scientific reports

Check for updates

OPEN

Analyzing the performance of biomedical time-series segmentation with electrophysiology data

Richard Redina[1,2]✉, Jakub Hejc[2,3], Marina Filipenska[1] & Zdenek Starek[2,4]

Accurate segmentation of biomedical time-series, such as intracardiac electrograms, is vital for understanding physiological states and supporting clinical interventions. Traditional rule-based and feature engineering approaches often struggle with complex clinical patterns and noise. Recent deep learning advancements offer solutions, showing various benefits and drawbacks in segmentation tasks. This study evaluates five segmentation algorithms, from traditional rule-based methods to advanced deep learning models, using a unique clinical dataset of intracardiac signals from 100 patients. We compared a rule-based method, a support vector machine (SVM), fully convolutional semantic neural network (UNet), region proposal network (Faster R-CNN), and recurrent neural network for electrocardiographic signals (DENS-ECG). Notably, Faster R-CNN has never been applied to 1D signals segmentation before. Each model underwent Bayesian optimization to minimize hyperparameter bias. Results indicated that deep learning models outperformed traditional methods, with UNet achieving the highest segmentation score of 88.9 % (root mean square errors for onset and offset of 8.43 ms and 7.49 ms), closely followed by DENS-ECG at 87.8 %. Faster R-CNN and SVM showed moderate performance, while the rule-based method had the lowest accuracy (77.7 %). UNet and DENS-ECG excelled in capturing detailed features and handling noise, highlighting their potential for clinical application. Despite greater computational demands, their superior performance and diagnostic potential support further exploration in biomedical time-series analysis.

Keywords Time-series Segmentation, Electrophysiology Study, Rule-based Delineation, Support Vector Machines, U-Net, Faster R-CNN, DENS-ECG

The analysis of biomedical time-series data, such as electrocardiograms (ECG), intracardiac electrograms (EGM), electroencephalograms (EEG) and electromyograms (EMG), plays a critical role in clinical diagnosis and monitoring. A crucial task in this domain is to segment these signals into distinct, interpretable segments, each representing a specific physiological state or event. Accurate delineation of segment boundaries is vital for extracting meaningful features and temporal patterns. This facilitates the identification of abnormal conditions and the assessment of treatment efficacy[1].

Reproducible segmentation is particularly crucial in scenarios that require real-time or millisecond-level precision analysis, such as during clinical procedures or emergency diagnostics. For example, in the analysis of ECGs, segmentation helps identify key cardiac events such as the P-wave, the QRS complex, and the T-wave, which are essential for diagnosing conditions such as arrhythmias and myocardial infarctions[2–4]. Similarly, EGM segmentation helps detect conduction abnormalities[5–7], map arrhythmia mechanisms[8,9], and guide subsequent interventions[10,11] during electrophysiology studies. Likewise, EEG segmentation is used to detect abnormal brain activity[12,13], while EMG segmentation helps to understand motor control in neuromuscular disorders[14,15].

The interpretation of biomedical time series can be complex. Extraneous signals such as noise, movement, or motion artifacts, and far-field potentials of adjacent tissues can affect the precision of the measurement[16,17]. These issues often necessitate manual annotation by experts. However, manual annotation is time-consuming, labor-

[1]Department of Biomedical Engineering, Brno University of Technology, Technicka 12, Brno 61600, Czech Republic. [2]International Clinical Research Centre, St. Anna's Faculty Hospital, Pekarska 53, Brno 60200, Czech Republic. [3]Children's Hospital, Faculty Hospital Brno, Cernopolni 9, Brno 61300, Czech Republic. [4]First Department of Internal Medicine/Cardioangiology, St. Anne's Hospital, Masaryk University, Pekarska 664/53, Brno 60200, Czech Republic. ✉email: 195715@vut.cz

intensive and prone to variability between observers[18,19]. To address these limitations, automated segmentation algorithms have been developed. These algorithms leverage signal processing and machine learning techniques to automatically detect transitions and segment boundaries. By reducing the reliance on human intervention, automated segmentation has the potential to improve efficiency, improve diagnostic precision, and support personalized medicine[20]. Specifically, artificial intelligence (AI) and deep learning methods have shown promise in automating these tasks with high accuracy[21].

This study aims to compare the performance of various segmentation methods, including threshold-based techniques, statistical models, and AI-based approaches. We will apply these methods to real-world intracardiac data obtained during electrophysiology studies, a vital procedure to understand the electrical properties of the heart[5,7]. In addition, we will explore the challenges and limitations related to reproducibility, especially in the presence of variable noise levels or superimposed electrical sources, both of which are commonly encountered in electrophysiology measurements[22].

Although there are several automated annotation methods[2,23–25], comprehensive investigations are lacking to evaluate detection performance and reproducibility in complex, high-noise scenarios. This is particularly relevant given the rapid advances in AI-driven analysis of complex biomedical data, propelling the clinical sector toward digital medicine. Although modern signal processing techniques can mitigate some limitations of manual annotation, extensive validation across diverse electrophysiological patterns is crucial to ensure reliability and clinical utility.

Methods

Experimental design

This is a comparative study performed on the retrospective single-platform intracardiac data from one healthcare centre. Performance and robustness in the presence of noise of atrial activity detection and delineation achieved by hand-crafted rules, support vector machine (SVM) and different architectures of deep convolutional neural networks (CNN) were evaluated and compared to each other. Each method was compared to achieve the most efficient design independent of human decision making.

Data

A database[26] of intracardiac signals obtained from 100 consecutive pediatric patients (median age 14 years, 52 men) indicated for catheter ablation was used at the Children's Hospital, University Hospital Brno, Brno, Czech Republic for algorithm training and evaluation. The local ethics review committee approved the use of the data for this study. Database consists of 326 preselected segments (8.5 s median length, IQR 6.4-12.2 s) including sinus rhythms (58.6 %), atrial (14.4 %) and ventricular (12.3 %) premature beats, AV node (13.8 %) and AV (12.3 %) re-entry tachycardia, atrial fibrillation (10.1 %), junctional rhythm (6.8 %), artificial pacing (16.6 %), ventricular pre-excitation syndrome (17.8 %), and others. The segment inclusion criteria were the presence of sinus rhythm, ongoing arrhythmia, irregular rhythm as a result of procedural intervention, and rhythms from artificial pacing.

Individual recordings consist of 5-lead bipolar intracardiac signals recorded by a deca-polar diagnostic catheter placed within the coronary sinus (CS) as a standard part of the electrophysiology procedure. Intracardiac signals were recorded using the WorkMateTM System 4.2 (Abbott Laboratories, Abbott Park, IL, USA) and exported in raw unfiltered form by custom software tool[27]. Each atrial electrogram was visually identified and manually delineated by an experienced electrophysiology expert. Considering the cost associated with manual annotation, the onset of the atrial complex was annotated as the earliest activation time among all CS channels. The offset of atrial activity was determined as the deactivation time of the latest electrogram. The interval between onset and offset is referred to as the local activation time (LAT). Each atrial electrogram was categorized into one of four groups: a) normal sinus beat (SB); b) abnormal atrial beat (AB); c) atrial fibrillation (AF); and d) stimulated beat (ST).

The recordings were randomly assigned to the training and validation sets (70:30 ratio) using stratification based on patient and arrhythmias to ensure a balanced and non-overlapping split. Eight rhythm categories, representing varying segmentation difficulties, served as strata. This split ratio was chosen heuristically to maximize training examples while maintaining a sufficiently large validation set for reliable performance evaluation, particularly given that the smallest rhythm group contained only 19 examples. A sensitivity analysis further validated this choice. The same randomization was used for all the methods studied. In machine learning approaches, the voltage scale of the electrograms was standardized beforehand using a Min-Max scaling. System-specific pre-processing steps, such as signal filtering and smoothing, are described in the respective sections below.

Rule-based method

Rule-based systems utilize domain-specific knowledge to extract the most discriminative electrogram characteristics, in which the local activation time can be measured by identifying a value falling below a predetermined voltage threshold. Due to its simplicity and computational efficiency, this approach is widely prevalent in current electrophysiology acquisition systems[28], often using a complex set of rules to eliminate diagnostically problematic electrograms. To suppress unwanted noise and, conversely, highlight extracted features, we have used the wavelet transform[29]), which is a well established method in the processing of various biomedical time series[30].

Specifically, we applied a single Daubechies 3 (db3) mother wavelet[31] with a pass-band frequency ranging from 125 Hz to 250 Hz to each of the five channels independently. Fundamentally, the transformed output can be understood as a correlation between the shape of the wavelet function and the local electrogram morphology, suggesting that numerous configurations of the method are possible. The selection of the appropriate wavelet

and other parameters is elaborated thoroughly in the "Optimal Design Selection" section. Subsequently, we estimate the smoothed nonnegative envelope[32] of the filtered signals using the Hilbert transform[33], followed by a Butterworth lowpass filter (critical frequency of 100 Hz). This step aims to reduce the number of oscillations and local extremes.

We employ peak-over-threshold (POT) and voltage-over-threshold (VOT) methods to localize the atrial complex and detect its boundaries within the transformed electrogram representation. The localization threshold in POT adapts dynamically based on the arithmetic mean of the five most recent peak magnitudes. The thresholds for detecting local activation and deactivation times in VOT were established by estimating the electrical baseline noise level.

