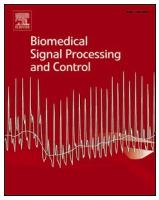
Biomedical Signal Processing and Control 94 (2024) 106274

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc



Multi-channel delineation of intracardiac electrograms for arrhythmia substrate analysis using implicitly regularized convolutional neural network with wide receptive field

Jakub Hejc [a,b,c,*], Richard Redina [a,b], Jana Kolarova [b], Zdenek Starek [a,d]

[a] International Clinical Research Center, St. Anne's University Hospital Brno, Pekařská 53, 656 91 Brno, Czech Republic
[b] Brno University of Technology, Department of Biomedical Engineering, Brno 601 90, Czech Republic
[c] University Hospital Brno, Children's Hospital, Department of Pediatrics, Brno 625 00, Czech Republic
[d] First Department of Internal Medicine/Cardioangiology, St. Anne's Hospital, Masaryk University, Pekařská 53, 656 91 Brno, Czech Republic

ARTICLE INFO

Keywords:
Cardiac electrophysiology
Local activation time
Arrhythmia mapping
Convolutional neural network
Wide receptive field
Dilated convolution

ABSTRACT

Objective: Automated segmentation of intracardiac electrograms and extraction of fundamental cycle length intervals is crucial for reproducible arrhythmia substrate analysis conducted during electrophysiology procedures. The objective of this study is to develop a robust, computationally efficient end-to-end model for the precise electrogram multi-channel delineation using a highly imbalanced dataset.

Methods: A temporal deep convolutional neural network (CNN) based on the UNet architecture incorporating convolutional layers of varying dilation rates was implicitly regularized through data augmentations (DAs), a domain specific Tversky loss function, and distinct labelling strategies for segments comprising atrial fibrillation (AF). An exploratory study utilizing Bayesian search was conducted to optimize architectural and loss function hyperparameters. The impact of dilated convolutions, data augmentations, and labelling strategies on the performance and generalization capability was assessed through an ablation study. The performance of different models was evaluated using a cross-validation procedure and two independent test datasets derived from two separate patient cohorts containing 326, 84, and 97 electrograms encompassing sinus rhythms, abnormal complexes during ongoing tachycardias, and stimulation protocols.

Results: A UNet model with optimized loss hyperparameters, a dilated receptive field, and atrial fibrillation (AF) annotated as a positive class (D-UNet-L) achieved an average Sørensen-Dice coefficient (SDC) of 84.9 % on recordings with regular atrial beats across test datasets, surpassing the performance of models without loss optimization (81.5 %), without dilated kernels (81.3 %), and with inversed AF labelling (77.5 %). Notably, the highest average accuracy (Acc) of 95.8 % for AF recordings was obtained by a model trained on negatively assigned AF segments, outperforming D-UNet-L (88.9 %), the model without loss optimization (81.5 %), and the model without dilated kernels (81.3 %). The reference D-UNet-L model exhibited overall root-mean-square errors of 8.3 and 9.0 ms across test datasets. Additionally, 61.5 % and 20.8 % of delineations exhibited absolute errors below 5 ms and 10 ms, respectively. Disabling data augmentation (DAs) resulted in a 2.7 % decrease in validation SDC and a 5.3 % increase in training SDC.“

Conclusion: Generalization capability across independent datasets was improved by employing exponentially weighted Tversky loss. The model's segmentation performance on longer sequences with atrial fibrillation was improved by incorporating dilated convolution kernels. Noise-aware and morphology data augmentations effectively mitigated overfitting potential in a limited training dataset. Label noise introduced by annotating atrial fibrillation sequences into a positive class strengthened regularization of the model, particularly in its ability to identify regular beats. However, it also negatively impacted performance on F-waves.

Significance: The proposed method diminishes the dependence on manual measurements in complex electrophysiology measurements and enhances utilization of the collected electrophysiology data. The findings underscore the substantial impact of intrinsic regularization techniques and establish guidelines to facilitate methodological decision-making processes.

* Corresponding author at: International Clinical Research Center, Interventional Cardiac Electrophysiology Group, St. Anne's University Hospital - International Clinical Research Center, Pekařská 53, 656 91 Brno, Czech Republic.
E-mail address: jakub.hejc@fnusa.cz (J. Hejc).

https://doi.org/10.1016/j.bspc.2024.106274
Received 30 June 2023; Received in revised form 25 December 2023; Accepted 20 March 2024
Available online 8 April 2024

1. Introduction

Catheter ablation has emerged as an effective and safe treatment for cardiac arrhythmias, resulting in sustained clinical improvements and enhanced quality of life for most patients [1]. Electrophysiology (EP) study is essential in investigating myocardial electrical properties and understanding arrhythmia mechanisms [2,3].

EP studies typically involve recording intracardiac electrograms (EGMs) from one or more locations on the endocardial surface in response to multiple stimulation protocols or during ongoing arrhythmias. Local activation times (LATs) of EGM complexes and inter-beat cycle lengths (CLs) are commonly measured to derive clinical parameters that aid in localizing arrhythmia sources [4–8]. For instance, tissue restitution curves [9] and effective refractory period [10] can be determined by employing dynamic S1-S2 stimulation protocols and measuring post-pacing CLs. Analysis of pacing and post-pacing CLs is commonly employed to distinguish between potential arrhythmia origins in atrial flutter (AFL) entrainment mapping [11], atrio-ventricular (AV) tachycardias [12], or in localizing conduction gaps during pulmonary vein isolation [13].

