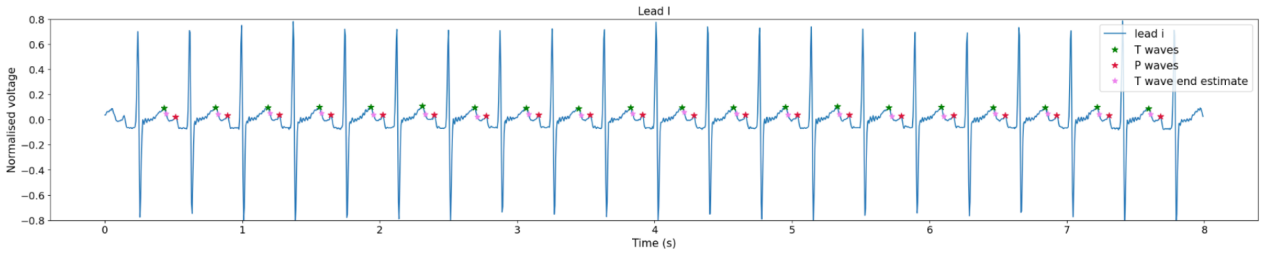


Figure 8: Selected lead (I) for P wave detection with annotated T waves, T wave end estimates and P waves.

3.3 VCG loops

Figure 9 provides examples of the vector loops of QRS complexes and SQ segments of the same patient showing SR and JET, respectively. The dominant vector found for each loop is also shown.

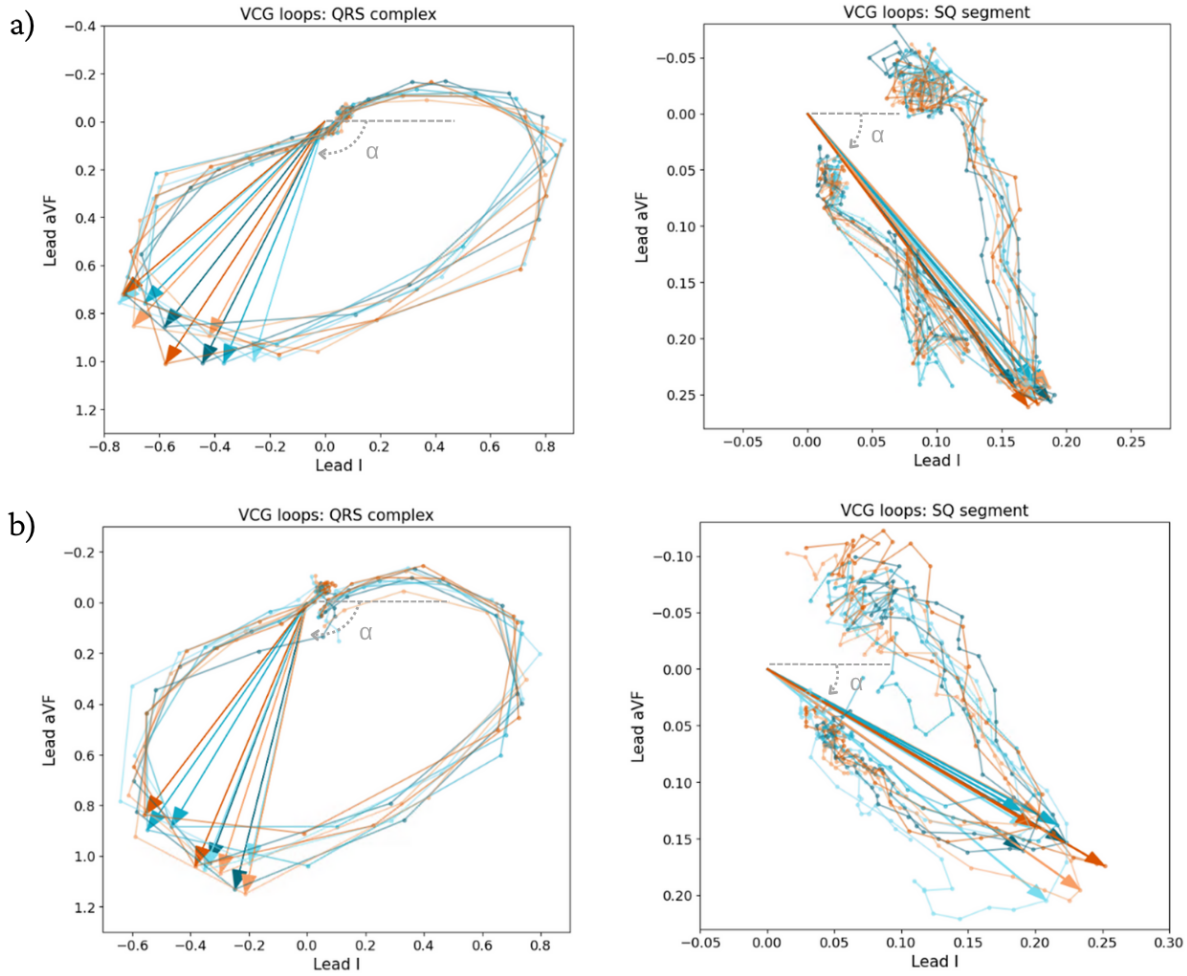


Figure 9: VCG loops of QRS complexes (left) and SQ segments (right) of a) SR and b) JET. 10 consecutive loops are shown in different colours. The dominant vector of each loop is illustrated with an arrow. α denotes the vector's angle.

3.4 Features

Table 2 presents the median and Q1 and Q3 of the feature values for SR and JET. The Pearson correlation matrix of all features is given in Figure 21 in Appendix A.4. It shows that pq_var is moderately correlated (>0.5) to RR_sd and p_ang_sd. Additionally, several VCG features exhibit high collinearity (>0.7), especially when describing the same signal segments (either SQ segments or QRS complexes). Particularly, the DTW distances are highly correlated to the derivative-based DTW distances.

Table 2: Feature values for SR and JET, given in median and first and third quartile values. Descriptions of all features can be found in Table 6 in Appendix A.3.

Feature	SR (Median (Q1-Q3))	JET (Median (Q1-Q3))
HR	130 (116 - 145)	157 (125 - 172)
RR std	0.90 (0.58 - 1.34)	2.96 (1.9 - 6.75)
p_valid	1 (1 - 1)	0.38 (0.12 - 0.62)
pq_var	0.42 (0.28 - 0.58)	21.1 (8.8 - 54.4)
p_angle_sd	3.97 (2.38 - 5.06)	6.05 (2.66 - 9.75)
qrs_mag_sd	0.034 (0.02 - 0.050)	0.051 (0.033 - 0.072)
qrs_angle_sd	7.09 (4.36 - 12.2)	7.36 (4.04 - 14.3)
sq_mag_sd	0.012 (0.009 - 0.017)	0.018 (0.012 - 0.0312)
sq_angle_sd	2.07 (1.47 - 3.00)	2.94 (2.32 - 4.47)
sq_distance	0.036 (0.027 - 0.047)	0.033 (0.025 - 0.054)
sq_distance_10	0.039 (0.031 - 0.059)	0.077 (0.042 - 0.103)
qrs_distance	0.11 (0.09 - 0.13)	0.13 (0.10 - 0.17)
qrs_distance_10	0.22 (0.18 - 0.24)	0.24 (0.20 - 0.30)
sq_der_distance	0.014 (0.011 - 0.022)	0.011 (0.007 - 0.014)
sq_der_distance_10	0.021 (0.018 - 0.022)	0.026 (0.020 - 0.035)
qrs_der_distance	0.082 (0.060 - 0.098)	0.094 (0.071 - 0.117)
qrs_der_distance_10	0.15 (0.13 - 0.17)	0.17 (0.14 - 0.20)
sq_ang_distance	0.16 (0.10 - 0.26)	0.11 (0.07 - 0.19)
sq_ang_distance_10	0.27 (0.20 - 0.45)	0.35 (0.21 - 0.57)
qrs_ang_distance	0.20 (0.17 - 0.28)	0.25 (0.17 - 0.35)
qrs_ang_distance_10	0.65 (0.52 - 0.77)	0.79 (0.63 - 1.00)

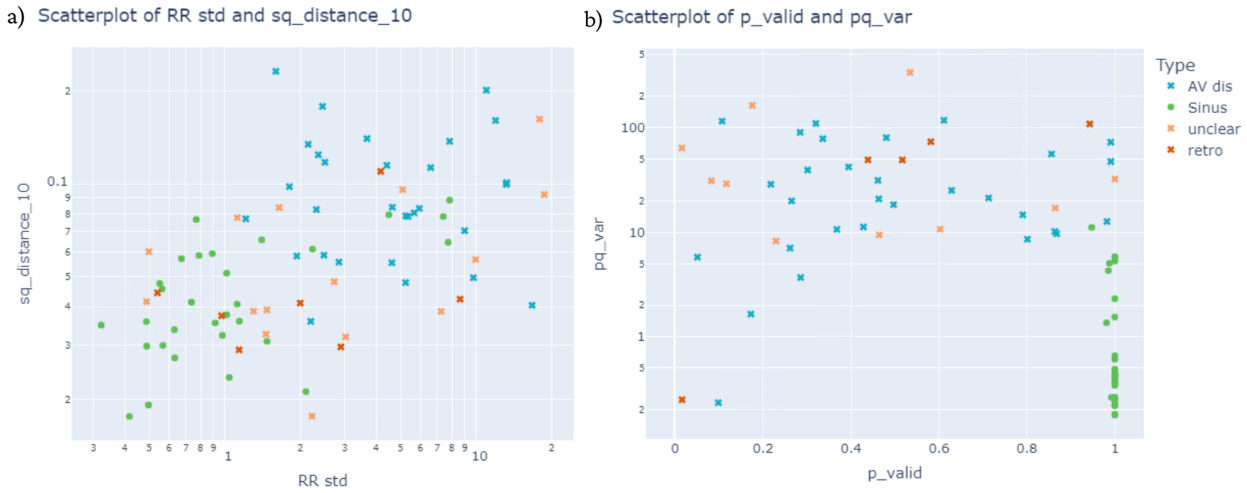


Figure 10: Two examples of scatterplots showing the distribution of features for a window size of 60 s. Different subtypes of JET are labelled. a) RR_std and $sq\_dtw\_distance\_10$. b) p_valid and pq_var.

Figure 10 presents two examples of scatter plots comparing feature values across SR and various JET subtypes: cases with AV dissociation (*AV dis*), with retrograde P waves (*retro*) and with an unclear AV relation (*unclear*). The plots indicate a clearer separation between JET and SR when using the features p_valid and pq_var, whereas the distinction appears less pronounced for RR_sd and $sq\_distance\_10$. Various additional scatter plots, visualising the feature values of the dominant vectors and the derivative-based and

angular distances, are provided in Appendix A.5. They show no clear separations between SR and JET. The MI scores, reflecting how much each feature reduces uncertainty about the classification label, are given in Figure 11, with p_valid and pq_var showing the highest scores.

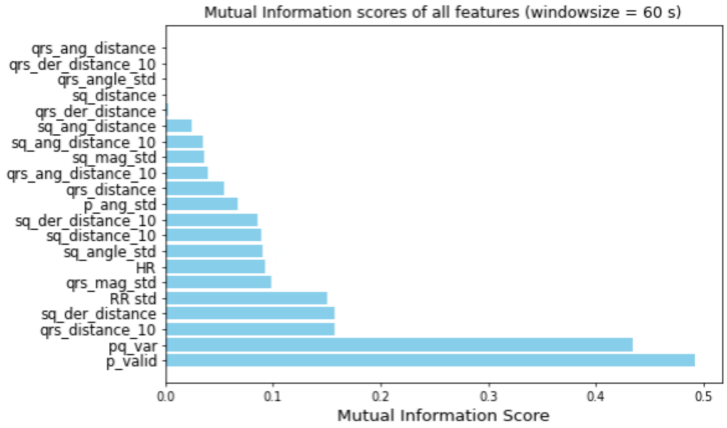


Figure 11: Mutual Information scores of all features, for a window size of 60 s.