Support vector machine with wavelet decomposition

The Support Vector Machine (SVM)[34] is a prominent example of supervised machine learning methods widely applied in various high-dimensional classification problems[35–37], including the automated analysis of electrocardiographic time-series[38]. Unlike rule-based approaches, SVM establishes its decision-making process through automated optimization, usually forming a hyperplane that maximizes the separation among distinct object groups, as exemplified in our case by atrial electrograms, ventricular far-field potentials, and noise. For implementation, we have chosen a linear variant of SVM whose main advantage is simple implementation in medical devices.

SVM requires handcrafted features that accurately represent the electrogram. In this context, we have opted to employ a similar feature extraction pipeline as for the rule-based method, involving the Symlet 2 (sym2) wavelet function[31] for wavelet decomposition and the estimation of the non-negative signal envelope. This process encompassed the decomposition of the electrogram into multiple representations with varying temporal and frequency resolutions, achieved through octave filter banks with pass-band frequency from 62 to 125 Hz and 125 to 250 Hz, respectively. The number of representations used was optimized as described in the following section. The final step removed detections smaller than the optimized boundary. The SVM's decision boundaries for one-hot decoding were trained on electrogram features from all five channels.

AI-driven approaches

Compared to previous approaches, deep learning methods[39] are capable of obtaining an optimal representation of highly complex data at various levels of abstraction through automated extraction of relevant features. This capability reduces the impact of inappropriate preprocessing resulting from a lack of domain knowledge. In the last decade, these methods have achieved significant success in areas such as image recognition[40,41], speech recognition[42], prediction of gene expression[43], in-silico drug modeling[44], and also analysis of electrocardiography data[45,46]. For comparison purposes, we have selected three architectures that are widely used in current deep learning research[4,47,48].

1D U-Net

The model used in this work is based on the 1D U-Net convolutional neural network (CNN)[49]) and was previously described in[47]. It combines low- and high-resolution features through a top-down pathway and lateral connections. To capture long-term signal variations and the surrounding context, we added an intermediate layer with dilated convolution kernels between the decoder and the last encoding stage, following the approach of[50]. The model accepts each electrogram channel as a separate feature, preprocessed only by filtering-out (5Hz high-pass filter) baseline wander, which is in a similar manner applied by all EP acquisition systems.

For the U-Net optimization, we used the exponentially-weighted Tversky loss function[51]. The function comprises three modulating parameters (α, β, and γ) that facilitate the discovery of an optimal problem-specific solution by rebalancing precision, recall, and the contribution of true positives difficult to identify. The model consists of a total of 492,332 trainable parameters. The architecture and optimization design choices are provided in detail in the supplementary materials.

Faster R-CNN

To transition from semantic segmentation (each signal sample is classified) to instance segmentation (contiguous signal segments are classified as distinct events), we adapted the Faster R-CNN architecture[48] for multichannel electrograms. The encoding part is, similarly to 1D U-Net, composed of residual layers with reducing temporal dimensions. By allowing for the localization and refinement of individual complex boundaries instead of per-sample classification, this approach may potentially improve the delineation accuracy. The model is optimised employing distinct loss functions: smooth L1 loss is used to regress the instance boundaries and binary cross-entropy is used to differentiate between signal instances and the background.

To mitigate the issue of less accurate boundary estimation in challenging positive cases, we improved the smooth loss of L1 with a performance-sensitive weight inspired by the Tversky loss. This modification penalises each positive instance proportionally to the complement of its Intersection over Union (IoU) with the ground truth. The exponent γ is a hyperparameter controlling the degree of penalization, allowing for flexible adjustment to balance the importance of accurate boundary estimation with overall model performance. The architecture and optimization design choices are provided in detail in the supplementary materials.

DENS-ECG

DENS-ECG was originally proposed by[4] for the segmentation of surface ECGs and was adapted for EGM analysis by our team. The architecture combines CNN with Gated Recurrent Units (GRU) aimed at processing long-term dependencies present in the sequential data. Unlike the U-Net, CNN extracts highly granular features corresponding to the original sampling frequency of the EGM signal by not incorporating any dimension

reduction and lateral connections in the top-down pathway, providing a straightforward way for the original data to final classification. The LSTM unit then generates high-level features combining morphological and temporal information to correctly identify EGM complexes.

To optimize the model, we used the same loss function as U-Net, with hyperparameters α, β, and γ. The final architecture, consisting of 1,069,953 trainable parameters, is detailed in supplementary materials.

Optimal design selection

Signal processing methods often require intricate parameter tuning, which can significantly impact their performance. In this study, Bayesian optimization (BO)[52] was used to determine the optimal parameters for each method, ensuring a fair and thorough comparison. BO is a model-based optimization approach that iteratively updates a surrogate model to guide the search for optimal hyperparameters efficiently (see Fig. 1).

For the wavelet method, we optimized the wavelet type, decomposition level, frequency bands, smoothing filter cutoff, and thresholds used in the rule-based system. In SVM, we focus on refining parameters such as the learning rate and the decision boundary margin. For U-Net, Faster R-CNN, and DenseNet, we fine-tuned both architecture and training parameters, including the number of layers, receptive field size, filter cardinality, learning rate, batch size, and Tversky loss hyperparameters. For Faster R-CNN to mitigate the class imbalance and improve training convergence, we optimized the ratio of negative to positive anchor boxes employed in background sampling using Bayesian optimization. We also rebalanced the cross-entropy loss using the same factor to compensate for the underrepresentation of positive samples. For a full list of optimized hyperparameters and search spaces, refer to the supplementary material. Optimization was performed using the Optuna 3.0.0 package[53].

Evaluation metrics

Only electrograms without the presence of atrial fibrillation (Non-AFib) were considered for evaluation. The algorithms were evaluated using two distinct approaches. Results were analysed both sample-wise, evaluating correctness of classification of each signal sample, and instance-wise, assessing deviation of activation measurement for each instance of atrial activity from the reference annotation. Sample-wise evaluation metrics included the overall Dice score, recall (Se), and precision (PPV) defined by the following equations:

$$Se = \frac{TP}{TP + FN}, \tag{1}$$

$$PPV = \frac{TP}{TP + FP}, \tag{2}$$

$$Dicescore = \frac{2 * TP}{2 * TP + FN + FP}, \tag{3}$$

where TP is the number of correctly detected samples in the signal, FP is the number of samples where atrial activity was detected despite not being present in the signal, and FN is the number of undetected samples with atrial activities that were present in the signal. The mean deviation error (ME) and the root mean square error (RMSE) were used to assess the onset and offset delineation performance and can be defined as

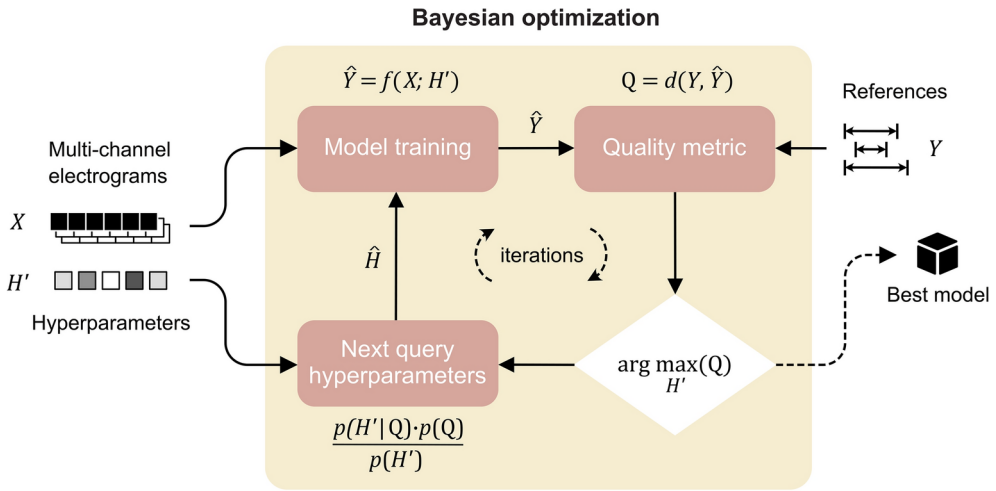

Fig. 1. The Bayesian optimization workflow. Function f(x; P) represents any of the analysed methods, each involving its unique intrinsic set of hyperparameters denoted as P. P is iteratively optimised to maximise quality metric Q representing the harmonic mean of precision and recall (Dice score). Q is constrained to a range between 0 and 1.

$$ME = \frac{1}{N}\sum_{i=0}^{N}(y_i' - y_i), \tag{4}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=0}^{N}(y_i' - y_i)^2}, \tag{5}$$

where N is the number of detected atrial beats, y_i is the ground truth time mark of onset or offset of the beat, and y_i' is the estimated time mark by the algorithm. Although positive and negative errors can cancel out in the mean error metric, it helps to assess whether the system tends to detect events earlier or later than their true occurrence. However, the RMSE metric provides additional information on the magnitude of these deviations.

Statistical analysis

The statistical analysis was performed on deviations in atrial activity onset and offset. First, we invoke the central limit theorem[54] supported by Berry-Esseen[55] bounds below 0.05, given the hundreds of LAT samples compared in this study. Consequently, we employed a two-sample F-test for equal variances[56] to compare the error value variances across different methods. The null hypothesis for this test was formulated as one-tailed to guide our analysis. In all statistical tests, a P-value of ≤ 0.05 was considered statistically significant. The analyzes were performed using MATLAB R2023a.