The localization of EGM complexes and the detection of both their onset and offset are crucial steps in extracting LATs and CLs, identifying earliest (EA) or latest activation (LA) in multiple electrode locations, and assisting in diagnosing the arrhythmia substrate. Despite advancements in automatic EGM segmentation, visual and manual assessment of LATs, EAs, LAs, and CLs continue to be an essential aspect of EP studies. The primary drawback of manual measurement is that it is time-consuming and highly subjective, particularly when EGMs contain fusion of atrial and ventricular complexes or superimposed artifacts from programmed stimulation [14]. In such cases, conventional filtering and feature extraction techniques may excessively suppress the component of interest, leading to detrimental effects on detection performance and potential loss of valuable data in complex mapping scenarios.

In this study, a modified deep convolutional neural network (CNN) architecture U-Net with an enhanced receptive field was proposed for automated delineation of atrial EGMs. The main objective of the method is to accurately detect onsets and offsets of atrial complexes in multi-channel EGMs to extract basic CLs within various diagnostic pipelines employed during EP studies, such as S1-S2 programmed stimulation, entrainment mapping, or early activation measurement. The method can automate measurements in several challenging scenarios, including superimposed atrioventricular (AV) complexes and highly irregular rhythms.

The model's receptive field was enhanced by incorporating dilated filter kernels to provide a longer temporal context. An intrinsic regularization method such as data augmentation, domain-specific loss functions and annotation strategies were thoroughly explored to identify robust end-to-end solution without computationally expensive preprocessing steps, and the factors that most benefit the detection and delineation accuracy of a deep neural model trained on limited imbalanced dataset. To the best of our knowledge, there is currently no generalized approach for atrial complex segmentation that employs multi-lead and inter-beat context.

2. Problem formulation

In this work, the multi-channel EGM delineation problem was reformulated as a binary semantic segmentation task. The objective is to find a mapping function g: $X \mapsto Y$, where g is a feature extractor, $X \in \mathbb{R}^d$ are input EGM sequences containing d channels, and Y ground truths of atrial complexes boundaries. Function g(X) models the probability of sample $y_i \in Y$ of being part of an atrial complex with respect to information from all d channels.

3. Related works

3.1. Beat localization and inter-beat interval analysis

While numerous ML methods have been widely adopted for analyzing arrhythmia-related morphological changes and classification of intracardiac EGMs [15,16], even the most recent methods for the delineation predominantly rely on fundamental feature extraction techniques and expertise-driven systems derived from cellular measurements or simulated data. A comprehensive survey of conventional algorithms can be found in Cantwell et al. [17].

Feature extraction generally encompasses strategies such as signal correlation with template banks [18], nonlinear operators to estimate instantaneous signal energy [14,19,20,21], wavelet transform [22], derivative-based methods [23], or Hilbert transform [22,24]. Subsequent delineation of the complex boundaries is performed in the transformed signal using adaptive voltage thresholds [14,22] or sets of morphological rules [23].

Nothstein et al. [14] proposed a method based on matched filtering and a nonlinear energy operator for segmenting isolated atrial EGMs and stimulation artifacts during S1-S2 protocols, enabling the measurement of pacing CLs and subsequent electrophysiological restitution curve analysis. El Haddad et al. [23] employed a 4-state machine algorithm comprising a set of detection rules and thresholds to determine the beginning and the end of atrial EGMs in the differentiated and rectified signal. Wavelet transform has been explored for delineating ventricular complexes in [22], wherein authors combined Hilbert transform and multiscale decomposition to identify maxima and minima [25] across different scales of wavelet coefficients, facilitating estimations of complex boundaries. These algorithms have been adapted for inter-beat CL measurement during atrial fibrillation [20,26,27].

To our knowledge, the only ML method for EGM delineation proposed to date is that of Hajimolahoseini et al. [28]. The algorithm employs a Gaussian mixture model to classify consecutive pairs of EGM inflection points as either a segment with atrial activation or an isoelectric line. The model utilizes 3-dimensional feature space constructed with morphology, energy-based and differential operators.

3.2. Datasets

The development of ML algorithms for automated EGM delineation has been hindered by the interdisciplinary nature of this task and limited access to clinical data. Most studies have employed proprietary datasets comprising a limited number of patients (often lower tens) with preselected signals representing atrial fibrillation [27], atrial tachycardias [27,23], or ventricular electrograms in patients with premature contractions [29]. None of these datasets has been made publicly available.

Currently, only two publicly available datasets contain EGM recordings: the Intracardiac Atrial Fibrillation Database (IAFDB) [30] and the Ann Arbor Electrogram Libraries (AAEL) [31]. However, neither of these datasets includes annotations of EGM complex onset and offset. While Prudat et al. [20] and Nothstein et al. [14] offer clinical recordings with annotations upon request, these datasets consist of short-term atrial fibrillation recordings from 35 patients [20] and a single recording during an S1-S2 pacing protocol [14], respectively. Nothstein et al. [14] further employed a synthesized electrogram with 152 annotated complexes, and Abdi et al. [32] provided simulated EGMs lacking annotations of complex boundaries.

The paucity of sufficiently diverse datasets poses challenges for training reliable ML models and for conducting fair comparisons of detection and delineation performance across different algorithms.

4. Materials and methods

4.1. Clinical datasets

The data employed in this study are retrospective and multicenter, encompassing two distinct patient populations who underwent interventional EP procedures. The first cohort comprises children and young adults (Y) examined at the Children's Hospital, University Hospital Brno, Brno, Czech Republic. The second cohort comprises middle-aged patients (M) who attended the EP center affiliated with the Department of Internal Medicine and Cardiology, University Hospital Brno, Brno, Czech Republic. The study was approved by the local ethics review committee.