3.5 Classification Model

3.5.1 Feature selection

The results of feature selection using 60-second windows are presented in Table 3. The number of features selected ranged from 1 to 5 for LR and from 3 to 8 for RF. While some features were consistently selected across folds, several others were only selected once. In RF models, the features p_valid, pq_var and RR_sd were selected in every fold. In LR, p_valid and pq_var were also frequently selected, appearing 5 and 4 times, respectively.

The results for other window sizes are summarised in Tables 9 and 10 in Appendix A.7. Across all window sizes, the results for both LR and RF models were comparable, with similar balanced accuracy scores. The features p_valid and pq_var (and RR_sd in LR) were consistently selected across all window sizes. It can be noted that when using 10-second windows, the number of selected features increased and the balanced accuracy slightly decreased.

For the final models, features p_valid and pq_var were selected for LR and p_valid, pq_var and RR_sd for RF. The feature importances and an illustration of the decision boundary of the LR model are provided in Appendix A.8.

3.5.2 Model performance

Model performance on the training set (for 60-second windows) is shown in Table 4. After hyperparameter optimisation, balanced accuracy scores of 0,989 and 0,988 were achieved for LR and RF, respectively. Performance results for other window sizes, along with the selected hyperparameters, are provided in Tables 12 and 13 in Appendix A.9. Using smaller window sizes generally led to slightly lower accuracy scores.

On the SR and JET labelled fragments of the validation dataset, the LR and RF model achieved balanced accuracy scores of 0,899 and 0,892, respectively. The other performance metrics are provided in Table 5. When comparing the LR and RF models, 4 predictions (2,2%) differed between them. Post-hoc analysis of the incorrect predictions showed that all were attributable to errors in peak detection. These errors stemmed from two main causes: variations in waveform morphology not accounted for during model development and interference from artefacts and/or noise. Within individual recordings, multiple prediction errors were often caused by the same underlying issue. In Appendix A.10, several examples are provided to illustrate these types of errors.

Table 3: Results of nested cross-validation using LR and RF models, with automatic feature selection and hyperparameter tuning. The mean and SD of the balanced accuracy score and the median and range of the number of features across the 5 outer loops are reported. The selected features and their count across these 5 folds are also given. 60-second windows were used.

	LR	**RF**
Balanced accuracy (mean (SD))	0,935 (0,057)	0,957 (0,039)
Number of selected features (median (range))	2 (1-5)	4 (3-8)
Feature	**Times selected (N)**	
p valid	5	5
pq var	4	5
RR std		5
sq der dist	1	3
sq ang std	1	
sq der dist 10	1	
qrs dist 10		1
p ang sd		1
HR		1
qrs ang dist 10		1
sq ang dist		1

Table 4: Performance of LR and RF models on the training dataset, after 5-fold stratified group cross-validation. 60-second windows were used. The LR model includes the features p_valid and pq_var, and the RF model includes p_valid, pq_var and RR_sd.

	LR	**RF**
Balanced accuracy	0,989	0,988
AUROC	0,997	0,992
FPR (%)	0,0	2,5
FNR (%)	2,2	0,0

Table 5: Performance of LR and RF models on validation dataset. The LR model includes the features p_valid and pq_var, and the RF model includes p_valid, pq_var and RR_sd.

	LR	**RF**
Balanced accuracy	0,899	0,892
AUROC	0,957	0,929
FPR (%)	9,5	10,9
FNR (%)	10,6	10,6

Besides the segments labelled as SR or JET, there were 8 fragments which contained premature ventricular contractions (PVC), ectopic atrial rhythms (EAT), single nodal beats or unclear rhythms. Of these, five were classified as JET and three as SR by both models. The presence of PVCs notably led to errors in peak detection. For cases with EAT or single nodal beats, the impact on the features (primarily pq_var) depended on the number of deviating beats. If the deviation was significant enough, it resulted in a JET prediction. Appendix A.10 provides several examples that demonstrate this.

4 Discussion

This study aimed to identify distinguishing ECG and VCG features between JET and SR and develop an automatic JET detection model, using monitor ECG data.

The features that proved to be most important to distinguish JET from SR, and that were included in the final models are p_valid, the number of detected P waves, and pq_var, the variance of the PQ interval, with RR_sd also included in RF. Although other features were selected in some folds during feature selection, their selection was highly dependent on the training data, indicating instability. These features likely reflect patient- or instance-specific characteristics rather than generalisable indicators of JET.

The developed models achieved high balanced accuracy scores on the training dataset, with LR and RF models attaining 0,989 and 0,988, respectively. In the secondary validation, lower balanced accuracy scores of 0,899 and 0,892 were achieved. These lower scores suggest overfitting in the initial model, with errors primarily originating from the peak detection stage. This highlights that peak detection is currently the most important limiting factor for the robustness and generalisability of this method.

4.1 Comparison to literature

Two studies have reported on automatic JET detection on ECG monitor data. Waugh et al. proposed the use of median P wave prominence and the interquartile range (IQR) of the PR interval for the classification of JET [14]. In a LR model, they achieved an AUROC of 0,93 for a test cohort of 25 patients. However, it should be noted that the included test cohort presented with relatively low frequency tachycardias. For cases of high frequency tachycardias, which is commonly observed in JET, their algorithm functioned less effectively. This may the result of the merging of P waves with T waves, making P waves difficult to detect. In a follow-up study, Ju et al. investigated a Deep Learning (DL) approach using Convolutional Neural Networks (CNNs), achieving an AUROC of 0,953 [15]. To address the interpretability limitations typical of DL models, they used Local Interpretable Model-Agnostic Explanations (LIME) to visualise which parts of each heartbeat influenced the model's predictions. They revealed that their model primarily relied on P wave features. One of the factors leading to classification errors was interpatient variability. False positive errors occurred in cases where P waves had such low prominences that they were not detected. False negatives occurred in cases of abnormal morphology.

These studies show considerable similarity to the results of this research. Our LR model achieved a comparable AUROC of 0,957, but with the added benefit of greater interpretability and transparency than DL models, making it more suitable as a bedside tool. Yet, it is important to note that methodological differences, such as their beat-by-beat classification versus our segment-based classification, and the use of different datasets, make direct performance comparisons difficult.

Nevertheless, a key conclusion drawn from all of these results is that P waves play a crucial role in the detection of JET. Like the two mentioned studies, this research also encountered challenges with reduced P wave prominence and morphological variation across patients. Additional challenges identified in this study, including the presence of noise and artefacts, and other arrhythmias, were not reported in literature, because segments exhibiting movement artifacts or unphysiological data were excluded in those studies, making their results less representable for real-world data and clinical use. The inclusion of a secondary validation in this study was very useful for assessing clinical applicability.

This study also introduces a broader methodological perspective by incorporating multi-lead analysis. Unlike prior models that relied solely on lead II, our approach which leverages all available leads offers potential advantages. It recognises that the maximum P wave amplitude may not always align with lead II, especially within this patient group with congenital cardiac defects and often multiple prior cardiac surgeries. Using multiple leads enhances the likelihood of reliably capturing the P wave. This principle also applies to other ECG metrics. Since P wave detection often relies on the accurate identification of other peaks, robust methods for detecting the peaks of the QRS complex and T wave are equally crucial. These ECG metrics also exhibit varying morphologies across individuals. Consequently, using the VCG to compute new leads in the optimal direction for these peaks represents an innovative approach that appears promising for enhancing the robustness of peak detection.

Given the challenge of distinguishing the P-wave from noise or other ECG components, this study aimed to incorporate additional features to enhance the robustness of JET detection. However, the VCG-derived features that were introduced did not contribute significantly to JET detection.

4.2 Interpretation of the results

VCG features

Several VCG-derived features were introduced, focusing primarily on variability across beats. These features generally received low MI scores and were not consistently selected during the feature selection process, indicating limited relevance for general classification. Nonetheless, exploratory scatter plots, when stratified by JET subtypes, suggested some patterns. For instance, while in general no clear separation is seen between SR and JET when plotting RR_sd and $sq\_distance\_10$ (Figure 10a), higher values for both features were seen in JET cases with AV dissociation. Higher RR_sd values are expected since JET can present as an irregular rhythm in case of both nodal beats and capture beats. Also, higher $sq\_distance$ values were anticipated due to the P wave displaying different relative timing with each beat, causing an inconsistent beat-to-beat morphology. However, there is still substantial overlap between JET and SR. This may be partially due to the changes in JET being too subtle to be captured effectively by this measure. In JET, the P wave may appear spread over the entire SQ segment, leading to higher distances, but there are also instances where it shifts only minimally in time. In such cases, if there is a difference in beat length, DTW may still achieve alignment with low distance scores. Conversely, noisy signals can lead to higher DTW distances, which could explain the SR samples with higher values (see Figure 25 in Appendix A.5 for a set of examples to visualise this). Although signal smoothing or additional filters could theoretically reduce the influence of noise, these approaches risk suppressing the already low-amplitude P wave, potentially obscuring diagnostically relevant information. Since the P wave is central to detecting JET, excessive filtering is undesirable. Balancing noise reduction with the preservation of important waveform features remains an important challenge in this research.

For distance measures of the QRS complex, aberrant conduction was hypothesised to be an additional cause for beat-to-beat variability next to the 'wandering' P wave. Aberrancy can occur as a result of sudden changes in cardiac frequency, where the length of the cardiac cycle changes without a compensatory change in the length of the refractory period [23]. These changes in frequency are particularly expected during the warm-up phase of JET, which is also when AV dissociation is commonly observed. However, only slightly higher $qrs\_distance$ values can be seen for JET ($AVdis$), with substantial overlap with SR (see Figure 22 in Appendix A.5). This finding may also be explained by the subtle beat differences combined with the presence of noise.

Altogether, despite the inclusion of additional measures of DTW derivative-based and angular distances, which are theoretically useful for capturing differences in beat shape and depolarisation direction, VCG features did not provide additional value in distinguishing JET from SR.

However, this does not preclude the potential of VCG-based features altogether. Future studies may benefit from computing alternative metrics from the VCG, such as directional changes during more specific signal segments, or applying these analyses selectively during rhythm transitions (e.g., onset of JET or return to SR). Moreover, incorporating additional ECG leads, particularly the precordial leads (V1–V6), may provide a more detailed spatial representation and thereby enhance the effectiveness of VCG-based features.

P wave features

As mentioned before, P waves and specifically the features p_valid and pq_var appeared to be most important for distinguishing JET from SR. Within the training dataset, the value of p_valid was consistently very close to 1 for all SR fragments. For some JET fragments, this value was also close to 1, but this was typically accompanied by a higher pq_var value. This elevated pq_var may reflect either dissociated P waves causing true PQ variability, or noise falsely detected as P waves, resulting in higher interval variance due to randomness. It is important to note that detecting noise is accepted within this model, as the pq_var value helps differentiate whether the detected peaks are true P waves corresponding to SR. Another key takeaway from this finding is that the model has been trained to require a high p_valid value for an SR prediction,

implying that missing P waves would likely result in a JET prediction. This underscores the model's strong reliance on accurate identification of P waves.

Some detection methods used in this study may not generalise to other arrhythmias. For instance, the P wave detection approach assumes consistent directional alignment, reflecting atrial activity from a single origin, typically the sinus node. This assumption does not hold in arrhythmias such as EAT. In such cases, selecting the lead with the lowest P wave angle variation may not yield reliable results.