Results

This section provides a detailed analysis of five algorithms, focusing on their performance in the segmentation and measurement of regular atrial electrograms. Performance was tested using the validation subset. RMSE was calculated for all five methods for all electrograms, as well as separately for sinus atrial beats (SB), abnormal atrial beats (AB), and complexes resulting from artificial atrial pacing (ST). The following findings elucidate the practical capabilities and limitations of each algorithm across various signal types and under diverse conditions, such as the presence of artificial noise.

Sample-wise and instance-wise performance

Table 1 compares the results of sample-wise segmentation for each method. The U-Net neural network dominates in the overall results represented by the Dice score with 88.9 % compared to other AI driven approaches and followed by SVM Rule-base and systems, respectively.

Figures 2 to 5 illustrate the performance of our models on representative signals from the validation data set. Chosen scenarios are commonly encountered during electrophysiological procedures and some of them could severely impact segmentation quality.

In Figure 2, U-Net and SVM demonstrate higher reproducibility of the beat-to-beat measurement during sinus rhythm, compared to the Rule-based method. Fig. 3 shows a misclassification of the ventricular far-field potential performed by the SVM. When the magnitude of the ventricular far field becomes more pronounced, a similar issue can also be observed for Rule-based processing, as is demonstrated in Figs. 4 and 5. Superimposed complexes are common causes of both detection failure and delineation failure. We demonstrate the behavior in such cases on electrograms with pacing artifacts (Fig. 4).

Table 2 summarises the onset and offset measurement errors. U-Net and DenseNet consistently outperforms SVM, rule-based methods and faster R-CNN. The ST group exhibits higher RMSE values, likely due to altered morphology of the atrial complex after pace.

Figure 6 illustrates the distributions of detection errors for each algorithm. The dispersion in detection errors estimated by U-Net is significantly lower compared to SVM (onset and offset p-value < 0.001), rule-based method (onset and offset p-value < 0.001) and faster R-CNN (onset and offset p-value < 0.001), indicating a consistently higher precision. The last method - DenseNet - achieved similar variability in detection error (onset and offset p-value > 0.05).

Robustness to artificial noise

Resilience to various noise scenarios is an essential property of any signal processing method. This section evaluates the performance of the algorithms in electrograms with different signal quality. To assess performance under different noise conditions, we artificially generated Gaussian white noise (GWN) and 50 Hz power line interference (PLI_{50}) to modulate the signal-to-noise ratio (SNR) of the signals to 20, 15, 10, 5, and 0 dB.

	Dice score [%]	Recall [%]	Precision [%]
Rule-based	77.7	80.4	79.4
SVM + WT	80.7	85.8	78.2
U-Net	**88.9**	**88.3**	**89.4**
Faster R-CNN	85.0	88.3	82.0
DENS-ECG	87.8	86.4	89.3

Table 1. Sample-wise evaluation results for each segmentation method.

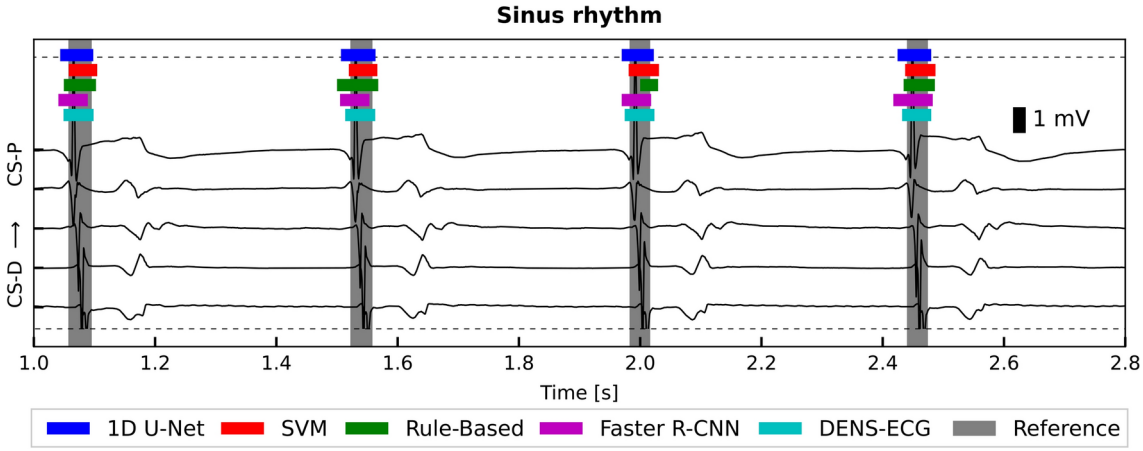

Fig. 2. Comparison of B-EGM segmentation performance for sinus rhythm. The figure illustrates the detection and delineation of atrial complexes in the presence of normal atrial rhythm and ventricular far-field potentials. The ground truth segments are marked with gray rectangles. The color-coded rectangles represent the segmentation results from the rule-based (green), support vector machine (red), and deep neural network methods based on U-Net (blue), R-CNN (violet), and DENS-ECG (cyan), respectively. The SVM and DNN methods demonstrated superior inter-beat performance in accurately identifying complex boundaries, which is essential for reproducibility of diagnostic protocols. Signal clipping marked by dashed lines was applied to prevent a high voltage signals from overflowing outside the plotted area.

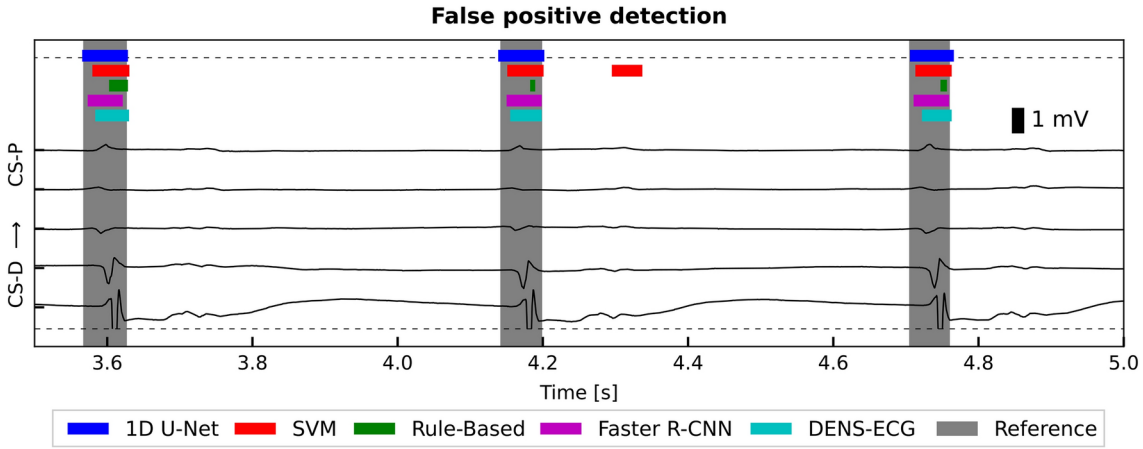

Fig. 3. Comparison of B-EGM segmentation performance for sinus rhythm with false positive detection. The figure illustrates the detection and delineation of low-amplitude atrial complexes recorded by catheter with decreased tissue-electrode contact. The DNN methods demonstrated superior inter-beat performance, while SVM falsely misclassified ventricular far-field potential as atrial complex. The ground truth segments are marked with gray rectangles. The color-coded rectangles represent the segmentation results from the rule-based (green), support vector machine (red), and deep neural network methods based on U-Net (blue), R-CNN (violet), and DENS-ECG (cyan), respectively. Signal clipping marked by dashed lines was applied to prevent a high voltage signals from overflowing outside the plotted area.

GWN represents stochastic processes of diverse origins, typically introduced by measurement instrumentation or radiofrequency current application during catheter ablation. PLI_{50} originates from the vicinity of the electromagnetic field of the power grid. Both sources were applied independently.

Figures 7 and 8 represent performance metrics in response to increasing noise levels (decreasing SNR). AI-driven approaches exhibit minimal metric variation for both types of interference, with a drop in the Dice score of only 5 % except in scenarios involving highly noisy signals. In contrast, the other two methods are more susceptible to interference. The Dice score for the rule-based method drops to almost zero for GWN at higher noise levels, while the Dice score for the SVM-based method drops by 10 % with PLI_{50}. The Rule-based method remains unaffected by PLI_{50} as the BO selected wavelet bands that exclude 50 Hz, whereas SVM does not inherently filter this frequency.