4.1.1. Data acquisition

EP data were recorded using Abbott WorkMate 4.2 (Y cohort) and Abbott WorkMate 4.3 (M cohort) systems at a sampling rate of 2000 Hz and a resolution of 78 nV/LSb. The acquisition unit included built-in hardware high-pass filter with a cutoff of 0.5 Hz and a notch filter at 50 Hz. 8-lead surface ECGs and raw U-EGMs from a decapolar diagnostic catheter inserted into the coronary sinus (CS) were extracted from acquired data for the purpose of this research.

4.1.2. Data management and clinical characteristics

The data were partitioned into a cross-validation and two independent hold-out sets using patient-oriented selection. This approach ensures that no subset shares EGMs from the same patient.

A 3-fold cross-validation training set (DSCV-Y) encompassing 326 short-term bipolar EGM (B-EGM) segments from 100 consecutive patients in Y group was utilized for model training. The dataset was previously described in [33]. Stratified randomization into three folds was employed to mitigate potential bias in quality metrics arising from the highly imbalanced representation of cardiac rhythms, procedure-related protocols, and EGMs complexes with varying segmentation difficulties. The optimal cross-fold distribution was assessed using relative entropy and Pearson's chi-square. Rhythms characterized by specific temporal sequences were selected as strata and are indicated in Table 1.

The first hold-out test set (DST-Y) comprises 84 EGM sequences from 20 consecutive patients from Y group. DST-Y is independent of DSCV-Y and underwent the EP procedure during a distinct time period. The second test set (DST-M) encompasses 97 EGM sequences from 30 subjects in the M group. Neither DST-Y nor DST-M were involved in training or hyperparameter tuning of the algorithm.

Each dataset comprises procedure-related events that have been validated and confirmed by an electrophysiology expert. Clinical characteristics of individual datasets are presented in Table 1.

Table 1
Clinical characteristics of cross-validation and hold-out datasets DSCV-Y, DST-Y, DST-M. Symbol † indicates strata used for randomization.

Dataset	DSCV-Y	DST-Y	DST-M
Subjects	100	20	30
Age [years]	14.0 (12.0–17.0)	15.0 (12.3–16.3)	58.5 (48.0–63.0)
Sex: males	52 (52.0 %)	10 (50.0 %)	16 (53.3 %)
Length of records [s]	8.5 (6.4—12.2)	7.9 (6.5–8.9)	9.8 (7.7–12.1)
Number of records	326	84	97
Arrhythmia type:			
Sinus rhythm†	191 (58.6 %)	40 (47.6 %)	49 (58.3 %)
Ventricular pre-excitation	58 (17.8 %)	21 (25.0 %)	0 (0.0 %)
Atrial premature beat†	47 (14.4 %)	8 (9.5 %)	13 (15.5 %)
AV node reentry tachycardia†	45 (13.8 %)	3 (3.6 %)	4 (4.8 %)
AV reentry tachycardia†	40 (11.3 %)	12 (14.3 %)	3 (3.6 %)
Ventricular premature beat†	40 (11.3 %)	4 (4.8 %)	16 (19.0 %)
Ventricular pacing	35 (10.7 %)	14 (16.7 %)	9 (10.7 %)
Atrial fibrillation†	26 (10.1 %)	5 (6.0 %)	15 (17.9 %)
Junctional rhythm†	22 (6.7 %)	1 (1.2 %)	0 (0.0 %)
Atrial pacing†	19 (5.8 %)	15 (17.9 %)	21 (25.0 %)
Left bundle branch block	9 (2.8 %)	2 (2.4 %)	2 (2.4 %)
Focal atrial tachycardia	8 (2.5 %)	3 (3.6 %)	0 (0.0 %)
Right branch block	7 (2.2 %)	2 (2.4 %)	0 (0.0 %)
Atrial flutter	5 (1.5 %)	0 (0.0 %)	6 (7.1 %)
1st degree AV block	3 (0.9 %)	0 (0.0 %)	0 (0.0 %)

4.1.3. Data annotation

Each B-EGM segment underwent visual examination and manual annotation of atrial complexes by a single experienced electrophysiology expert. Since the primary focus was on extracting CLs, early, and late activations, the "global" onset of an atrial complex was annotated as the earliest activation detected across all B-EGM channels. The "global" offset of the atrial complex was determined as the transition into an electrical isoline following the latest local activation. Global annotation poses a challenge in cases of atrial fibrillation F-waves, which are typically dispersed across individual leads without clearly defined boundaries. In such instances, the beginning and end of the arrhythmia itself were marked as the boundaries for atrial activity (see Fig. 1).

Atrial complexes were categorized into one of four groups based on its characteristics: a) normal sinus beat (SB); b) abnormal atrial beat (AB); c) atrial fibrillation (AF); and d) stimulated atrial beat (ST). Each category presents a varying level of detection and delineation difficulty. For instance, the ST group includes B-EGMs with pacing artifacts often situated near the atrial complex. The AB group contains arrhythmic complexes, frequently superimposed on ventricular far-field signals. Similarly, each recording is distinguished between the presence of an AF segment (WAF) and its absence (WOAF). Overall, the datasets comprise of 3,624; 3,710 and 701 annotated regular atrial complexes in SB, AB, and ST group.

From a clinical standpoint, such categorization is not essential for intracardiac interval measurement. In this work, the categories are employed exclusively for comprehensive evaluation of the delineation performance under diverse scenarios. The cumulative relative durations and aggregated individual durations of atrial complexes from each group are presented in Fig. 2.

4.2. Data preprocessing

4.2.1. Feature extraction and standardization

Feature extraction is an integral part of many machine learning (ML) methods, aiming to eliminate redundant components and transform data into a more discriminative feature space. Two key concepts exist: manual feature engineering and automatic feature extraction, particularly associated with deep learning (DL) techniques.