Peak detection

Beyond P wave detection, accurate peak detection more broadly forms the foundation for many components of this research. Some features are directly calculated from the detected R peaks, and accurate detection of the QRS complex is essential for signal segmentation, which is followed by analysis of the vector loops. Additionally, the computation of the fictive lead for T wave detection relies on QRS cancellation, and P wave detection depends on accurate identification of T waves. Altogether, this creates numerous dependencies, meaning that errors in one step can propagate through the pipeline. However, peak detection was not the primary focus of this research and was not directly manually validated, unlike in previous work. One key advantage of this study was that the dataset encompassed a variety of patients with different ECG morphologies, including signals with high irregularities, inconsistent QRS complex amplitudes, and conduction disorders such as BBBs. During model development, these abnormalities were taken into account, allowing for the optimisation of the peak detection methods over a wide range of pathophysiologies. Several iterations on earlier approaches improved robustness (as illustrated in Appendix A.2). Despite these improvements, errors in peak detection persisted in the secondary validation, highlighting the need for further refinement.

These errors were not only linked to morphological variations but also to the presence of noise and artefacts. A contributing factor was the use of bedside monitor ECG data, which is inherently prone to signal disturbances such as baseline wander, drift, and muscle activity, often caused by respiration, movement, or poor electrode contact [24, 25, 26]. Although the use of this data type is considered a strength of this study, enhancing clinical relevance and real-world applicability, it introduces practical challenges that must be addressed through robust preprocessing and pipeline design.

Model performance

High performance was achieved in this study, with respect to annotations provided by a single expert. LR and RF models gave comparable results. RF was included to capture potential nonlinear patterns. However, since JET and SR were almost perfectly separable using only two features, as shown by the LR model, the added complexity of RF may not be required for this specific classification task. In this context, LR offers the advantage of simplicity and interpretability, making it well-suited for clinical implementation. More complex models such as RF may become valuable when expanding to multi-class classification involving other arrhythmias.

Since the model was trained exclusively on SR and JET fragments, a positive prediction may not be highly specific to JET. Given the included features, the model is better described as detecting non-sinus rhythms that lack consistent atrial activity preceding ventricular depolarisation. Consequently, other arrhythmias, such as AV block or re-entry tachycardias, may also be classified as JET.

Performance metrics from the secondary validation were calculated only for SR and JET fragments, in line with the study's primary focus. However, a few additional fragments containing other rhythms were also processed. In several of these cases, the model indeed predicted JET, although this was partially due to peak detection errors that distorted the extracted feature values. Nonetheless, the number of such fragments was too limited to draw any conclusions about the model's behaviour on other arrhythmias.

The limited dataset remains a key factor influencing the generalisability of the validation results. The training set was also imbalanced, with more JET than SR fragments. Although this imbalance could have increased the risk of false positives, this was not specifically observed in the secondary validation results.

The model performance across different window sizes (Tables 12 and 13) indicates that 60-second windows yield the highest performance, though the differences are relatively small. The lower performance

observed with 10-second windows may be attributed to insufficient deviations in p_valid and pq_var over shorter intervals. In particular, in cases where the nodal rate is only slightly higher than the atrial rate, usually during the warm-up phase of JET, p_valid might remain close to 1 and pq_var very low, resulting in a SR classification. Additionally, shorter window sizes led to a larger number of selected features. This could be due to increased overlap in p_valid and pq_var values between SR and JET, necessitating more features to distinguish between the two rhythms. Given the higher overall performance and expected greater robustness, the 60-second window was used for further validation.

4.3 Limitations

This study has several methodological and data-related limitations. The labelling process, conducted by a single paediatric intensivist, introduces subjectivity. When assessing the ECG, clinicians' opinions may vary. Moreover, as arrhythmias evolve over time, it is challenging to define clear thresholds for labelling a rhythm as JET. For example, a rhythm may initially appear as SR but show occasional nodal beats, complicating the decision of when it should be classified as JET. Furthermore, the question can be raised whether the different phases of JET over time (warm-up, high frequency, cooling-down) should be treated as separate classes. While various subtypes (retro, AV dis, unclear) were labelled for exploratory analysis, all JET fragments were treated as one single class. Splitting the data into distinct classes may have revealed new patterns. However, with the current sample size the study was not powered to conduct such subtype analysis. Furthermore, given that only two features were sufficient to achieve high accuracy in the complete dataset and that significant overlap was observed between SR and JET in the scatter plots of other features, it is unlikely that splitting the subtypes would have given substantially different results in this study.

Additionally, the ECG data in this study was sampled at a frequency of 200 Hz. The American Heart Association recommends 150 Hz as minimum bandwidth and 500 Hz as minimum sampling rate for recording paediatric ECGs [27]. Rijnbeek et al. even recommend a higher sampling rate, after investigating errors in maximum QRS amplitudes for different low-pass filter cut-offs [28]. They found that the youngest children, who have the highest frequency content, also exhibit more errors. For children under 1 year old, similar to the population in this study, they suggest a minimum bandwidth of 250 Hz and recommend a sampling frequency of at least 1000 Hz. Insufficient sampling frequencies can lead to altered wave amplitudes, thus limiting data quality. Using data with higher sampling frequencies could improve the quality of the data and may enhance the performance of features. Especially those representing beat-to-beat variability, which aim to capture very subtle differences, may benefit from improvement of data quality.

4.4 Recommendations

Several steps are recommended for further research. First, improving the robustness of the overall pipeline is crucial, as secondary validation revealed that many errors stemmed from inaccurate peak detection. Some mistakes occurred due to deviating waveform morphologies not encountered during model development, which highlights the need to expand the dataset to include more patient data, allowing for further iterations. Other mistakes occurred with high levels of noise or many artefacts, which could be addressed by improving preprocessing. However, in these situations the P wave is often distorted or obscured by noise, which cannot be completely restored by additional preprocessing.

Given the noise and artefacts commonly encountered in the monitor ECG, achieving a perfect peak detection method is unlikely. Integrating automatic signal quality assessment (SQA) could help identify unreliable segments and prevent misleading predictions. Numerous SQA techniques have been proposed in literature, ranging from simple thresholding and morphological change detection to advanced machine learning and signal decomposition techniques [29]. There are several straightforward approaches that could be implemented within the model pipeline, to review peak detection accuracy. For instance, the RR_sd value can provide insight into potential issues with R peak detection. While higher values may indicate rhythm irregularities, exceeding a specific threshold could signal errors in R peak detection. Similarly, for T wave detection, high variability in the ST interval could point to potential issues in the identification of T waves. Incorporating such checkpoints and refraining from making predictions when certain thresholds are exceeded could enhance the model's robustness. If these simple measures are insufficient to address incorrect predictions caused by poor data quality, more complex SQA techniques should be explored.

If manual features prove insufficient to ensure robust and specific classification, the use of DL should be further explored, as it offers the major advantage of capturing more complex patterns. Instead of a single-beat approach, as done by Ju et al., using multi-beat segments (as in this study) or incorporating memory over longer time windows may improve classification by capturing temporal dynamics. Clinical rhythm interpretation often relies on such patterns: for instance, a sudden onset of high-frequency tachycardia may indicate re-entry tachycardia, while a gradual increase following AV dissociation is more suggestive of JET [30]. DL models with memory components, such as long short-term memory (LSTM) models, could help mimic this contextual reasoning, offering a way to integrate temporal patterns into classification [31]. However, their effectiveness would depend on having sufficient and representative temporal data.

Before clinical deployment, alarm logic should be refined. Since alarm fatigue is a well-recognised issue in the PICU, an excessive number of false positives should be prevented [32, 33]. Currently, the probability threshold for a JET prediction is set at 0.5. Adjusting this threshold, combined with requiring multiple consecutive windows to be classified as JET, could help reduce unnecessary alerts.

If validated for long-term use, the model could support studies into JET treatment strategies, such as antiarrhythmic drugs. A study by Sylva et al. on eight patients treated with ivabradine for JET identified an association between higher ivabradine serum levels and lower heart rates [34]. However, the direct impact on JET remains unexplored. With automated rhythm detection, long-term ECG data could be used to better understand how JET evolves in response to medication over time. These insights could contribute to the development of pharmacodynamic models that link drug levels to rhythm patterns.

5 Conclusion

This study demonstrates that P wave-related features, specifically the number of detected P waves and the variance of the PQ interval, are most effective for distinguishing JET from SR in critically ill children with CHD admitted to the PICU. In contrast, VCG-derived features based on beat-to-beat variability did not contribute meaningfully to classification.

The LR and RF models achieved high balanced accuracy scores on the training set (0,989 and 0,988, respectively). However, secondary validation revealed reduced performance (0,899 and 0,892), primarily due to errors in peak detection. These findings underscore that accurate and robust peak detection remains the key limitation for generalising this method to broader clinical use.

References

[1] P.-L. Bernier et al. “The Challenge of Congenital Heart Disease Worldwide: Epidemiologic and Demographic Facts”. en. In: *Seminars in Thoracic and Cardiovascular Surgery: Pediatric Cardiac Surgery Annual* 13.1 (Jan. 2010), pp. 26–34. ISSN: 10929126. DOI: 10.1053/j.pcsu.2010.02.005.

[2] J. I. Hoffman and S. Kaplan. “The incidence of congenital heart disease”. en. In: *Journal of the American College of Cardiology* 39.12 (June 2002), pp. 1890–1900. ISSN: 07351097. DOI: 10.1016/S0735-1097(02)01886-7.

[3] *Cyanotic Heart Disease - StatPearls - NCBI Bookshelf*.

[4] M. S. Kabbani et al. “Critical arrhythmia in postoperative cardiac children: Recognition and management”. en. In: *Avicenna Journal of Medicine* 07.03 (July 2017), pp. 88–95. ISSN: 2231-0770, 2249-4464. DOI: 10.4103/ajm.AJM_14_17.

[5] R. I. Kylat and R. A. Samson. “Junctional ectopic tachycardia in infants and children”. en. In: *Journal of Arrhythmia* 36.1 (Feb. 2020), pp. 59–66. ISSN: 1880-4276, 1883-2148. DOI: 10.1002/joa3.12282.

[6] M. Ashraf and A. Goyal. “Junctional Ectopic Tachycardia”. eng. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing, 2025.

[7] M. Alasti et al. “Junctional ectopic tachycardia (JET)”. en. In: *Journal of Arrhythmia* 36.5 (Oct. 2020), pp. 837–844. ISSN: 1880-4276, 1883-2148. DOI: 10.1002/joa3.12410.

[8] *Junctional Ectopic Tachycardia Workup: Laboratory Studies, Electrocardiography, Imaging Studies*.

[9] L. Mildh et al. “Junctional ectopic tachycardia after surgery for congenital heart disease: incidence, risk factors and outcome”. en. In: *European Journal of Cardio-Thoracic Surgery* 39.1 (Jan. 2011), pp. 75–80. ISSN: 10107940. DOI: 10.1016/j.ejcts.2010.04.002.

[10] N. Sasikumar, R. K. Kumar, and S. Balaji. “Diagnosis and management of junctional ectopic tachycardia in children”. en. In: *Annals of Pediatric Cardiology* 14.3 (July 2021), pp. 372–381. ISSN: 0974-2069. DOI: 10.4103/apc.apc_35_21.

[11] N. A. Haas and C. K. Camphausen. “Impact of early and standardized treatment with amiodarone on therapeutic success and outcome in pediatric patients with postoperative tachyarrhythmia”. en. In: *The Journal of Thoracic and Cardiovascular Surgery* 136.5 (Nov. 2008), pp. 1215–1222. ISSN: 00225223. DOI: 10.1016/j.jtcvs.2008.04.011.

[12] A. Dodge-Khatami. “Impact of junctional ectopic tachycardia on postoperative morbidity following repair of congenital heart defects”. In: *European Journal of Cardio-Thoracic Surgery* 21.2 (Feb. 2002), pp. 255–259. ISSN: 10107940. DOI: 10.1016/S1010-7940(01)01089-2.