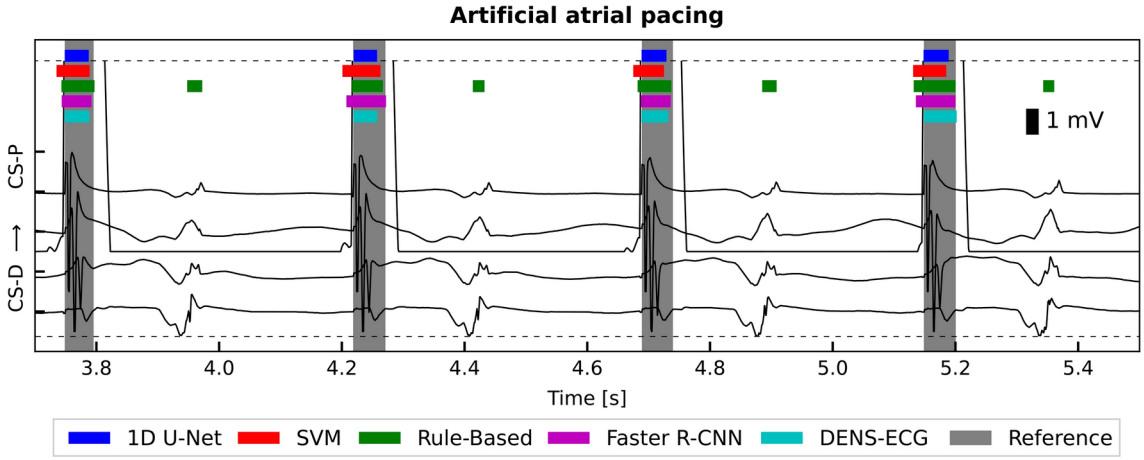

Fig. 4. Comparison of B-EGM segmentation performance for paced rhythm. The figure illustrates the detection and delineation of high-amplitude ventricular far-field potentials and atrial complexes fused with pacing artifacts due to the artificially paced rhythm. The SVM and R-CNN methods overestimated the beginning of the complex due to the presence of pacing artifacts. The rule-based method also mistakes prominent far-field activity as an atrial complexes. The ground truth segments are marked with gray rectangles. The color-coded rectangles represent the segmentation results from the rule-based (green), support vector machine (red), and deep neural network methods based on U-Net (blue), R-CNN (violet), and DENS-ECG (cyan), respectively. Signal clipping marked by dashed lines was applied to prevent a high voltage signals from overflowing outside the plotted area.

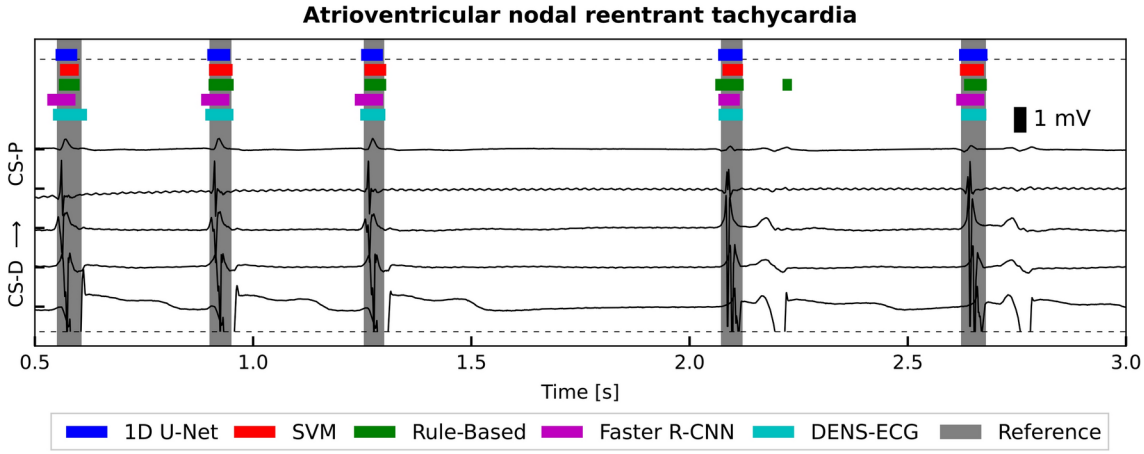

Fig. 5. Comparison of B-EGM segmentation performance for reentry tachycardia. The figure illustrates the detection and delineation of superimposed atrial complexes and ventricular far-field potentials (first three complexes) during an ongoing tachycardia. The ground truth segments are marked with gray rectangles. The color-coded rectangles represent the segmentation results from the rule-based (green), support vector machine (red), and deep neural network methods based on U-Net (blue), R-CNN (violet), and DENS-ECG (cyan), respectively. Each method demonstrated capability of identifying superimposed atrial complexes, which is essential for interval measurement and arrhythmia mapping during electrophysiology procedure. Signal clipping marked by dashed lines was applied to prevent a high voltage signals from overflowing outside the plotted area.

Figures 9 and 10 illustrate the progression of the ME and RMSE with respect to the SNR. Figure 9 reveals the vulnerability of the rule-based method to GWN in both metrics. RMSE systematically increases up to more than 15 ms after reaching a critical level of 20 dB, while showing inconsistent bias across different SNR levels with elevated ME in approximately half of cases above 10 ms and below -5 ms for atrial beat onsets and offsets, respectively. Both SVM and AI models show relatively stable performance up to an SNR of 10 dB. From 10 dB lower, the bias represented by ME gradually shifts by approximately 3 ms for all methods, while RMSE for the onset and offset is relatively stable in SVM, and almost doubles in the case of the U-Net model.

	$RMSE_{onset}$ [ms]				$RMSE_{offset}$ [ms]			
Experiment	**All**	**SB**	**AB**	**ST**	**All**	**SB**	**AB**	**ST**
Rule-based	14.52	14.39	14.25	16.39	10.39	10.83	9.13	12.82
SVM + WT	11.53	10.97	10.39	18.19	10.61	10.00	11.57	9.89
U-Net	8.43	8.23	**8.33**	11.61	**7.49**	5.98	**8.92**	**8.86**
Faster R-CNN	12.69	11.47	13.91	11.31	12.62	10.58	14.45	11.01
DENS-ECG	**8.37**	**7.83**	8.70	**9.79**	7.90	**5.39**	9.17	12.8

Table 2. Instance-wise evaluation results of delineation accuracy for different delineation methods in sinus beat (SB), abnormal beat (AB) and stimulated beat (ST) group. RMSE - root mean squared error.

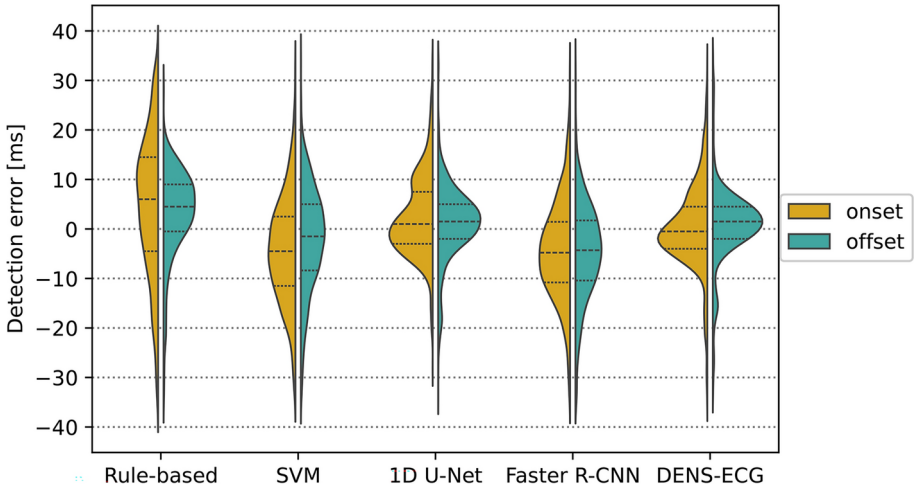

Fig. 6. Detection error distribution. Violin plots showing the distribution of the detection error for the onset (left part) and offset (right part) detection methods. As observed, the U-Net and DenseNet methods exhibit the smallest variability, indicating it is the most stable method for both onset and offset detection.

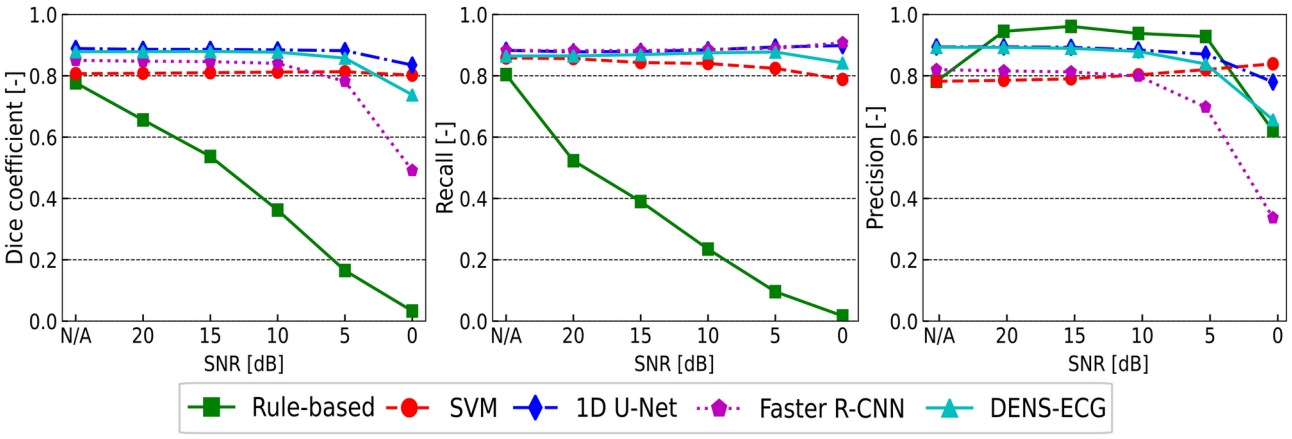

Fig. 7. Models' performance for GWN noise. The trend of the monitored metrics (Dice score, recall and precision) as a function of increasing GWN level for each method. The rule-based method is very sensitive to increasing noise levels and starts to gradually fail even with low noise addition.