Common feature engineering approaches for biological time series involve time–frequency decomposition, typically performed using short-time Fourier transform (STFT) or wavelet transform, as well as statistical methods for sparse reconstruction, such as independent component analysis [34]. While these methods can enhance the performance of DL models, particularly for highly complex data like electroencephalogram [35] and speech signals [36], they introduce computational overhead that may not be desirable for mission-critical EP monitoring applications. In such cases, it may be more advantageous for DL models to leverage the embedding space by incorporating prior knowledge via implicit regularization [37]. Therefore, the preprocessing stage in this work was limited to a few basic steps based on established procedural standards, while denoising and feature extraction were incorporated within the model and augmented using several optimization paradigms discussed in subsequent sections.

U-EGMs are susceptible to remote electromagnetic fields, including activation of distant cardiac cells or powerline interference. To mitigate the impact of these distant sources, 5-lead B-EGMs were constructed by subtracting U-EGMs from consecutive pairs of electrodes. This approach represents the simplest technique for reducing spatially correlated

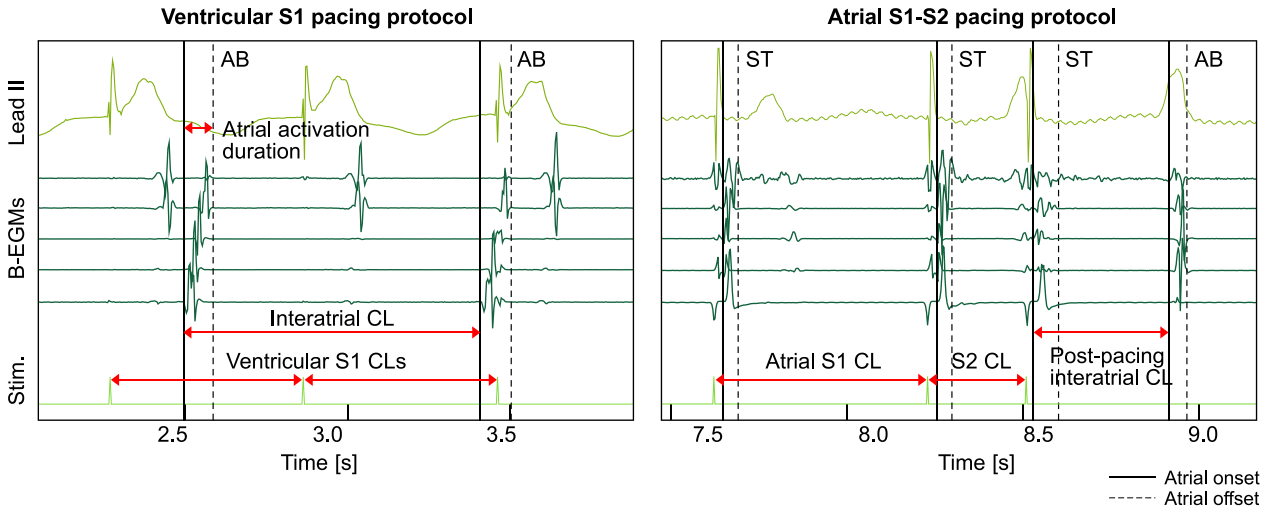


Fig. 1. Representative B-EGM signals from an internal dataset showcasing characteristic pacing protocols during EP studies, atrial beat categories AB and ST, and commonly measured intervals. In the left panel, a prolonged interatrial CL during ventricular pacing suggests dissociation between the ventricles and atria; in the right panel, a shortened post-pacing interatrial CL reveals possible substrate for supraventricular tachycardia.

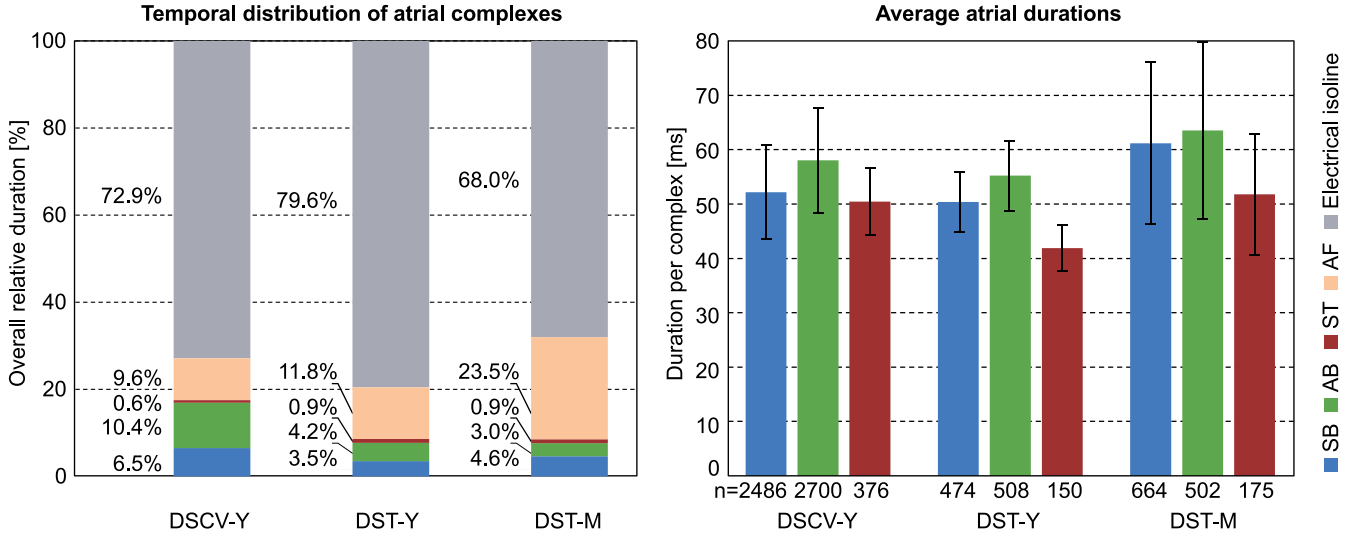


Fig. 2. The left panel displays the imbalanced temporal distribution of atrial B-EGMs across the four categories, namely SB (normal sinus beat), AB (abnormal atrial beat), ST (stimulated atrial beat), and AF (atrial fibrillation) in the DSCV-Y, DST-Y and DST-M datasets. The AF group is the least common B-EGM type in all three datasets. The right panel shows the total count of individual atrial B-EGMs and their durations (expressed as mean and standard deviation) for each category. SB, AB, and ST exhibit very short durations compared to regions of electrical inactivity. AF rhythms are typically much longer than isolated atrial complexes and were therefore excluded from the figure to maintain clarity.