[13] G. Izumi et al. “Perioperative junctional ectopic tachycardia associated with congenital heart disease: risk factors and appropriate interventions”. en. In: *Heart and Vessels* 37.10 (Oct. 2022), pp. 1792–1800. ISSN: 0910-8327, 1615-2573. DOI: 10.1007/s00380-022-02074-3.

[14] J. L. S. Waugh et al. “A novel automated junctional ectopic tachycardia detection tool for children with congenital heart disease”. In: *Heart Rhythm O2* 3.3 (June 2022), pp. 302–310. ISSN: 2666-5018. DOI: 10.1016/j.hroo.2022.02.014.

[15] Y. Ju et al. “A multimodal deep learning tool for detection of junctional ectopic tachycardia in children with congenital heart disease”. en. In: *Heart Rhythm O2* 5.7 (July 2024), pp. 452–459. ISSN: 26665018. DOI: 10.1016/j.hroo.2024.04.014.

[16] E. Frank. “An Accurate, Clinically Practical System For Spatial Vectorcardiography”. en. In: *Circulation* 13.5 (May 1956), pp. 737–749. ISSN: 0009-7322, 1524-4539. DOI: 10.1161/01.CIR.13.5.737.

[17] J. Meijer. “Application of VCG on the Detection, Classification or Characterization of Cardiac Arrhythmias: a Scoping Review”. Unpublished work. 2024.

[18] G. Raaijmakers. “Automated electrocardiogram interpretation for the detection of postoperative junctional ectopic tachycardia at the pediatric intensive care unit”. Unpublished work. 2024.

[19] J. P. Do Vale Madeiro et al. “Techniques for Noise Suppression for ECG Signal Processing”. en. In: *Developments and Applications for ECG Signal Processing*. Elsevier, 2019, pp. 53–87. ISBN: 978-0-12-814035-2. DOI: 10.1016/B978-0-12-814035-2.00009-8.

[20] D. Erb. *pybaselines: A Python library of algorithms for the baseline correction of experimental data*. Mar. 2025. DOI: 10.5281/ZENODO.5608581.

[21] V. Tuzcu and S. Nas. “Dynamic Time Warping As a Novel Tool in Pattern Recognition of ECG Changes in Heart Rhythm Disturbances”. In: *2005 IEEE International Conference on Systems, Man and Cybernetics*. Vol. 1. Waikoloa, HI, USA: IEEE, 2005, pp. 182–186. ISBN: 978-0-7803-9298-4. DOI: 10.1109/ICSMC.2005.1571142.

[22] *An introduction to Dynamic Time Warping*.

[23] D. H. Singer and R. E. Ten Eick. “Aberrancy: Electrophysiologic aspects”. en. In: *The American Journal of Cardiology* 28.4 (Oct. 1971), pp. 381–401. ISSN: 00029149. DOI: 10.1016/0002-9149(71)90002-6.

[24] U. Satija, B. Ramkumar, and M. S. Manikandan. “A Review of Signal Processing Techniques for Electrocardiogram Signal Quality Assessment”. In: *IEEE Reviews in Biomedical Engineering* 11 (2018). Publisher: Institute of Electrical and Electronics Engineers (IEEE), pp. 36–52. ISSN: 1937-3333, 1941-1189. DOI: 10.1109/rbme.2018.2810957.

[25] G. Friesen et al. “A comparison of the noise sensitivity of nine QRS detection algorithms”. In: *IEEE Transactions on Biomedical Engineering* 37.1 (1990). Publisher: Institute of Electrical and Electronics Engineers (IEEE), pp. 85–98. ISSN: 0018-9294. DOI: 10.1109/10.43620.

[26] G. D. Clifford. “ECG Statistics, Noise, Artifacts, and Missing Data”. In: *Advanced Methods and Tools for ECG Data Analysis*. 2006, pp. 55–99.

[27] J. J. Bailey et al. “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.” en. In: *Circulation* 81.2 (Feb. 1990). Publisher: Ovid Technologies (Wolters Kluwer Health), pp. 730–739. ISSN: 0009-7322, 1524-4539. DOI: 10.1161/01.cir.81.2.730.

[28] P. R. Rijnbeek, J. A. Kors, and M. Witsenburg. “Minimum Bandwidth Requirements for Recording of Pediatric Electrocardiograms”. en. In: *Circulation* 104.25 (Dec. 2001). Publisher: Ovid Technologies (Wolters Kluwer Health), pp. 3087–3090. ISSN: 0009-7322, 1524-4539. DOI: 10.1161/hc5001.101063.

[29] P. Kumar and V. K. Sharma. “Detection and classification of ECG noises using decomposition on mixed codebook for quality analysis”. en. In: *Healthcare Technology Letters* 7.1 (Feb. 2020). Publisher: Institution of Engineering and Technology (IET), pp. 18–24. ISSN: 2053-3713, 2053-3713. DOI: 10.1049/htl.2019.0096.

[30] B. P. Knight. *Atrioventricular nodal reentrant tachycardia - UpToDate*.

[31] Y. Yu et al. “A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures”. en. In: *Neural Computation* 31.7 (July 2019). Publisher: MIT Press, pp. 1235–1270. ISSN: 0899-7667, 1530-888X. DOI: 10.1162/neco_a_01199.

[32] B. J. Drew et al. “Insights into the Problem of Alarm Fatigue with Physiologic Monitor Devices: A Comprehensive Observational Study of Consecutive Intensive Care Unit Patients”. en. In: *PLoS ONE* 9.10 (Oct. 2014). Ed. by L. G. Tereshchenko, e110274. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0110274.

[33] H. Herrera and D. Wood. “Battling Alarm Fatigue in the Pediatric Intensive Care Unit”. en. In: *Critical Care Nursing Clinics of North America* 35.3 (Sept. 2023), pp. 347–355. ISSN: 08995885. DOI: 10.1016/j.cnc.2023.05.003.

[34] M. Sylva et al. “Serum Ivabradine Associates With Heart Rate Reduction But Enteral Exposure Is Unpredictable in Post surgical JET”. en. In: *JACC: Clinical Electrophysiology* 9.8 (Aug. 2023), pp. 1821–1823. ISSN: 2405500X. DOI: 10.1016/j.jacep.2023.05.019.

[35] *Clinical Practice Guidelines : Basic Paediatric ECG interpretation*.

[36] C. R. Vázquez-Seisdedos et al. “New approach for T-wave end detection on electrocardiogram: Performance in noisy conditions”. en. In: *BioMedical Engineering OnLine* 10.1 (Dec. 2011), p. 77. ISSN: 1475-925X. DOI: 10.1186/1475-925X-10-77.

A Appendix

A.1 Previous Research

In previous research, peak detection was applied to find the QRS peaks, T wave and P wave. Figure 12 provides a flowchart showing the different analysis steps. From the original ECG signals, a fictive lead was created in the direction of the highest amplitude of the signal. In this fictive lead, all peaks were detected and their prominences were fitted using a Gaussian model, assuming a bimodal distribution. The minimum between the two peaks of this distribution was used as prominence threshold for R peaks. Q and S peaks were subsequently detected by finding the peaks in the inverted fictive lead within 60 ms before and after the R peak, respectively. QRS complexes were then removed from the signal, after which a new fictive lead was created. In this lead, local maxima were detected and labelled as T waves. The start point of Q and the end of T were both identified using a 'trapezium's area approach', as proposed in literature. Following this, the signal from Q start points to T end points was removed by setting the signal to zero. Finally, another fictive lead was created, aimed to be in the direction of the P wave. P waves were detected using a similar approach as for R peaks: the prominence threshold was determined using a Gaussian fit model.

Using these peaks, the following features were computed: HR, SD of PR interval and fraction of P waves. Additionally, the VCG was created (using lead I and aVF), and the SD was computed of the vector length of the initial 50 ms of the QRS complex. Finally, a decision tree for the detection of JET was manually created, as illustrated in figure 13.

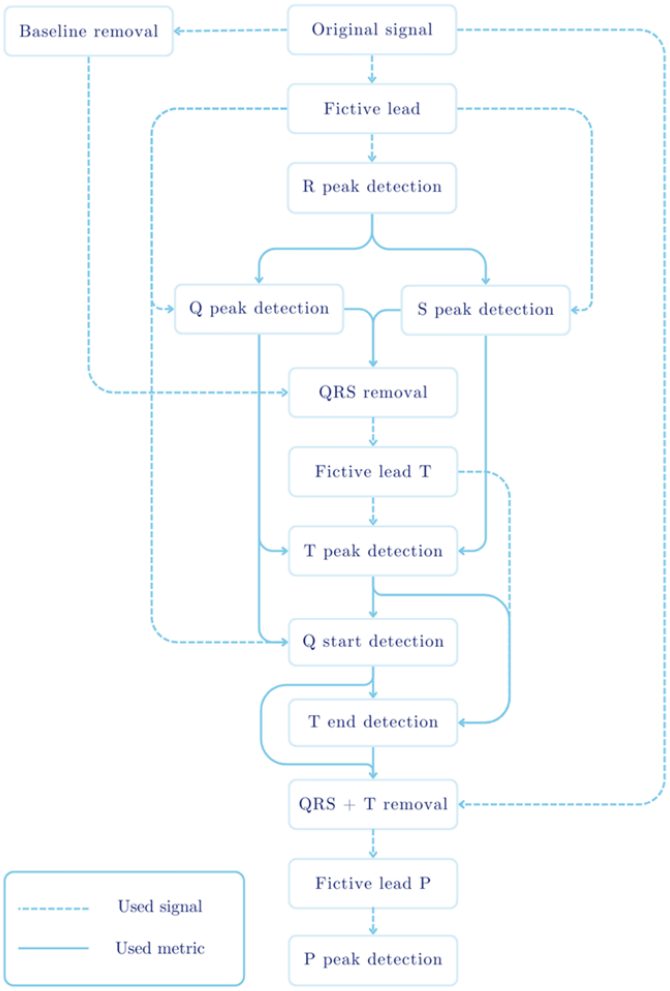


Figure 12: Flowchart of peak detection used in previous research [18].

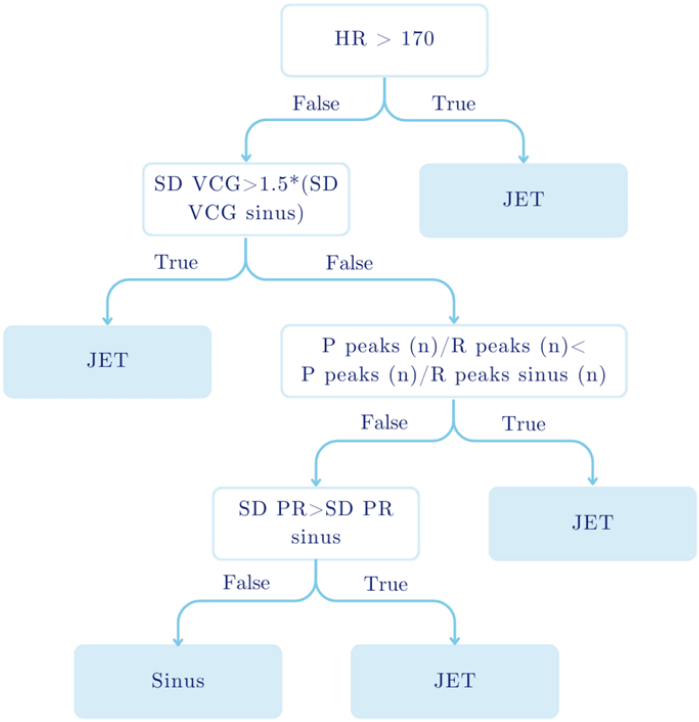


Figure 13: Decision tree for the classification of JET, developed in previous research [18].