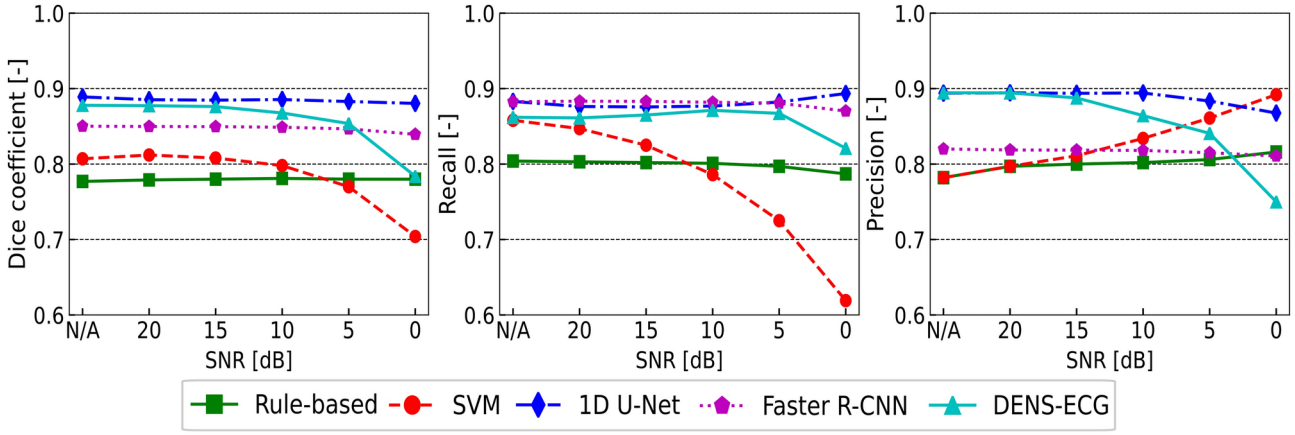

Fig. 8. Models' performance for PLI_{50}**noise.** The trend of the monitored metrics (Dice score, recall and precision) as a function of increasing PLI_{50} level for each method. In this comparison, the SVM method performs worst, as it becomes unreliable at higher noise intensities and fails to correctly detect atrial activity in the signals.

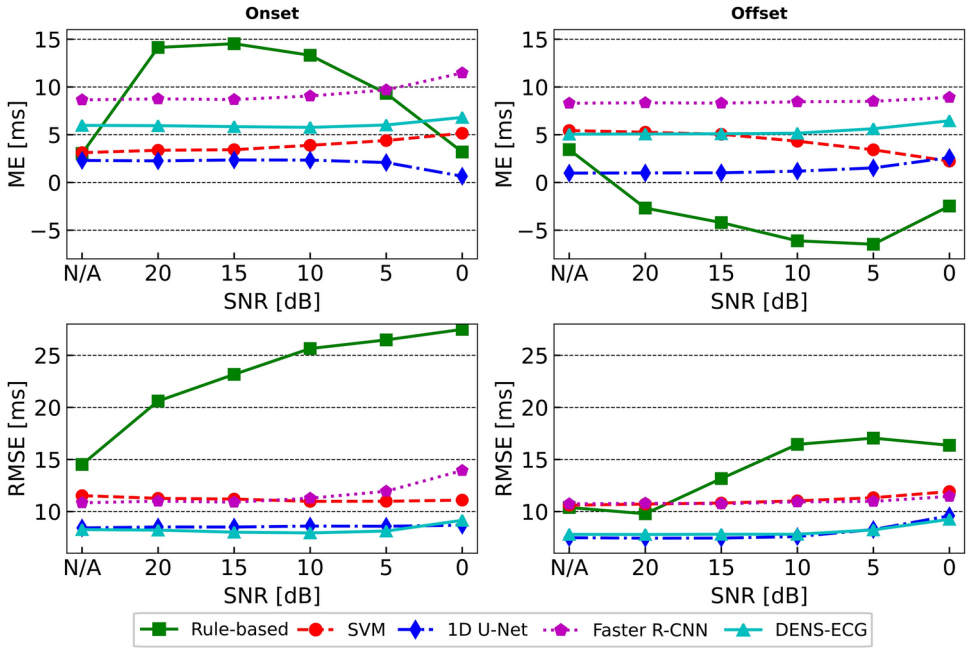

Fig. 9. Models' detection error for GWN noise. The trend of the ME and RMSE as a function of increasing GWN level for each method. Similar to Figure 7, it can be seen that the rule-based method exhibits considerable variability in the resulting delineation with increasing noise intensities.

Figure 10 shows minimal changes in RMSE for the five methods, with the best method (U-Net) achieving a consistent RMSE of around 8 ms at each SNR. The general resilience to PLI_{50} noise is also good for the SVM and rule-based methods, with RMSE remaining stable in the range of 10 to 15 ms even for poor signal quality.

Computing latency

We have evaluated processing times for three distinct EGM segments lasting 1, 10 and 100 seconds. The evaluation was performed on Ubuntu 20.04.5 LTS, 12 core Intel®Xeon®X5650 2.67GHz 12MB cache, 48MB DDR3 ECC with built-in Python 3.8 timer. The general results are reported in table 3 as mean±standard deviation from multiple trials.

For 1 second segments, which is approximately the scenario used in a beat-to-beat analysis, the rule-based method provided an order of magnitude faster processing time reaching 6.62±0.59 ms compared to 45.76±3.48 ms, 32.63±2.80, 30.77±5.76, 204.03±18.39 for SVM, U-Net, Faster R-CNN and DenseNet respectively.

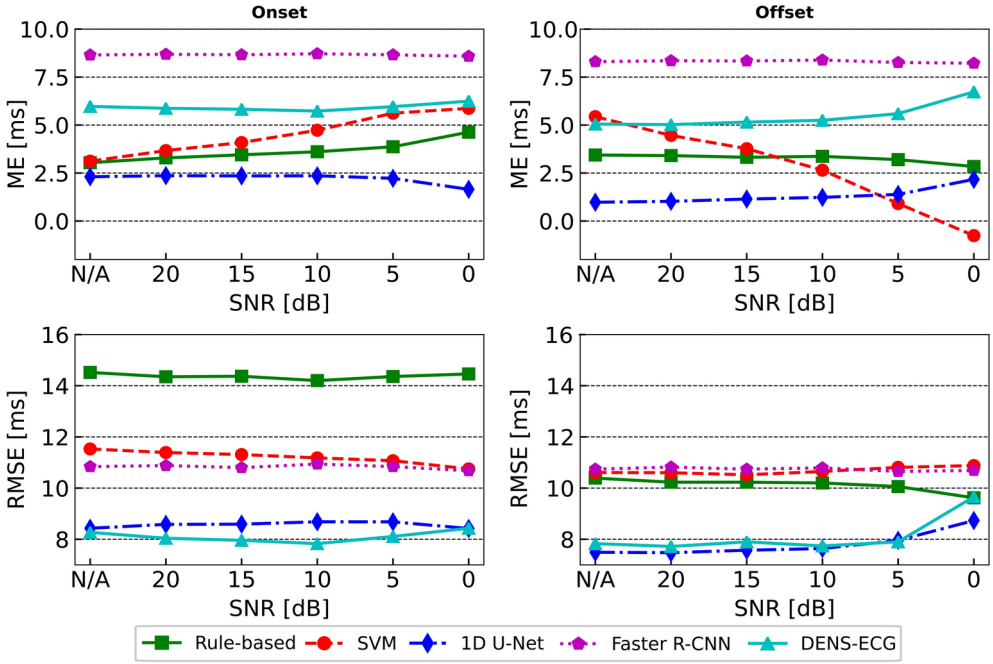

Fig. 10. Models' detection error for PLI_{50}**noise.** The trend of the ME and RMSE as a function of increasing PLI_{50} level for each method. Compared to Figure 9, which shows considerable variability in the outputs of the rule-based method, there is not much change in the SVM model in this case. Thus, in the context of Figure 8, it can be said that the SVM method only has a detection problem in the noisy data, not in the delineation of the atrial activity found.

	Trainable parameters	Processing latency times [ms]		
Method		**B-EGM 1 s**	**B-EGM 10 s**	**B-EGM 100 s**
Rule-based		6.62 ± 0.59	47.89 ± 3.10	457.03 ± 24.71
SVM + WT	11	45.76 ± 3.48	406.64 ± 16.58	3933.41 ± 159.27
U-Net	492,332	32.63 ± 2.80	72.74 ± 6.26	408.18 ± 17.74
Faster R-CNN	1,068,396	30.77 ± 5.76	91.84 ± 7.76	477.61 ± 53.60
DENS-ECG	1,069,953	204.03 ± 18.39	1949.17 ± 97.42	18696.88 ± 607.96

Table 3. Comparison of processing latency times for rule-based method, SVM and U-Net employing input signals of various lengths. Reported results are averages of 326 runs using randomly selected EGM segments.

For processing longer EGM sequences, the U-Net method proved to be the most reliable. For a 100 second sequence, it achieved the lowest latency of 408.18±17.74 ms, compared to other methods.

Discussion

Artificial intelligence (AI) has become a buzzword in recent years, with many systems claiming superiority over others. However, the terms encompass a wide range of machine learning (ML) algorithms with varying strengths and weaknesses. Although ML methods can substantially improve the efficiency of clinical workflows during atrial mapping and catheter ablation, it is important to evaluate these systems in a standardized complex environment to determine which one is best suited for a specific user case.