sources in a multi-channel setup and enhancing high-frequency components in the signal [38]. B-EGMs remain a clinical standard for measuring procedurally important CLs, despite the significant influence of wavefront direction on their morphology and interpretation.

To address potential baseline drift and residual low-frequency components, B-EGMs were filtered using a high-pass finite impulse response (FIR) filter with a 5 Hz cutoff frequency, preserving the most relevant frequencies in the range of 5–250 Hz [39].

B-EGMs exceeding ± 12.436 mV were clipped to eliminate extreme voltage oscillations caused by mechanical interference and pacing artifacts. B-EGMs were then rescaled using z-score standardization:

$$B-EGMscaled = (B-EGM-\mu)/\sigma \quad (1)$$

where μ represents the DC component removed by the FIR filter and σ is the standard deviation of the dataset.

4.2.2. *Morphology aware augmentation*

Data augmentation (DA) allows to increase morphological variability of atrial B-EGMs and incorporate the domain-specific knowledge into learned data representation. It is a fundamental strategy used to subject the model to more challenging learning conditions [40,41], particularly in small-sized datasets.

A B-EGM transformation can be expressed as a triple $T = (t, p, m)$, where $t \in \mathscr{T}$ is the transformation, $p \in \mathscr{P}$ is the probability and $m \in \mathscr{M}$ represents the intensity of transformation [42]. Then, DA pipeline is a set of stochastic transformations $\pi = \{T_1, \ldots, T_k\}$, which is applied to each B-EGMs with probability p and magnitude m.

In this work, simple affine transformations of B-EGMs are proposed to modulate underlying bipole directly representing waveform direction, speed, and distance from electrodes. For every B-EGM i-th sample $x_{c,i}$ in c-th channel, the voltage scaling t_{scale}, polarity flipping t_{flip}, temporal translation t_{trans}, and linear temporal scaling $t_{stretch}$ can be expressed as:

$$x' = tscale(x) = mscalexc, n \tag{2}$$

$$x' = tflip(x) = -xc, n \tag{3}$$

$$x' = ttrans(x) = xc, n + mtrans \tag{4}$$

$$x' = ttretch(x) = xc, n^{*}mstretch \tag{5}$$

To emulate a spectrum of B-EGM spatio-temporal activation patterns, the DA consisted of a transformation technique t_{rot} that involved the randomized translation between channels. The translation from channel c to c' can be expressed as:

$$c' = (c + mrot)\%d \tag{6}$$

where symbol % denotes modulo operator and d represents the number of channels.

4.2.3. Noise aware augmentation

A corrupted version of B-EGMs was generated by adding noise N. Subsequently, a robust mapping g: $(X + N) \mapsto Y$ is trained to effectively remove noise from the extracted features. Additive Gaussian white noise, simulated power line interference and pacing artifacts were incorporated as transformations T, which stochastically modulated the targeted noise power P_N and signal-to-noise ratio (SNR) of individual signals SNR_{X+N} by:

$$PN = PX/10^{(}SNRX + N/10) \tag{7}$$

For a more detailed list and description of the parameters used in the DA pipeline, please refer to the Supplementary materials.

4.3. Model architecture

4.3.1. Residual encoder

A temporal UNet [43] architecture incorporating pre-activated residual units with bottleneck (2-factor compression) as proposed by He et al. [44] was employed to model the mapping g: X ↦ Y. The architecture is depicted in Fig. 3 and provided in detail in the Supplementary materials. To effectively process all B-EGMs at once, residual units were modified by replacing original 2D kernels with 1D counterparts, as previously described in our work [45,46].

An anti-aliased strided convolution [47] was employed as a residual pooling operator. Since the data enters the model at the original sampling frequency of 2000 Hz, an aggressive 4-factor downsampling preceded by a 17×1 kernel was employed as the model's input stem to enhance computational efficiency.

4.3.2. Feature pyramid decoder

The latent features z_i from each residual layer L_i were compressed using 1×1 convolutions to ensure a consistent number of channels throughout the network. A pyramid feature network (FPN) [48] was employed to effectively combine low-level and high-resolution zi features into a single feature map that serves as the input to the multilayer perceptron (MLP) classifier. To address the computational overhead associated with conventional UNet deconvolution kernels, the temporal resolutions of two consecutive differently scaled z_i features were aligned using linear interpolation and then combined by element-wise addition.

4.3.3. Wide receptive layer

Dilated convolutions offer a computationally efficient approach for gaining a comprehensive understanding of the surrounding context [49] and modeling long-term variations [50].