A.2 Peak detection: typical examples

R peaks

Figure 14 illustrates two approaches for R peak detection in the fictive lead. In Figure 14a, the method based on previous research involves identifying all peaks and their prominences, followed by applying a prominence threshold derived from a Gaussian fit (as described in Appendix A.1). This example highlights a key limitation: prominence alone may not reliably distinguish R peaks from T waves. In contrast, Figure 14b shows the method used in this study, which also uses the fictive lead but applies a Gaussian fit to the mean steepness (i.e., smoothed derivative) around each peak to define a slope-based threshold. This approach improves the ability to discriminate R peaks from T waves.

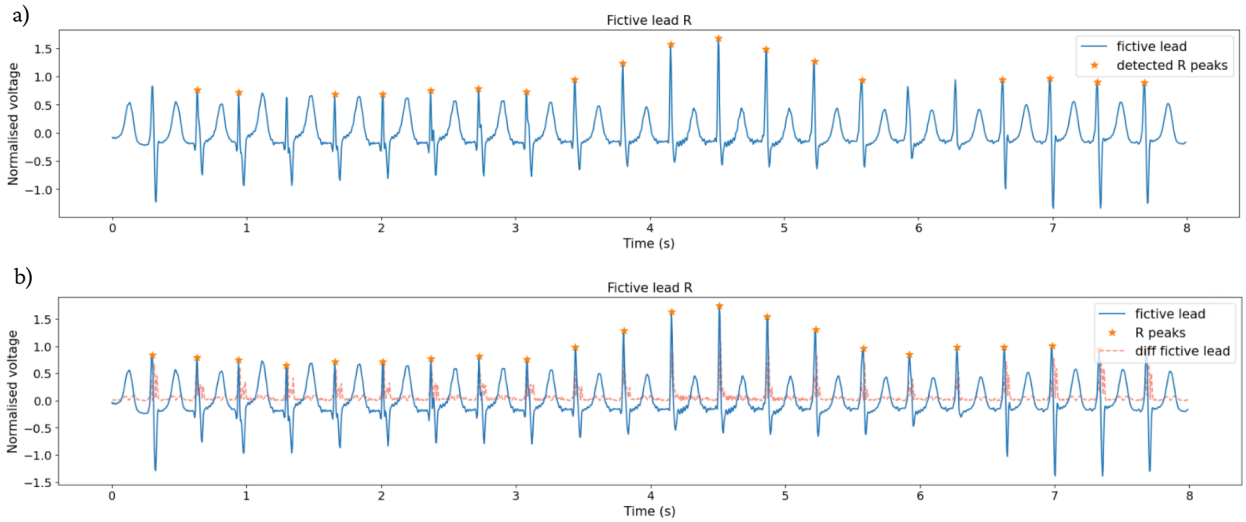


Figure 14: R peak detection in the fictive lead. a) Detection using a prominence-based threshold. b) Detection using a steepness-based threshold derived from the local signal slope. The first derivative of the fictive lead (diff) is shown to illustrate the discriminative steepness of R peaks.

Q and S peaks

Figure 15 illustrates different Q and S peak detection approaches in a patient with a BBB. In figure 15a, the method from previous work is applied, where Q and S peaks are detected in the inverted fictive lead within a 60 ms window around the R peak. The latest negative peak is selected for Q and the minimum value for S. In this example, Q peaks are detected too early due to the altered QRS morphology. Figure 15b demonstrates that simply broadening the search window does not resolve this issue. Selecting the last Q and the lowest S in a wider window results in inaccurate detection, with Q peaks occurring too late and S peaks too far from the actual QRS end. To address this, the method developed in this study selects peaks based on local steepness, requiring a steep positive slope following Q and a steep negative slope preceding S. As shown in Figure 15c, this improves peak detection in signals with conduction abnormalities.

Q start and S end

Figure 16a shows Q start detection when using the trapezium method proposed in previous work, resulting in mistakenly identifying the P waves as Q start points. In this study, Q start is identified as the point before Q where the smoothed derivative of the signal exceeds a certain threshold. Also, the S end point is added (needed for QRS segmentation for VCG analysis). S end is identified as the point after the S peak where the derivative drops below a second threshold. The resulting detections are shown in Figure 16b.

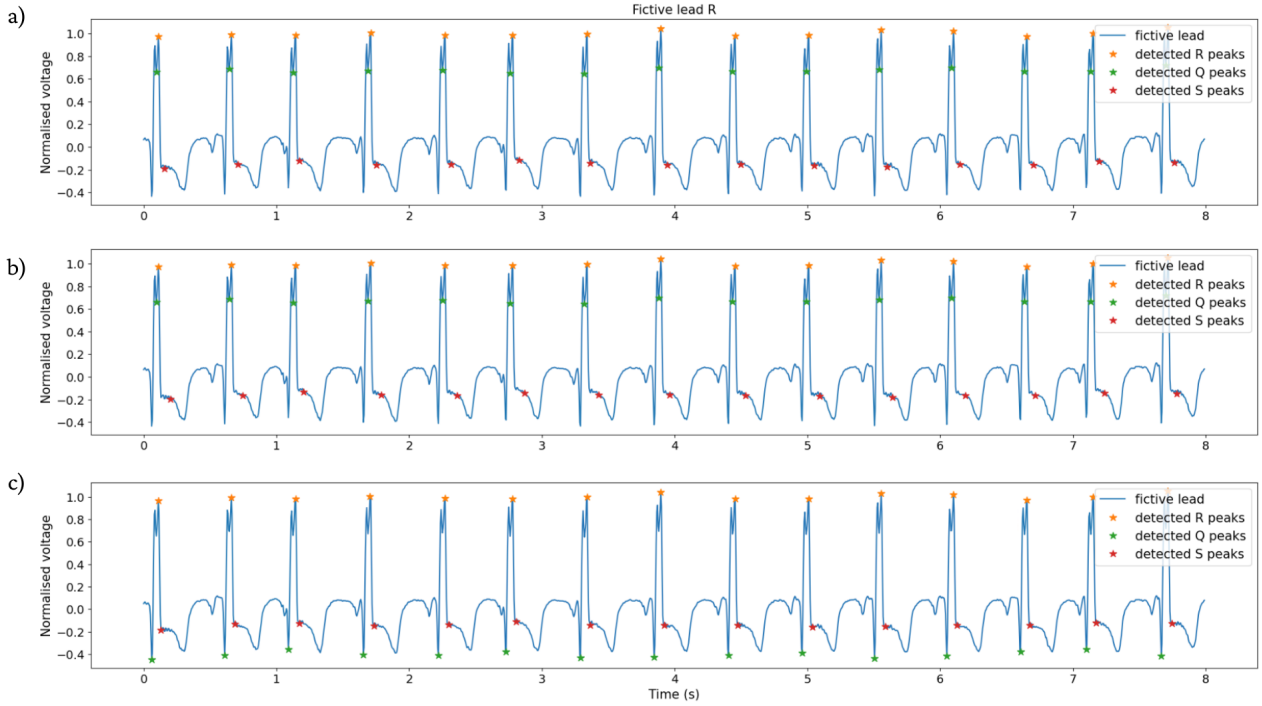


Figure 15: Q, R, and S peak detection in the fictive lead for a signal exhibiting a BBB. a) Detection using a fixed window and selection rules based on timing and amplitude. b) Broadened search window using the same rules, leading to incorrect peak assignment. c) Detection based on steepness criteria, improving accuracy in altered QRS morphology.

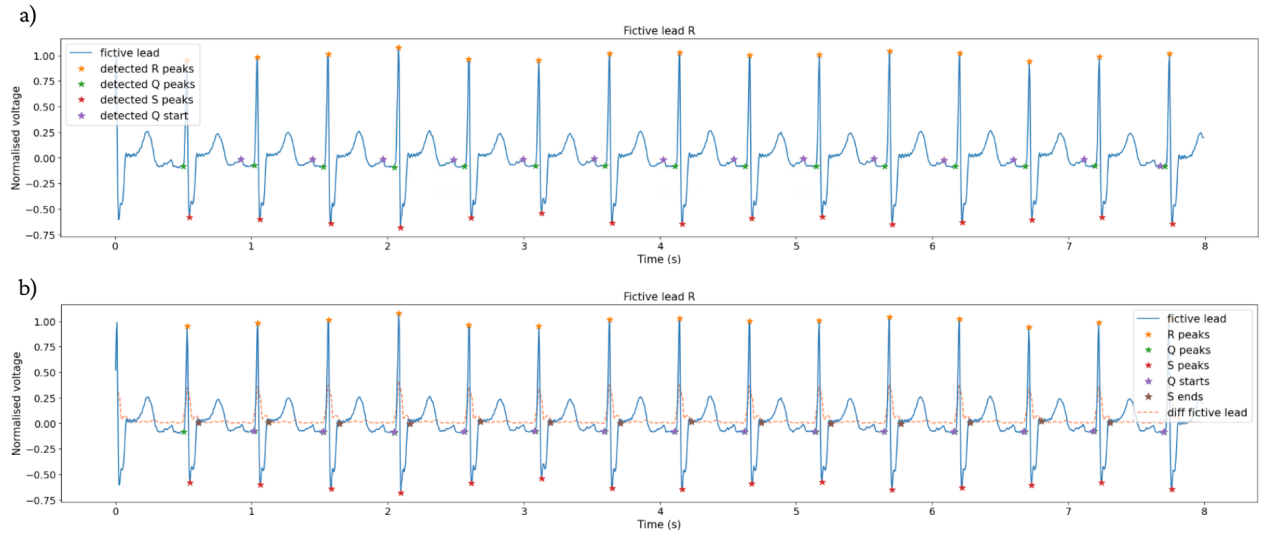


Figure 16: Q start and S end detection in the fictive lead. a) Trapezium-based method from previous work, with misclassification of P waves as Q start. b) Detection method using thresholds on the smoothed absolute derivative. The derivative (diff) is shown to visualise the slope-based detection.

T waves

In previous research, the highest amplitude of the fictive lead T within each SQ interval was marked as the T wave. However, the presence of flat or low-amplitude T waves, which can be a normal finding in newborns, can give errors when applying this method [35]. Figure 17 provides an example where T waves were lower in amplitude than P waves. Consequently, when selecting the highest peak in the fictive lead T, P waves are mistakenly labelled as T waves. When choosing the first of the two most prominent peaks, as applied in this study, T waves are correctly detected.

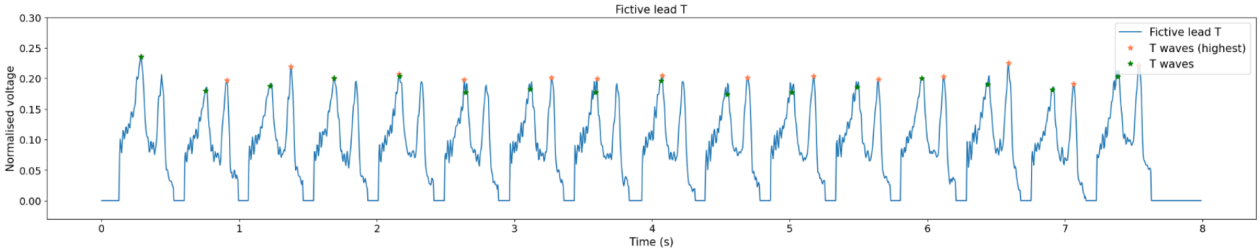


Figure 17: Fictive lead T with detected T waves when selecting the highest peak.

T wave end

Figure 18 shows an example of T wave end points detected when using the trapezium area approach as used in previous research [18, 36]. It can be seen that the P waves closely follow the T waves and that they are partially superimposed. When using this T wave end point as a starting point for searching the P wave, the P wave would be missed. It also shows the first zero crossings of the fictive lead T (e.g. inflection points), which are used in this study.