Our research compared five methods for segmenting intracardiac atrial electrograms (EGM) and measuring LAT in multichannel signals recorded from within the coronary sinus.

Conventional methods[23,57,58] are typically based on temporal or spectral characteristics, focusing mainly on finding well-defined local extrema that characterize the position of the entire region of interest, such as the highest peak, barycenter or minimum of the first derivative ($-\frac{dV}{dt}$). The accuracy of such methods has been shown to be sufficient for common cases, but it usually needs manual revision under more challenging conditions. Thus, less precise measurement can be expected in various mapping frameworks, such as post-pacing interval analysis or mapping of early activation in accessory pathways and ventricular tachycardias. For example, the authors[23] achieved an overall measurement error of 22±24 ms.

In this context, we have shown that ML and AI methods, and especially deep neural networks, can provide better performance in diverse procedural settings, even when dealing with low signal quality. To our knowledge,

this is the first study to provide such a thorough evaluation of methods for the detection of early and late LATs in intracardiac signals, extensively refined by using Bayesian optimization.

The impact of model complexity on performance

Each method, rule-based, SVM, and convolutional neural networks (U-Net, Faster R-CNN, DenseNet), presents distinct advantages and disadvantages for measuring LATs. The rule-based method sets a threshold for each channel and measures when the signal crosses it, then aggregates the outputs. Although easy to implement, it fails to capture complex multidimensional dependencies between channels, leading to poorer detection and less precise measurements, as shown in our results. This is evident in Figure 5, where far-field ventricular activity in two of the five channels resulted in a false detection.

SVM separates data into classes, either an atrial complex or background noise, by finding a hyperplane in the feature space. Convolutional neural networks learn hierarchical filters that extract features directly from the electrogram. These methods capture complex multidimensional dependencies across channels, improving accuracy. All four methods utilize the five available channels, but interpret the interlead and temporal context differently. SVM relies on manual feature extraction via the wavelet transform, producing EGM descriptors independent of each channel. Its temporal receptive field is limited by the width of the mother wavelet, complicating the analysis of the propagation of the activation wavefront in larger tissue areas measured by multiple catheters.

Conversely, U-Net, Faster R-CNN, and DenseNet undertake automatic feature extraction while concurrently analyzing interlead correlations. It offers an adjustable receptive field that allows analysis on a beat-to-beat scale and even wider[59]. However, this always comes at the expense of an increased computational load.

Multi-channel recordings are common across medicine where detailed data acquisition is required. In electrophysiology mapping, especially with high-density catheters, these recordings are crucial. ML and AI methods can take advantage of these data to improve the measurement performance and beat-to-beat reproducibility of early and late LATs, as shown in the results. This is particularly important for mapping the exit sites of arrhythmias and the origins of the triggered activity. However, models that synthesize features from all channels may obscure individual LATs, complicating the measurement of spatial wave propagation during a single beat. Addressing this requires either augmenting the dataset with LAT annotations for each channel or refining models with advanced optimization techniques. As a result, rule-based methods may still be relevant in scenarios where general knowledge, rather than dataset-specific properties, is crucial.

Dense sample-wise classification models, commonly employed for semantic segmentation, may struggle with generating coherent segmentations, particularly when dealing with noisy or poorly defined object boundaries[60–62]. To address this in the semantic models (U-Net, DENSE-ECG, and SVM+WT) explored in this work, a post-processing step involving segment merging and the removal of short segments was implemented. Specifically, segments shorter than 8ms (SVM+WT) and 20ms (U-Net, DENSE-ECG) were discarded, while segments closer than 0ms (SVM+WT) and 48ms (U-Net, DENSE-ECG) were merged. These thresholds were optimized using BO, maximizing the Dice score within a defined search space for each parameter (details in the Supplementary Material). Although conceptually simple, this post-processing step further improved the performance of the semantic models. The Dice score improved by 1% for SVM+WT, 0.6% for U-Net, and 0.4% for DENSE-ECG. It was not applied to the instance segmentation models (Faster R-CNN and rule-based method) as their inherent mechanisms, such as Non-Maximum Suppression in Faster R-CNN and the filtering rules in the rule-based approach, effectively handle fragmented detections. Although these improvements were observed in the overall Dice scores, analysis of the RMSE of the detected boundaries revealed mixed results. Changes in RMSE ranged from 0.05 to 0.6 ms across different models and boundary types (onset vs. offset), suggesting that post-processing may not uniformly improve accuracy for all instances, but rather target specific problematic instances. For models like Support Vector Machines, where architectural modifications are less straightforward, post-processing remains a crucial tool. However, for deep learning models, exploring regularization techniques, including penalties for segment fragmentation, may be a potentially more integrated and effective approach.

Unlike the semantic models, Faster R-CNN, as an object detection approach, naturally avoids segment fragmentation by identifying regions of interest as single, cohesive objects. However, this leads to a distinct error distribution for the detection of onset and offset (see Fig. 6). Faster R-CNN exhibited a broader and arguably less desirable error distribution compared to the semantic models. This likely stems from the multi-task loss function, which may prioritize electrogram localization over precise boundary delineation. This suggests an advantage for semantic models in applications demanding high-precision boundary detection, motivating further investigation into loss functions that better balance detection accuracy with boundary precision in Faster R-CNN.

As demonstrated by studies on noise removal in both surface[63] and intracardiac[64] recordings, acquired signals commonly have a signal-to-noise ratio (SNR) of around 10 dB, which requires adept processing. Interestingly, the optimal frequency bands for the SVM model did not optimally suppress powerline interference (PLI_{50}), likely because the slow rolloff rate insufficiently attenuated the noise. In contrast, the rule-based method benefited from using a single frequency band at higher frequencies to effectively cancel PLI_{50}, but this made the system more sensitive to Gaussian white noise (GWN). As evident in Figure 7, the observed performance metrics dropped rapidly with higher levels of GWN. Additional filtering techniques specifically designed to alleviate PLI_{50} and smooth out GWN could benefit both approaches, but at the risk of electrogram distortion and a sacrifice in measurement accuracy. AI driven methods proved to be the most stable in terms of performance, capable of extracting good electrogram representations even at SNR levels below the considered threshold of 10 dB. However, this is at the cost of high model complexity.

Generalising study results

The lack of comprehensive public intracardiac electrogram data sets (EGMs) is a significant challenge in the development and evaluation of algorithms for the detection of electrograms and the measurement of cardiac activation times. Performance is highly dependent on the quantity and clinical variability[65] of the electrograms. Other reported studies[23,57,66] in this field have tested algorithms using proprietary electrogram recordings from only 8 to 31 patients. This makes it difficult to mutually compare results and highlight strengths and weaknesses without extensive evaluation on various public datasets. Although El Haddad[23] reports that the vast majority of LAT onset detection deviations were around 10 ms, our results are comparable or even better, demonstrating the potential of our approach for more accurate detection. Furthermore, some studies[67–69] have focused on developing ML algorithms only for the classification of specific types of rhythm, such as atrial fibrillation. Although it could help improve the technical aspects of complex mapping procedures, such algorithms are limited in their scope and cannot be used for the automation of other types of analysis. Our study addresses these problems using more than 320 electrograms from 100 patients that cover a wide range of rhythms including atrial fibrillation, pacing sequences, and different electrogram morphologies. This allows the development of a more generalized and likely preferable solution for the measurement and analysis of intracardiac electrograms, as demonstrated in atrial fibrillation recordings by the U-Net model in our previous work[48].

Limitations

It is essential to acknowledge several limitations that can impact the relevance of our findings to a broader context. Despite the data set comprising various cardiac rhythm disorders and a substantial number of subjects compared to similar studies, it mainly includes pediatric patients referred for electrophysiology procedures. This may introduce bias and may not accurately represent the general population undergoing catheter ablation.

Algorithms were tested only using coronary sinus electrograms, chosen for its challenging morphology that contains atrial and ventricular complexes. Different outcomes may be observed in other anatomical locations or in electrically modified tissue.

It should be noted that there are various methods available for enhancing and extracting electrogram features, particularly when combined with threshold-based measurement. Conducting a comprehensive comparison of all available preprocessing techniques went beyond the scope of this study. It is reasonable to anticipate variations in performance across different preprocessing algorithms. This study primarily focused on comparing the performance of established systems using hand-crafted rules with that of more complex machine learning models. The results based on our thresholding method shouldn't be generalised for all knowledge-based systems.

When employing complex ML algorithms, particularly in mission-critical applications such as real-time analysis during interventional procedures, it is essential to take into account computational processing latency. Numerous solutions are available to reduce the end-to-end response time of the system, such as the use of dedicated acceleration chip designs, process parallelism, and model quantization/compression. Due to implementation differences, our study did not thoroughly compare computational performance, and results can vary depending on the software and libraries used.

Conclusion

This paper presents a comparative analysis of five distinct methods for segmenting intracardiac recordings acquired through a coronary sinus catheter. Our objective is to provide a comprehensive assessment of each method's merits and limitations, representative of key trends in biological signal processing: digital signal processing, machine learning, and deep learning.

In particular, AI-driven models show remarkable performance compared to alternative methods. It excelled in both the Dice score evaluation and the instance-wise assessment, demonstrating its robustness and precision in delineating the onset and offset of atrial activity.

Deep learning methods show great potential to enhance segmentation and analysis in a variety of medical applications. Their potential for tasks such as signal processing, electroanatomical mapping, and arrhythmia classification could significantly advance fields such as electrophysiology or other domains in which biological time series are used.