This is particularly crucial in handling varying activation patterns observed during EP studies, including polymorphic and often randomly distributed complexes associated with atrial fibrillation. To effectively capture long-term B-EGM variations in such cases, an intermediate receptive layer f_R: $z_5 \mapsto {}'z_5$ with a set of 1×3 dilated convolutions was integrated into the model (D-UNet) between the decoder and the last encoding stage (see Fig. 3). The schematic representation of the f_R is

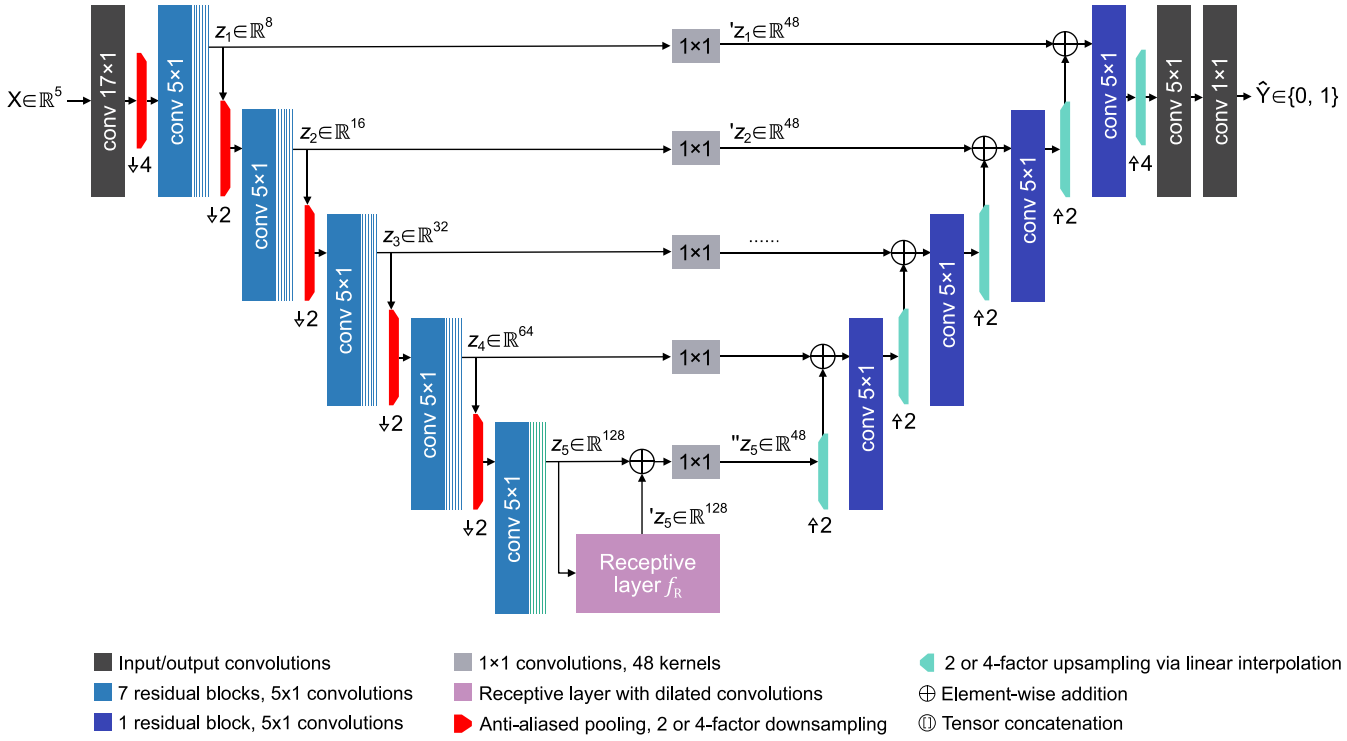


Fig. 3. The proposed temporal D-UNet incorporates five residual layers L, each comprising 5×1 1D convolutional (conv) kernels. Each encoding layer L consists of seven residual blocks with bottleneck connections, while a single residual block per each L was employed in the FPN. The downward and upward arrows alongside integers represent factors of temporal downsampling and upsampling, respectively. The extracted latent features are denoted by z_i. Receptive layer f_R consists of 3×1 dilated kernels.

depicted in Fig. 4.

Latent features z_5 are independently processed through multiple parallel branches employing dilated convolutions with exponentially increasing receptive fields. Individual receptive fields are modulated by the dilation factor $d = 2^k$, where $k \in \{x \in \mathrm{Z} \mid x \geq 1\}$, and k controls the rate of dilation growth. The receptive features $z_5^{(i)}$ are concatenated and compressed using 1×1 convolutions to form the transformed output $'z_5 = f_R(z_5)$ followed by an element-wise operation $z_5 + 'z_5$.

4.4. Optimization experiments

4.4.1. Optimal design selection (D-UNet model)

Bayesian optimization (BO) was employed to systematically identify the D-UNet architecture's optimal configuration. The search focused on optimizing the most relevant hyperparameters, namely the depth of the encoder (D_E), the depth of the residual layers in the encoder (D_L), the size of the filters in the residual layers (F_L), the size of the input convolution filters (F_{IN}), the number of 1x1 expansion/compression filters (N_C), and the cardinality (C_R) of the receptive layer f_R.

Non-linear activations, batch normalization, pooling operators, number of warm-up iterations, and other design choices were iteratively adjusted based on experimentation and maintained constant across all execution runs. More detailed BO settings and search spaces are provided within the Supplementary materials.

4.4.2. Performance-sensitive loss for imbalanced data (D-UNet-L model)

The loss function based on continuous approximation of the Sørensen-Dice coefficient [51] is a common choice for medical semantic segmentation tasks with deep neural networks, mitigating the negative impact of label sparsity. This study employs a generalized form approximating the Tversky index (TI), originally proposed by Abraham et al. [52] and further extended by Zhao et al. [53] to rebalance the computational cost based on the model's performance.