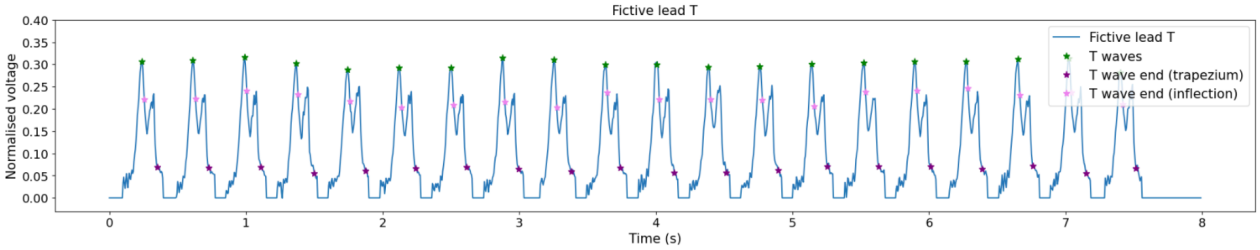


Figure 18: Fictive lead T with detected T waves and T wave end points using the trapezium are method.

P waves

For P wave detection, previous work proposed QRST cancellation, followed by the computation of a new fictive lead and application a prominence threshold based on a Gaussian fit. However, this method is highly sensitive to both noise and residual signal content after QRST cancellation. If the end of the T wave is not accurately detected, the remaining signal may still contain T wave components. As a result, the direction used to compute the fictive lead can become distorted, causing the lead to misalign with the true P wave orientation. This makes the method vulnerable to overlapping signals and noise. An illustration of this issue is provided in Figure 19, where residual T wave signal dominates the fictive lead and causes peak detection errors.

To address this issue, the current study avoids computing a fictive lead for P wave detection. Instead, peaks are identified directly in the original ECG leads I, II, and III. A decision rule is then applied to select the lead with the clearest P wave morphology. Figure 20 shows the improved result on the same signal.

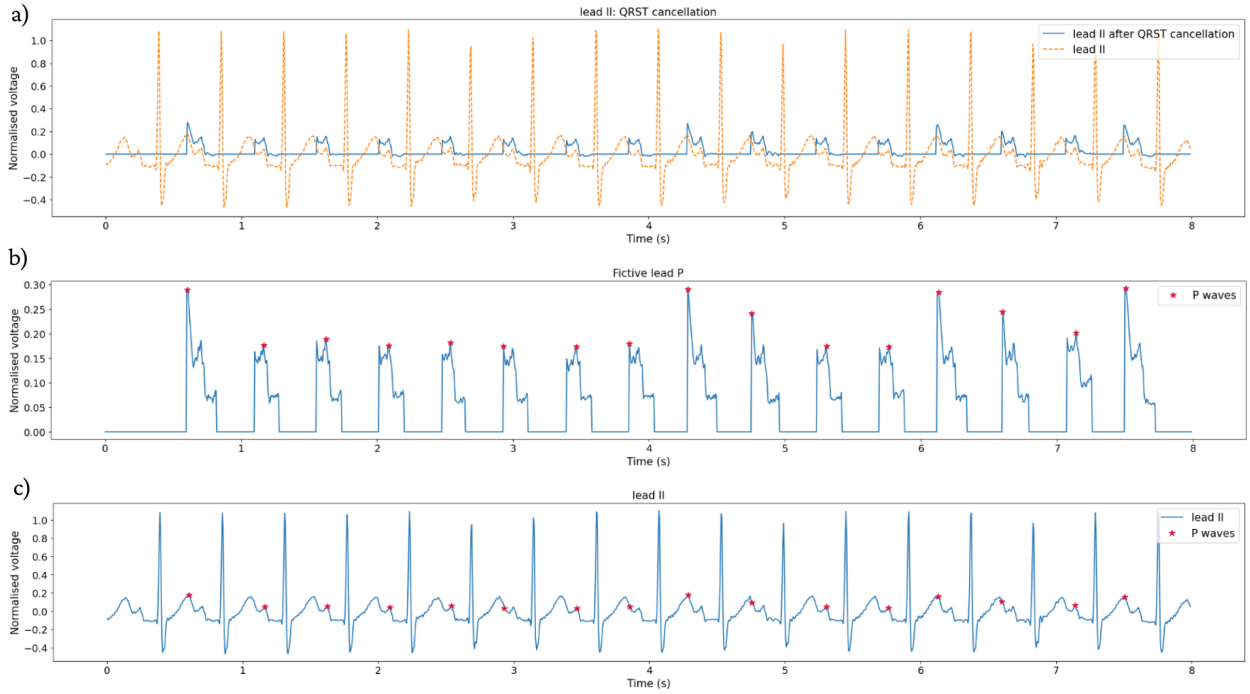


Figure 19: P wave detection using a fictive lead. a) Illustration of QRST cancellation in lead II. b) Fictive lead P with detected P waves based on prominence limit. c) The same detected P waves, demonstrated on lead II.

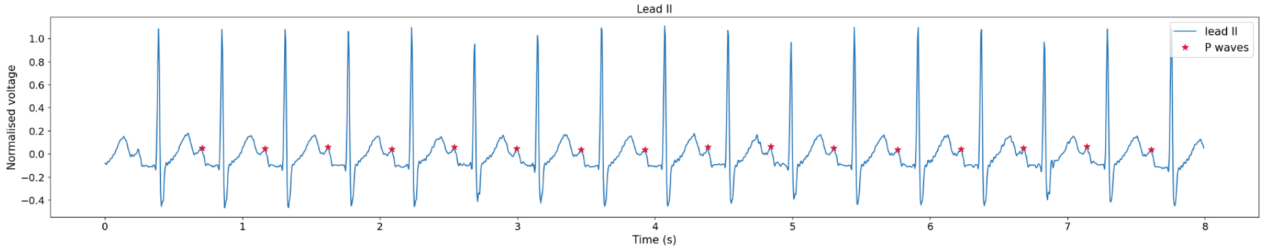


Figure 20: Lead II with detected P waves using this study's method.

A.3 Feature descriptions

Table 6: Overview of features with brief descriptions.

Feature	Description
HR	Heart rate
RR sd	SD of RR intervals
p_valid	number of identified P waves (relative to number of QRS complexes)
pq_var	variance of PQ intervals
p_ang_sd	SD of P wave angles
qrs_mag_sd	SD of dominant vector magnitude of QRS complex
qrs_angle_sd	SD of dominant vector angle of QRS complex
sq_mag_sd	SD of dominant vector magnitude of SQ segment
sq_angle_sd	SD of dominant vector angle of SQ segment
sq_distance	DTW (Euclidean) distance of SQ segment
sq_distance_10	DTW (Euclidean) distance of SQ segment - highest 10% of local distances
qrs_distance	DTW (Euclidean) distance of QRS complex
qrs_distance_10	DTW (Euclidean) distance of QRS complex - highest 10% of local distances
sq_der_distance	DTW derivative-based (Euclidean) distance of SQ segment
sq_der_distance_10	DTW derivative-based (Euclidean) distance of SQ segment - highest 10% of local distances
qrs_der_distance	DTW derivative-based (Euclidean) distance of QRS complex
qrs_der_distance_10	DTW derivative-based (Euclidean) distance of QRS complex - highest 10% of local distances
sq_ang_distance	DTW angular distance of SQ segment
sq_ang_distance_10	DTW angular distance of SQ segment - highest 10% of local distances
qrs_ang_distance	DTW angular distance of QRS complex
qrs_ang_distance_10	DTW angular distance of QRS complex - highest 10% of local distances

A.4 Feature correlations

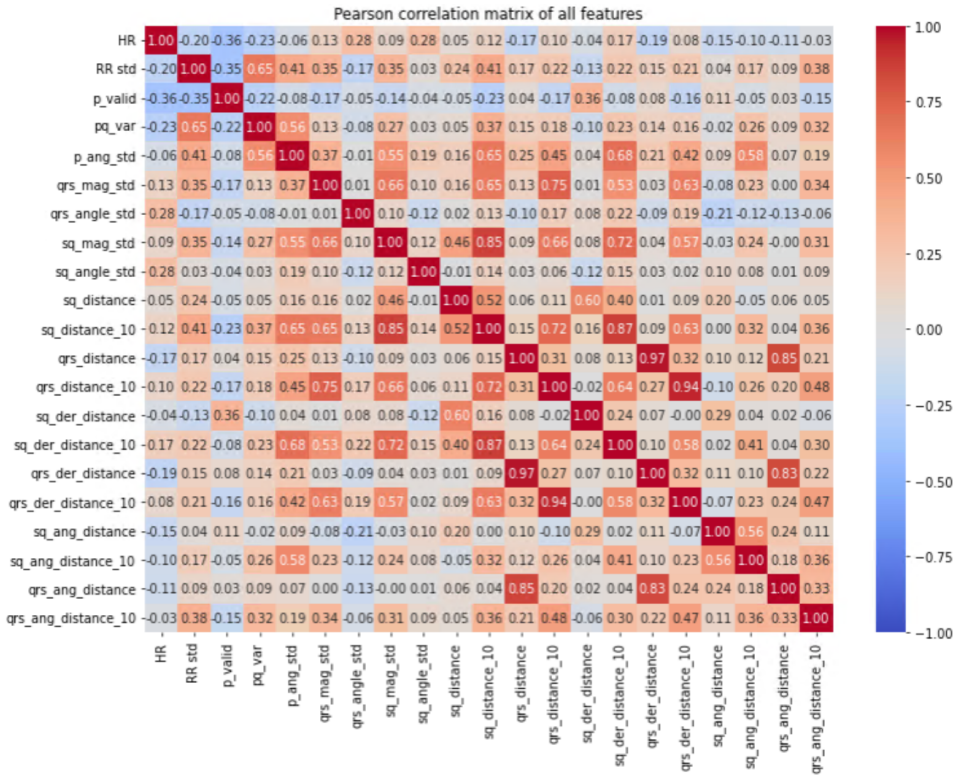


Figure 21: Pearson correlation matrix of all features, when using a window size of 60 s.

A.5 Feature trends

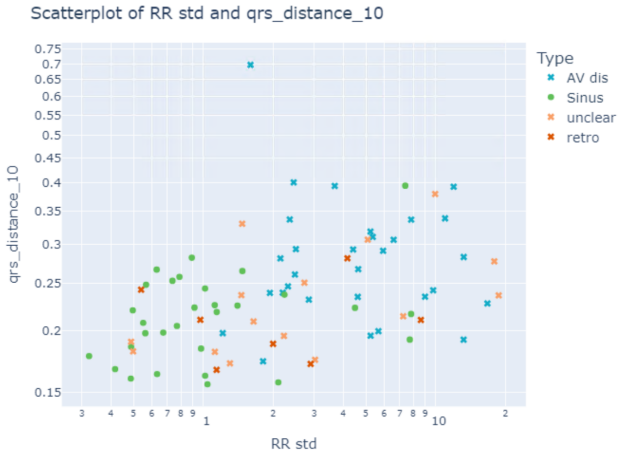


Figure 22: Scatter plot showing the distribution of features RR_sd and $qrs\_distance\_10$ (window size of 60 s). Different subtypes of JET are labelled.