Data availability

Data acquisition was carried out with the approval of the ethics committee of the University Hospital Brno, reference number 05-090920/EK. The dataset used and analyzed during the current study available from the corresponding author on reasonable request.

Received: 18 November 2024; Accepted: 13 February 2025
Published online: 06 April 2025

References

1. Ivaturi, P. et al. A comprehensive explanation framework for biomedical time series classification. *IEEE Journal of Biomedical and Health Informatics* **25**, 2398–2408. https://doi.org/10.1109/JBHI.2021.3060997 (2021).
2. Alcaine, A. et al. A wavelet-based electrogram onset delineator for automatic ventricular activation mapping. *IEEE Transactions on Biomedical Engineering* **61**, 2830–2839. https://doi.org/10.1109/TBME.2014.2330847 (2014).
3. Martinez, J., Almeida, R., Olmos, S., Rocha, A. & Laguna, P. A wavelet-based ECG delineator: evaluation on standard databases. *IEEE Transactions on Biomedical Engineering* **51**, 570–581. https://doi.org/10.1109/TBME.2003.821031 (2004).
4. Peimankar, A. & Puthusserypady, S. DENS-ECG: A deep learning approach for ECG signal delineation. *Expert Systems with Applications* **165**, 113911. https://doi.org/10.1016/j.eswa.2020.113911 (2021).

5. Issa, Z. F., Miller, J. M. & Zipes, D. P. *Clinical arrhythmology and electrophysiology: a companion to Braunwald's heart disease* (Elsevier, Philadelphia, PA, 2024), fourth edition edn. OCLC: 1381742307.
6. Kaiser, D. W. et al. The precise timing of tachycardia entrainment is determined by the postpacing interval, the tachycardia cycle length, and the pacing rate: Theoretical insights and practical applications. *Heart Rhythm* **13**, 695–703. https://doi.org/10.1016/j.hrthm.2015.11.032 (2016).
7. Shenasa, M., Hindricks, G., Callans, D. J., Miller, J. M. & Josephson, M. E. *Cardiac Mapping* (Wiley, 2019), 1 edn.
8. Pascale, P. et al. Pattern and timing of the coronary sinus activation to guide rapid diagnosis of atrial tachycardia after atrial fibrillation ablation. *Circulation: Arrhythmia and Electrophysiology* **6**, 481–490. https://doi.org/10.1161/CIRCEP.113.000182 (2013).
9. Steven, D., Seiler, J., Roberts-Thomson, K. C., Inada, K. & Stevenson, W. G. Mapping of atrial tachycardias after catheter ablation for atrial fibrillation: Use of bi-atrial activation patterns to facilitate recognition of origin. *Heart Rhythm* **7**, 664–672. https://doi.org/10.1016/j.hrthm.2010.01.009 (2010).
10. Haïssaguerre, M. et al. Localized Sources Maintaining Atrial Fibrillation Organized by Prior Ablation. *Circulation* **113**, 616–625. https://doi.org/10.1161/CIRCULATIONAHA.105.546648 (2006).
11. Hongo, R. Catheter Ablation of Scar-mediated Ventricular Tachycardia: Are Substrate-based Approaches Replacing Mapping?. *Journal of Innovations in Cardiac Rhythm Management* **10**, 3707–3715. https://doi.org/10.19102/icrm.2019.100603 (2019).
12. Fustes, O. J. H. et al. Somatosensory evoked potentials in clinical practice: a review. *Arquivos de Neuro-Psiquiatria* **79**, 824–831. https://doi.org/10.1590/0004-282x-anp-2020-0427 (2021).
13. Lantz, G. et al. Space-oriented segmentation and 3-dimensional source reconstruction of ictal EEG patterns. *Clinical Neurophysiology* **112**, 688–697. https://doi.org/10.1016/S1388-2457(01)00479-5 (2001).
14. Costa-Garcia, A., Itkonen, M., Yamasaki, H., Shibata-Alnajjar, F. & Shimoda, S. A Novel Approach to the Segmentation of sEMG Data Based on the Activation and Deactivation of Muscle Synergies During Movement. *IEEE Robotics and Automation Letters* **3**, 1972–1977. https://doi.org/10.1109/LRA.2018.2811506 (2018).
15. McKinnon, M. L., Hill, N. J., Carp, J. S., Dellenbach, B. & Thompson, A. K. Methods for automated delineation and assessment of EMG responses evoked by peripheral nerve stimulation in diagnostic and closed-loop therapeutic applications. *Journal of Neural Engineering* **20**, 046012. https://doi.org/10.1088/1741-2552/ace6fb (2023).
16. Clayson, P. E., Baldwin, S. A. & Larson, M. J. ow does noise affect amplitude and latency measurement of event-related potentials (ERPs)? A methodological critique and simulation study. *Psychophysiology* **50**, 174–186. https://doi.org/10.1111/psyp.12001 (2013).
17. Paradeshi, K., Scholar, R. & Kolekar, P. D. U. Removal of EMG Artifacts from Multichannel EEG Signal Using Automatic Dynamic Segmentation and Adaptive Thresholding with Multilevel Decomposed Wavelets. *IOSR Journal of Electrical and Electronics Engineering* **12**, 30–35. https://doi.org/10.9790/1676-1203043035 (2017).
18. Cazzaniga, G. et al. Improving the Annotation Process in Computational Pathology: A Pilot Study with Manual and Semi-automated Approaches on Consumer and Medical Grade Devices. *Journal of Imaging Informatics in Medicine*[SPACE]https://doi.org/10.1007/s10278-024-01248-x (2024).
19. Tizhoosh, H. R. et al. Searching Images for Consensus. *The American Journal of Pathology* **191**, 1702–1708. https://doi.org/10.1016/j.ajpath.2021.01.015 (2021).
20. Khalifa, M. & Albadawy, M. AI in diagnostic imaging: Revolutionising accuracy and efficiency. *Computer Methods and Programs in Biomedicine Update* **5**, 100146. https://doi.org/10.1016/j.cmpbup.2024.100146 (2024).
21. Faust, O., Hagiwara, Y., Hong, T. J., Lih, O. S. & Acharya, U. R. Deep learning for healthcare applications based on physiological signals: A review. *Computer Methods and Programs in Biomedicine* **161**, 1–13. https://doi.org/10.1016/j.cmpb.2018.04.005 (2018).
22. Estes, N. M. *EP News: Clinical. Heart Rhythm* **17**, 1054. https://doi.org/10.1016/j.hrthm.2020.04.002 (2020).
23. El Haddad, M. et al. Algorithmic detection of the beginning and end of bipolar electrograms: Implications for novel methods to assess local activation time during atrial tachycardia. *Biomedical Signal Processing and Control* **8**, 981–991. https://doi.org/10.1016/j.bspc.2012.11.005 (2013).
24. Nothstein, M. et al. CVAR-Seg: An Automated Signal Segmentation Pipeline for Conduction Velocity and Amplitude Restitution. *Frontiers in Physiology* **12**, 673047. https://doi.org/10.3389/fphys.2021.673047 (2021).
25. Sánchez, J. et al. Using Machine Learning to Characterize Atrial Fibrotic Substrate From Intracardiac Signals With a Hybrid in silico and in vivo Dataset. *Frontiers in Physiology* **12**, 699291. https://doi.org/10.3389/fphys.2021.699291 (2021).
26. Ředina, R. et al. Arrhythmia Database with Annotated Intracardial Atrial Signals from Pediatric Patients Undergoing Catheter Ablation. https://doi.org/10.22489/CinC.2022.282 (2022).
27. Hejc, J., Redina, R., Kolarova, J. & Starek, Z. Epycon: A Single-Platform Python Package for Parsing and Converting Raw Electrophysiology Data into Open Formats. https://doi.org/10.22489/CinC.2023.315 (2023).
28. Wang, Z.-Y., Xiang, Z.-R., Zhi, J.-Y., Ding, T.-C. & Zou, R. A novel physiological signal denoising method coupled with multispectral adaptive wavelet denoising(MAWD) and unsupervised source counting algorithm(USCA). *Journal of Engineering Research* **12**, 175–189. https://doi.org/10.1016/j.jer.2023.07.016 (2024).
29. Hejč, J., Vítek, M., Ronzhina, M., Nováková, M. & Kolářová, J. A Wavelet-Based ECG Delineation Method: Adaptation to an Experimental Electrograms with Manifested Global Ischemia. *Cardiovascular Engineering and Technology* **6**, 364–375. https://doi.org/10.1007/s13239-015-0224-z (2015).
30. Daubechies, I. The wavelet transform, time-frequency localization and signal analysis. *IEEE Transactions on Information Theory* **36**, 961–1005. https://doi.org/10.1109/18.57199 (1990).
31. Lee, G., Gommers, R., Waselewski, F., Wohlfahrt, K. & O'Leary, A. PyWavelets: A Python package for wavelet analysis. *Journal of Open Source Software* **4**, 1237. https://doi.org/10.21105/joss.01237 (2019).
32. Ouali, M. A., Ghanai, M. & Chafaa, K. Upper envelope detection of ECG signals for baseline wander correction: a pilot study. *TURKISH JOURNAL OF ELECTRICAL ENGINEERING & COMPUTER SCIENCES* **26**, 803–816. https://doi.org/10.3906/elk-1705-165 (2018).
33. Benitez, D., Gaydecki, P., Zaidi, A. & Fitzpatrick, A. The use of the Hilbert transform in ECG signal analysis. *Computers in Biology and Medicine* **31**, 399–406. https://doi.org/10.1016/S0010-4825(01)00009-9 (2001).
34. Burges, C. J. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery* **2**, 121–167. https://doi.org/10.1023/A:1009715923555 (1998).
35. Nasiri, J. A., Naghibzadeh, M., Yazdi, H. S. & Naghibzadeh, B. ECG Arrhythmia Classification with Support Vector Machines and Genetic Algorithm. In *2009 Third UKSim European Symposium on Computer Modeling and Simulation*, 187–192, https://doi.org/10.1109/EMS.2009.39 (IEEE, Athens, Greece, 2009).
36. Rabee, A. & Barhumi, I. ECG signal classification using support vector machine based on wavelet multiresolution analysis. In *2012 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA)*, 1319–1323, https://doi.org/10.1109/ISSPA.2012.6310497 (IEEE, Montreal, QC, Canada, 2012).
37. Turnip, A., Ilham Rizqywan, M., Kusumandari, D. E., Turnip, M. & Sihombing, P. Classification of ECG signal with Support Vector Machine Method for Arrhythmia Detection. *Journal of Physics: Conference Series* **970**, 012012. https://doi.org/10.1088/1742-6596/970/1/012012 (2018).
38. Smisek, R. et al. Multi-stage SVM approach for cardiac arrhythmias detection in short single-lead ECG recorded by a wearable device. *Physiological Measurement* **39**, 094003. https://doi.org/10.1088/1361-6579/aad9e7 (2018).
39. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444. https://doi.org/10.1038/nature14539 (2015).