The loss introduces three hyperparameters α, β, and γ to amplify the contribution of misclassified samples. The exponent γ penalizes the model for misidentified hard-positive samples, further addressing sub-optimal segmentation performance in underrepresented objects. For the tensor $\widehat{Y}$ representing raw model scores and one-hot encoded labels $\widehat{Y}$, the focal Tversky loss $\mathscr{L}_{FTL}$ can be defined as:

$$\mathscr{L}_{FTL} = \sum_n [1 - (\widehat{Y} \odot T)/\{\widehat{Y} \odot T + \alpha \cdot [(1 - \widehat{Y}) \odot T] + \beta \cdot [\widehat{Y} \odot (1 - T)]\}]^{\gamma}, \tag{8}$$

where operator $\odot$ represents the Hadamard product of tensors.

A wide range of similarity metrics, including Jaccard and Sørensen-Dice indices, can be explored by adjusting α, β, and γ, creating a performance-sensitive loss function.

The asymmetry in the feature learning process associated with difficult-to-classify B-EGMs was addressed by optimizing parameters α, β, and γ using BO framework. A domain-specific $\mathscr{L}_{FTL}$ was then employed to train rebalanced D-UNet model (D-UNet-L). More detailed BO settings and search spaces are provided within the Supplementary materials.

4.4.3. Receptive layer ablation study (UNet-L model)

A straightforward ablation study was conducted to evaluate the impact of the proposed receptive layer f_R on the B-EGM delineation performance, particularly in the WAF signal group. The extended receptive field was disabled in the D-Unet-L model (UNet-L) by setting all the dilation factors d to 1 (no dilation). Layer f_R including non-dilated convolution filters remained in the architecture to preserve the same number of trainable parameters.

4.4.4. Label noise regularization (D-UNet-F0 model)

Globally assigning the positive class (for details see Section 4.1.3) to all AF segments in the WAF group (Y^+_{AF}) presents a significant optimization challenge. Individual B-EGM channels still contain atrial waves separated by isoelectric lines, introducing label noise during the learning process.

Removal of the entire WAF group from training dataset would seem a logical step to focus on delineating regular complexes, but it comes at the expense of discarding valuable semantically rich data. Label noise has been shown to mitigate memorization in deep neural networks [54], suggesting that an appropriate labeling strategy could potentially leverage this effect.

To evaluate the implicit regularization impact of the WAF labeling, separate D-UNet model (D-UNet-F0) was trained on the DSCV-Y dataset with all AF segments reassigned to the negative class Y^-_{AF}. Consistent with the optimization goal, this reassignment was also applied during validation and testing on the DSCV-Y, DST-Y, and DST-M datasets.

4.5. Statistical analysis

4.5.1. Sample-level evaluation metrics

The performance of each semantic segmentation model was assessed individually using established metrics such as the Sørensen-Dice coefficient (SDC), accuracy (Acc), recall, and precision. Additionally, Matthew's correlation coefficient (MCC) was employed as a measure of agreement between the models and the rater, in accordance with the recommendation of Butkuviene et al. [55] for sparse segmentation tasks. The metrics can be expressed as follows:

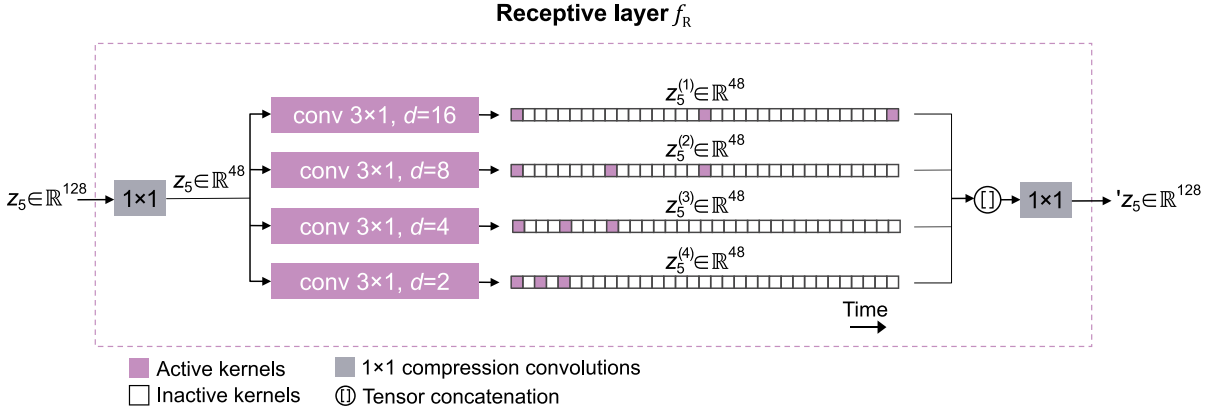


Fig. 4. Schematic representation of the receptive layer f_R incorporating four parallel 3×1 dilated convolutions (conv) into the D-UNet architecture. The dilation factor d increases exponentially across branches, with optimal d determined using Bayesian optimization. Latent features z_5 represent the output of the last encoding residual layer L; $'z_5 = f_R(z_5)$ denotes transformed features with temporal context expanded by dilated convolutions.

$$\mathrm{SDC} = 2\mathrm{TP}/(2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}), \tag{9}$$

$$\mathrm{Acc} = (\mathrm{TP} + \mathrm{TN})/(\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}), \tag{10}$$

$$\mathrm{Recall} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN}), \tag{11}$$

$$\mathrm{Precision} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FP}), \tag{12}$$

$$\mathrm{MCC} = [(\mathrm{TP}\cdot\mathrm{TN}) - (\mathrm{FP}\cdot\mathrm{FN})] \Big/ [(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})]^{\frac{1}{2}}, \tag{13}$$

where TP, TN, FP, and FN denote B-EGM samples classified as true-positives, true-negatives, false-positives, and false-negatives, respectively.