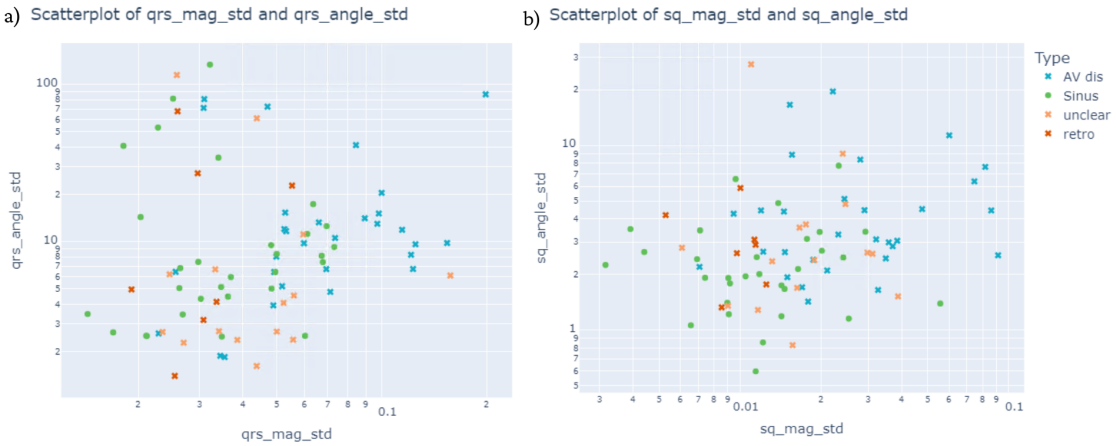


Figure 23: Scatter plots showing the distribution of VCG features of the dominant vector (window size of 60 s). Different subtypes of JET are labelled. a) SD of magnitude and angles of QRS complex. b) SD of magnitude and angles of SQ segment.

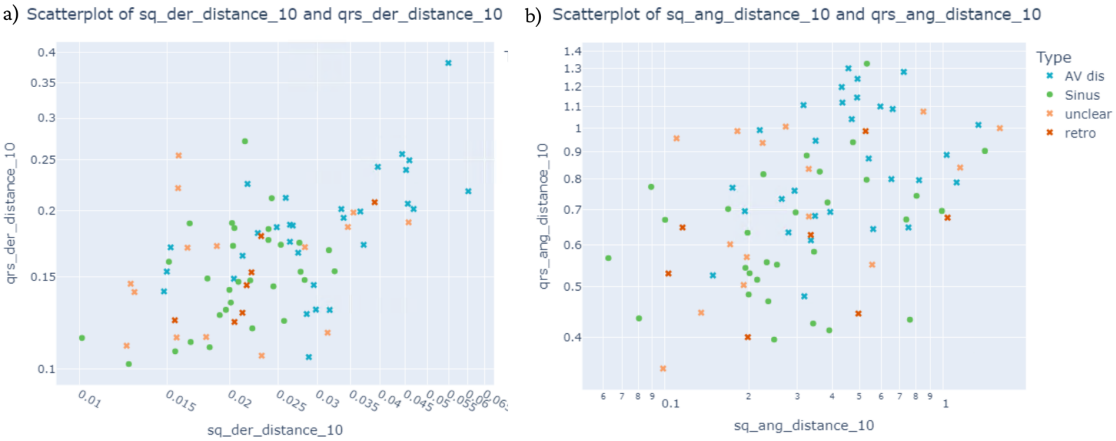


Figure 24: Scatter plots showing the distribution of DTW distance measures of QRS loops (y-axis) and SQ segments (x-axis) (window size of 60 s). Different subtypes of JET are labelled. a) Derivative-based distance b) Angular distance.

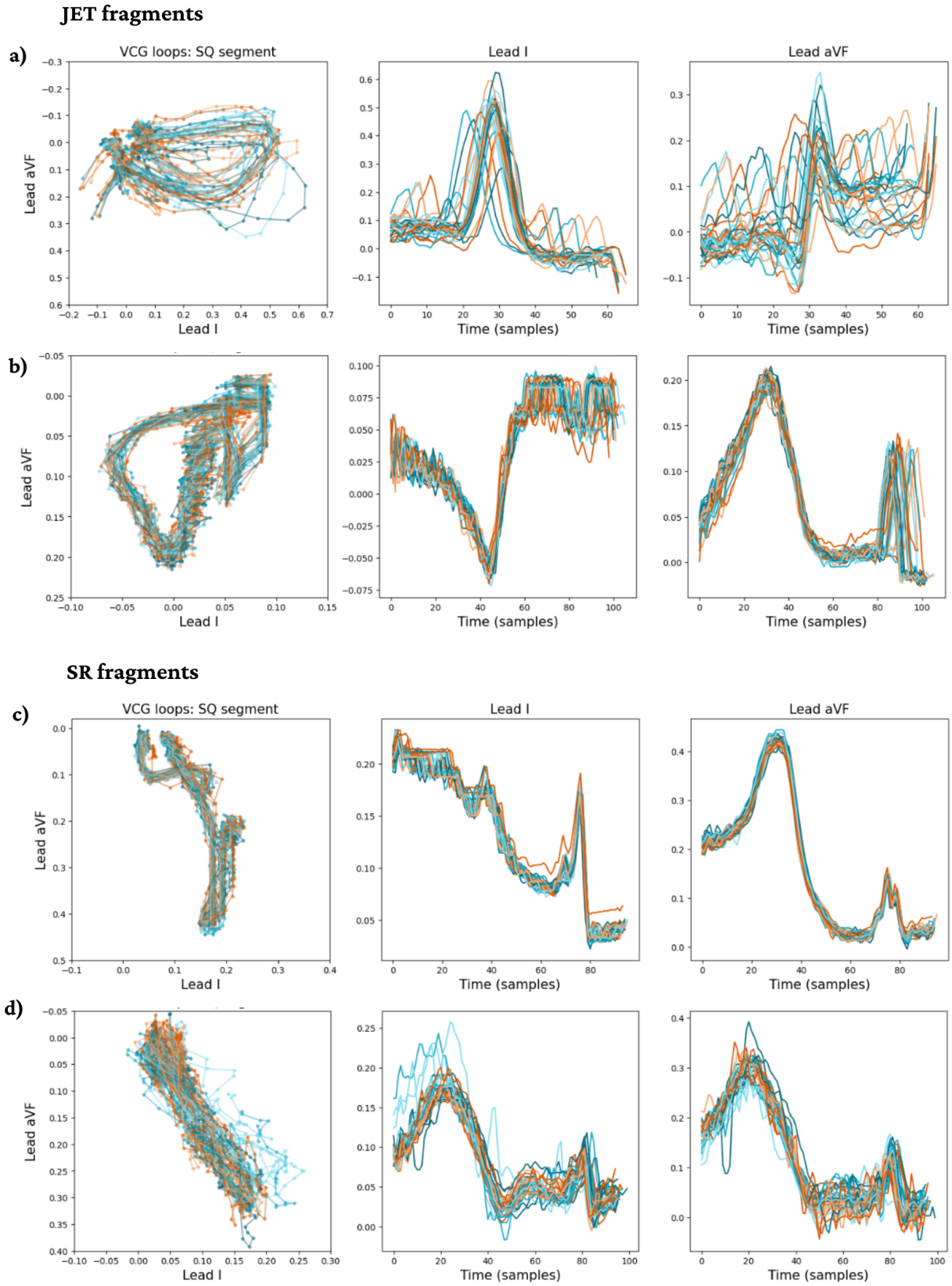


Figure 25: SQ segments of four different data fragments. On the left, the VCG loop is shown. The middle and right figures show lead I and aVF, respectively. The 30 loops, evenly sampled over 60-second windows, are presented in different colours. a) JET with AV dissociation, with high DTW distance. b) JET with AV dissociation, with low DTW distance. c) SR with low DTW distance. d) SR with high DTW distance.

A.6 Model hyperparameters

Table 7: Parameters used for LR model optimisation.

	Parameter	**Values**
Feature selection (RFE)	n_features_to_select	range(1, n_features+1)
LR model	C	[0.01, 0.1, 1, 10, 100]

Table 8: Parameters used for RF model optimisation.

	Parameter	**Values**
Feature selection (SelectFromModel)	threshold	[mean, median, 0.01]
RF model	n_estimators	[10, 50, 100, 200]
	max_depth	range(1,11)

A.7 Feature selection

Table 9: Results of nested cross-validation using a Logistic Regression model, with automatic feature selection (RFE) and hyperparameter tuning (grid search). The mean and SD of the balanced accuracy score and median and range of the number of features across the 5 outer loops are reported. The selected features and their count across these 5 folds are also given.

	Window size (s)			
	60	**30**	**20**	**10**
Balanced accuracy (mean (SD))	0,935 (0,057)	0,935 (0,029)	0,957 (0,023)	0,930 (0,030)
Number of selected features (median (range))	2 (1-5)	2 (1-4)	2 (1-2)	2 (2-6)
Feature	**Times selected (N)**			
p valid	5	5	5	5
pq var	4	4	4	5
sq der dist	1			
sq ang std	1	1		1
sq der dist 10	1			
qrs ang std		1		
RR std				1
qrs dist 10				1
qrs mag std				1

Table 10: Results of nested cross-validation using a Random Forest model, with automatic feature selection (based on feature importances) and hyperparameter tuning (grid search). The mean and SD of the balanced accuracy score and median and range of the number of features across the 5 outer loops are reported. The selected features and their count across these 5 folds are also given.

	Window size (s)			
	60	**30**	**20**	**10**
Balanced accuracy (%) (mean (SD))	0,957 (0,039)	0,952 (0,044)	0,957 (0,029)	0,920 (0,013)
Number of selected features (median (range))	4 (3-8)	3 (3-8)	3 (3-4)	3 (3-16)
Feature	**Times selected (N)**			
p valid	5	5	5	5
pq var	5	5	5	5
RR sd	5	5	5	5
sq der dist	3	1		1
p ang sd	1	1		2
HR	1	2	1	2
qrs dist 10	1			1
qrs ang dist 10	1	1		1
sq ang dist	1			1
sq der dist 10		1		
sq dist 10			1	2
sq ang sd				2
qrs mag sd				2
sq dist				1
qrs ang std				1
qrs dist				1
sq ang dist 10				1

A.8 Final models

Table 11 presents the feature contributions of the final LR and RF model. The left column shows the coefficients from the LR model (interpretable as the direction and relative strength of association), while the right column displays the feature importances as computed by the RF model (indicating the relative contribution to model prediction based on impurity reduction).

Figure 26 illustrates the decision boundary of the LR model.

Table 11: Coefficients and feature importance scores for final LR and RF model.

	LR coefficients	**RF feature importances**
p_valid	-8,53	0,46
pq_var	11,86	0,34
RR_sd	x	0,19

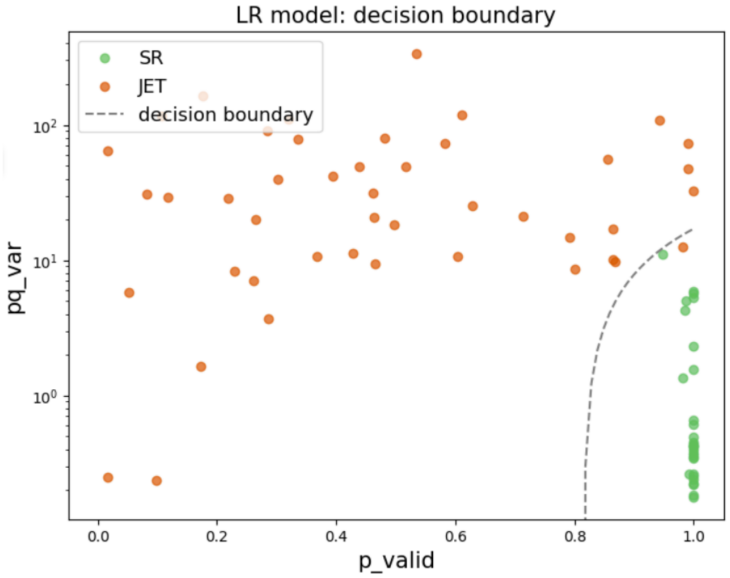


Figure 26: Illustration of decision boundary of LR model.

A.9 Model performance

Table 12: Performance of LR models on the training dataset using the following features: p_valid and pq_var. The optimised hyperparameters are also given.

	Window size (s)			
	60	**30**	**20**	**10**
Balanced accuracy	0,989	0,969	0,976	0,950
AUROC	0,997	0,992	0,994	0,987
FPR (%)	0,0	1,3	2,2	3,1
FNR (%)	2,2	4,9	2,6	6,9
hyperparameters:				
C	100	10	100	10

Table 13: Performance of RF models on the training dataset using the following features: p_valid, pq_var and RR_sd. The optimised hyperparameters are also given.