40. Farabet, C., Couprie, C., Najman, L. & LeCun, Y. Learning Hierarchical Features for Scene Labeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35**, 1915–1929. https://doi.org/10.1109/TPAMI.2012.231 (2013).
41. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. *Communications of the ACM* **60**, 84–90. https://doi.org/10.1145/3065386 (2017).
42. Hinton, G. et al. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. *IEEE Signal Processing Magazine* **29**, 82–97. https://doi.org/10.1109/MSP.2012.2205597 (2012).
43. Leung, M. K. K., Xiong, H. Y., Lee, L. J. & Frey, B. J. Deep learning of the tissue-regulated splicing code. *Bioinformatics* **30**, i121–i129. https://doi.org/10.1093/bioinformatics/btu277 (2014).
44. Ma, J., Sheridan, R. P., Liaw, A., Dahl, G. E. & Svetnik, V. Deep Neural Nets as a Method for Quantitative Structure-Activity Relationships. *Journal of Chemical Information and Modeling* **55**, 263–274. https://doi.org/10.1021/ci500747n (2015).
45. Somani, S. et al. Deep learning and the electrocardiogram: review of the current state-of-the-art. *EP Europace* **23**, 1179–1191. https://doi.org/10.1093/europace/euaa377 (2021).
46. Stracina, T., Ronzhina, M., Redina, R. & Novakova, M. Golden Standard or Obsolete Method? Review of ECG Applications in Clinical and Experimental Context. *Frontiers in Physiology* **13**, 867033. https://doi.org/10.3389/fphys.2022.867033 (2022).
47. Hejc, J., Redina, R., Kolarova, J. & Starek, Z. Multi-channel delineation of intracardiac electrograms for arrhythmia substrate analysis using implicitly regularized convolutional neural network with wide receptive field. *Biomedical Signal Processing and Control* **94**, 106274. https://doi.org/10.1016/j.bspc.2024.106274 (2024).
48. Ren, S., He, K., Girshick, R. & Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks (2016). ArXiv:1506.01497 [cs].
49. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation, https://doi.org/10.48550/ARXIV.1505.04597 (2015). Version Number: 1.
50. Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K. & Yuille, A. L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs, https://doi.org/10.48550/ARXIV.1606.00915 (2016). Version Number: 2.
51. Abraham, N. & Khan, N. M. A Novel Focal Tversky loss function with improved Attention U-Net for lesion segmentation, https://doi.org/10.48550/ARXIV.1810.07842 (2018). Version Number: 1.
52. Han, S., Eom, H., Kim, J. & Park, C. Optimal DNN architecture search using Bayesian Optimization Hyperband for arrhythmia detection. In *2020 IEEE Wireless Power Transfer Conference (WPTC)*, 357–360, https://doi.org/10.1109/WPTC48563.2020.9295590 (IEEE, Seoul, Korea (South), 2020).
53. Akiba, T., Sano, S., Yanase, T., Ohta, T. & Koyama, M. Optuna: A Next-generation Hyperparameter Optimization Framework, https://doi.org/10.48550/ARXIV.1907.10902 (2019). Version Number: 1.
54. Kwak, S. G. & Kim, J. H. Central limit theorem: the cornerstone of modern statistics. *Korean Journal of Anesthesiology* **70**, 144. https://doi.org/10.4097/kjae.2017.70.2.144 (2017).
55. Callaert, H. & Janssen, P. The Berry-Esseen Theorem for U-Statistics. *The Annals of Statistics* **6**, 417–421 (1978).
56. Zhou, Y., Zhu, Y. & Wong, W. K. Statistical tests for homogeneity of variance for clinical trials and recommendations. *Contemporary Clinical Trials Communications* **33**, 101119. https://doi.org/10.1016/j.conctc.2023.101119 (2023).
57. Dubé, B. et al. Automatic detection and classification of human epicardial atrial unipolar electrograms. *Physiological Measurement* **30**, 1303–1325. https://doi.org/10.1088/0967-3334/30/12/002 (2009).
58. Trigano, T. & Luengo, D. Intracardiac ECG pulse localization using overlapping block sparse reconstruction. *Biomedical Signal Processing and Control* **79**, 103921. https://doi.org/10.1016/j.bspc.2022.103921 (2023).
59. Gong, S. et al. Dilated FCN: Listening Longer to Hear Better, https://doi.org/10.48550/ARXIV.1907.11956 (2019). Version Number: 1.
60. Kim, K., Chalidabhongse, T. H., Harwood, D. & Davis, L. Real-time foreground-background segmentation using codebook model. *Real-Time Imaging* **11**, 172–185. https://doi.org/10.1016/j.rti.2004.12.004 (2005).
61. Parks, D. H. & Fels, S. S. Evaluation of Background Subtraction Algorithms with Post-Processing. In *2008 IEEE Fifth International Conference on Advanced Video and Signal Based Surveillance*, 192–199, https://doi.org/10.1109/AVSS.2008.19 (IEEE, Santa Fe, NM, USA, 2008).
62. Waheed, K., Weaver, K. & Salam, F. A robust algorithm for detecting speech segments using an entropic contrast. In *The 2002 45th Midwest Symposium on Circuits and Systems, 2002. MWSCAS-2002.*, vol. 3, III–328–III–331, https://doi.org/10.1109/MWSCAS.2002.1187039 (IEEE, Tulsa, OK, USA, 2002).
63. Blanco-Velasco, M., Weng, B. & Barner, K. E. ECG signal denoising and baseline wander correction based on the empirical mode decomposition. *Computers in Biology and Medicine* **38**, 1–13. https://doi.org/10.1016/j.compbiomed.2007.06.003 (2008).
64. Martínez, M., Ródenas, J., Alcaraz, R. & Rieta, J. J. Study on the Alternatives to Reduce High-Frequency Noise from Invasive Recordings of Atrial Fibrillation. https://doi.org/10.22489/CinC.2017.111-048 (2017).
65. Alwosheel, A., Van Cranenburgh, S. & Chorus, C. G. Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis. *Journal of Choice Modelling* **28**, 167–182. https://doi.org/10.1016/j.jocm.2018.07.002 (2018).
66. Veshchezerova, D., Bars, C. & Seitz", J. Cycle Length Estimation Using Accurate Adaptive Detection of Local Activations in Atrial Intracardiac Electrograms. https://doi.org/10.22489/CinC.2022.142 (2022).
67. Liao, S. et al. Deep learning classification of unipolar electrograms in human atrial fibrillation: application in focal source mapping. *Frontiers in Physiology* **12**, 704122. https://doi.org/10.3389/fphys.2021.704122 (2021).
68. Rodrigo, M., Rogers, A. J., Ganesan, P., Alhusseini, M. & Narayan, S. M. Abstract 14742: Deep Learning of Intracardiac Electrograms in Atrial Arrhythmia. *Circulation* **142**, https://doi.org/10.1161/circ.142.suppl_3.14742 (2020).
69. Rodrigo, M. et al. Machine learning classifies intracardiac electrograms of atrial fibrillation from other arrhythmias. *Journal of the American College of Cardiology* **77**, 279. https://doi.org/10.1016/S0735-1097(21)01638-7 (2021).

Acknowledgements

Brno Ph.D. Talent Scholarship Holder funded by the Brno City Municipality. This publication was written at Masaryk University as part of the project "Novel imaging, computing and analytical methods in cardiovascular diseases diagnostics and monitoring" number MUNI/A/1410/2022 with the support of the Specific University Research Grant, as provided by the Ministry of Education, Youth and Sports of the Czech Republic in the year 2023.

Author contributions

R.R. and J.H. conceived and conducted the experiments. R.R., J.H., and Z.S. prepared data. R.R., J.H., and M.F. analyzed the results. All authors reviewed the manuscript.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-90533-y.

Correspondence and requests for materials should be addressed to R.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.