	Window size (s)			
	60	**30**	**20**	**10**
Balanced accuracy	0,988	0,960	0,958	0,938
AUROC	0,992	0,989	0,993	0,982
FPR (%)	2,5	3,9	5,7	8,0
FNR (%)	0,0	4,1	2,7	4,4
hyperparameters:				
max depth	4	4	5	8
n estimators	100	10	100	50

A.10 Secondary validation: examples of misclassification

Different morphologies

Figure 27 illustrates an example of an SR rhythm, where errors occurred in the detection of Q start. In this signal, the PR interval is very short, and the expected isoelectric line is absent. Since the Q start is identified by locating a point with a low signal derivative, this criterion fails in some instances, causing the algorithm to mistakenly mark the peak of the P wave as the Q start. Consequently, several P waves are missed.

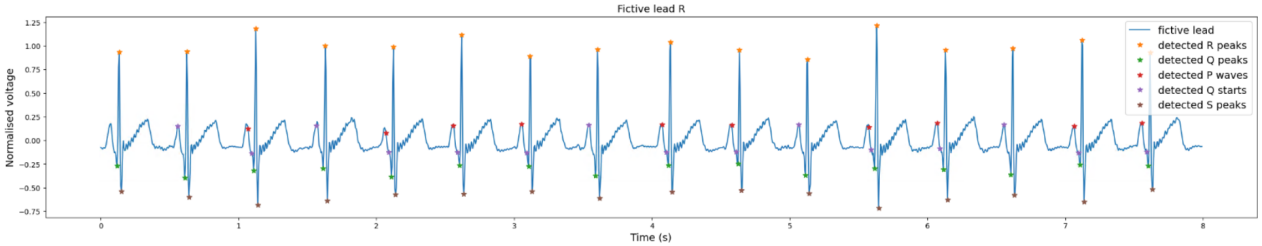


Figure 27: SR fragment illustrating mistakes in Q start detection, resulting in missed P waves.

Figure 28 provides another example of a signal with deviating morphology. T waves are substantially higher than the QRS complexes. Their increased amplitude, combined with their sharp peaks, leads to misidentification as R peaks. As a result, certain peaks actually belonging to the QRS complex are incorrectly identified as P waves. This leads to high p_valid and low pq_var values, resulting in an SR classification.

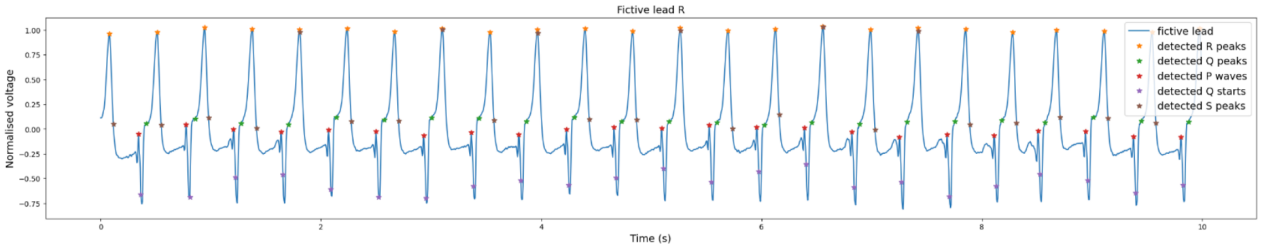


Figure 28: JET fragment. T waves were the most dominant in this signal, and are identified as R peaks by the detection model.

Noise and artefacts

Figure 29 illustrates a noisy SR signal in which several T and P waves are incorrectly detected. In some cases, the T wave is identified too early and subsequently the actual T wave is classified as a P wave. Additionally, the presence of high-frequency noise with relatively large amplitude makes it difficult to accurately identify the P wave. This results in a high pq_var value, leading to a JET classification.

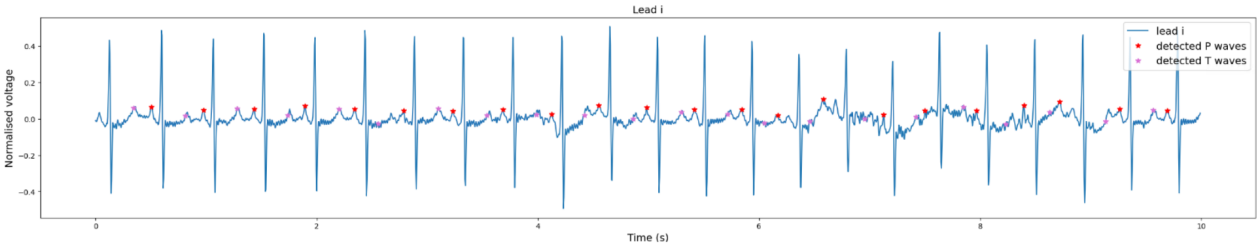


Figure 29: SR fragment with several errors in T and P wave detection, caused by a noisy signal.

Figure 30 shows a signal affected by an artefact, leading to an incorrect R peak detection and two missed R peaks. Elsewhere in the signal, the presence of noise causes some P waves to be detected at incorrect time points. Additionally, two P waves are missed, possibly due to their low prominence.

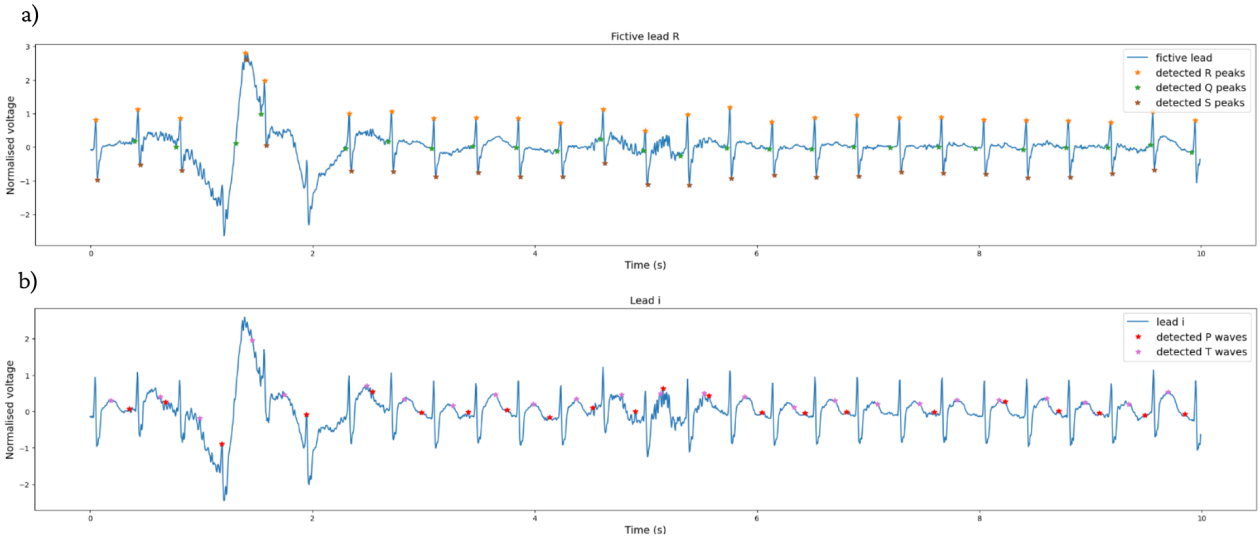


Figure 30: SR fragment illustrating mistakes in R peak detection around an artefact. Additionally, several mistakes in P peak detection are made. a) Fictive lead R with QRS peaks. b) Lead I with T and P waves.

Other arrhythmias

In figure 31, en example is given of a signal with several PVCs. Due to the high amplitude of these PVCs, the fictive lead R is constructed in a direction that does not align with the predominant orientation of the normal QRS complexes. As a result, many R peaks are missed, influencing the detection of the other peaks as well. In this case, the calculated heart rate is extremely low, but since there is only one 'window' where P waves are detected, p_valid is 1 and pq_var is 0, leading to a SR prediction. In case of more than two PVCs, pq_var becomes higher, resulting in a JET prediction.

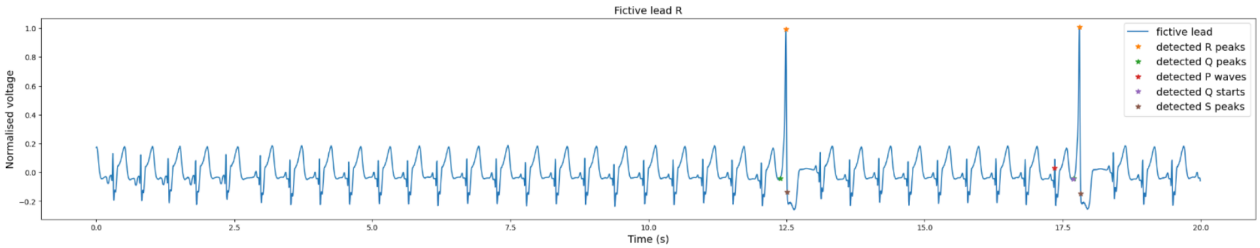


Figure 31: Signal with PVCs, leading to mistakes in R peak detection. Fictive lead R is shown with detected QRS peaks and P waves.

Figure 32 illustrates a SR signal with some extra beats, which may be nodal or atrial beats. This causes several absent P waves and some irregularity. In this case, it only slightly affects the feature values and the predicted class is SR.

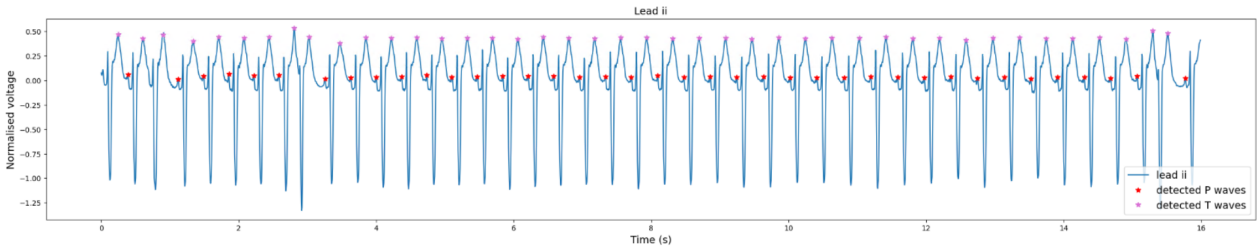


Figure 32: SR signal with some extra beats, either of nodal or atrial origin.

In figure 33, a rhythm is shown which changes into an AET, where the direction of the atrial depolarization changes. It causes an increase in the value of pq_var and leads to a JET prediction.

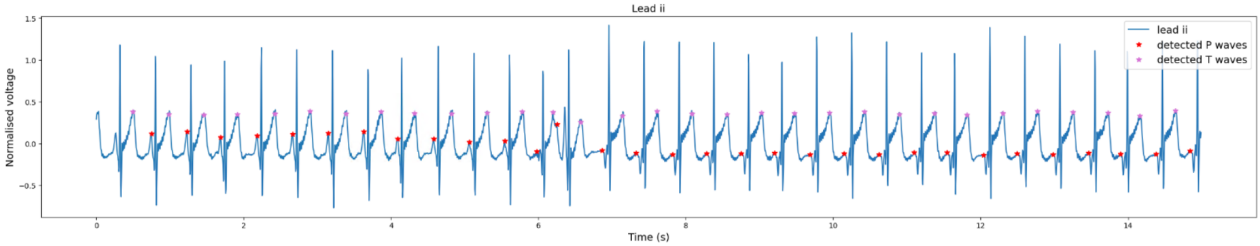


Figure 33: Signal converting to an atrial rhythm.