4.5.2. Instance-level evaluation metrics

To assess delineation errors across distinct SB, AB, and ST beat types, each continuous segment proposed by the model was handled as an instance of an atrial B-EGM, and its onset and offset were determined by identifying the first and last samples exceeding the threshold (limit 0.5). AF segments were excluded from the evaluation. In accordance with the ANSI/AAMI-EC57:1998 standard, only detections within the tolerance interval of ± 35 ms were considered for the analysis. The tolerance was chosen based on twice the standard deviation of the duration of the AB, SB, and ST beats in the DSCV-Y dataset.

Conventional regression metrics, including absolute errors (AE), mean error (ME), and root-mean-square error (RMSE), were employed to evaluate the delineation accuracy. The metrics can be formulated as follows:

$$\mathrm{AE_i} = |\widehat{y}_i - y_i|, \tag{14}$$

$$\mathrm{ME} = \frac{1}{N}\sum_{i}^{N}(\widehat{y}_i - y_i), \tag{15}$$

$$\mathrm{RMSE} = \left[\frac{1}{N}\sum_{i}^{N}(\widehat{y}_i - y_i)^2\right]^{\frac{1}{2}}, \tag{16}$$

where $\widehat{y}_i$ is the position of either onset or offset estimated by the model, and y_i is the corresponding reference position provided by the rater.

4.5.3. Delineation error grading

Delineation AEs were graded according to the methodology proposed by Haddad et al. [23] to assess the distribution of errors with clinically meaningful implications. This grading system, based on the evaluation of multiple raters, categorizes AEs as follows: highly accurate delineation for onsets and offsets detected by the algorithm with AEs ≤ 5 ms, accurate delineation for AEs ≤ 10 ms, and inaccurate detection for AEs greater than 10 ms.

4.5.4. Hypothesis testing

The Wilcoxon signed-rank test was employed to compare repeated measurements between two scenarios, where applicable. The relative importance (RI) of hyperparameters in achieving the highest model quality was determined using the functional analysis of variance (fANOVA) approach [56].

4.6. Implementation details

The AdamW algorithm [57] with decoupled L_2 regularization [58] was employed for the model optimization. Default moving average weights β_1 and β_2 (0.9 and 0.999 respectively), as recommended by Kingma et al. [57], were used since no other combinations improved the performance. The weights and biases of convolutional layers were initialized using the Kaiming method [59]. The normalization layers were initialized with c = 1 for weights and c = 0 for biases. During the first 8 training epochs, a linear warm-up phase was implemented. The initial learning rate (0.005) and a batch size (24) were determined through BO. The learning rate was scheduled using a reduce-on-plateau strategy with a decaying factor of 0.1. Training pipeline and auxiliary processes were implemented in the Python 3.8.1 environment [60] using the PyTorch 1.12.1 framework [61]. BO experiments were conducted using the Optuna 2.10.1 [62]. The model training was performed on the NVIDIA RTX 3060 12 G graphics card.

5. Results

5.1. Hyperparameters search

5.1.1. Optimal spatial design (D-UNet)

The optimal and lightweight architecture of the D-UNet with a total of 413,740 trainable parameters was identified within 200 BO runs. The performance assessed by SDC demonstrated a consistent decline with increasing model capacity. In terms of depths D_E and D_L, the optimized D-UNet model primarily benefited from 5 to 6 residual layers (RI 0.65) and 6–7 residual layers within a single layer L (RI 0.06). In terms of the model's width, performance deteriorated with larger F_L kernel sizes and an overparameterized number of features in the N_C, with the optimal range being 5–7 kernel size (RI 0.06) and 32–48 features (RI 0.13), respectively. Interactions between hyper-parameters together with their contributions to the monitored metric are depicted in Fig. 5.

5.1.2. Optimal loss function parameters (D-UNet-L)

The optimal domain-specific $\mathscr{L}_{FTL}$ weights were found within 80 BO repetitions, achieving optimal values of α = 0.4, β = 0.6, and γ = 0.5. Interactions between the weights and their contributions to SDC, precision, and recall are visualized in Fig. 6. Fig. 6 demonstrates that the parameters α and β, according to Equation (8), enable direct rebalancing of precision and recall directly within the training session (see Fig. 6).

The monitored SDC was primarily influenced by the selection of γ (RI 0.75). While a relatively broad interval with optimal values between 0.25 and 1.00 was discovered, a clear trend emerged toward increasing the relative contribution of B-EGMs with low Tversky index. This holds true for all $\gamma \ll 1.0$.

5.2. Comparison of sample-level performance

5.2.1. Sample-level performance on WOAF signals

A sample-level comparison of the proposed UNet designs is summarized in Table 2. While cross-validation results on DSCV-Y demonstrated comparable performance across all models, key metrics on the test datasets DST-Y and DST-M revealed a distinct advantage for D-UNet-L, which employs both dilated convolutions and domain-specific $\mathscr{L}_{FTL}$. D-UNet-L obtained SDC values of 88.0 % and 81.8 % on DST-Y and DST-M, respectively, indicating improved generalization capability. Inactivation of the receptive layer resulted in a 2.6 % and 4.4 % reduction in SDC on both DST-Y and DST-M, respectively. Among all the models, D-UNet-F0 trained with relabelled Y_{AF}^{-} exhibited notably lower SDC and MCC values, with decreases of 8.1 % and 6.8 % in SDC, and 0.089 and 0.059 in MCC, compared to D-UNet-L on the test sets DST-Y and DST-M. This demonstrates the positive regularization effect induced when the model is subjected to more complex B-EGMs associated with WAF signals.

5.2.2. Sample-level performance on WAF signals

Table 3 summarizes the segmentation performance on B-EGMs comprising atrial fibrillation. The D-UNet-L model achieved the best cross-validation performance with SDC of 93.5 ± 3.3 % and MCC of 0.725 ± 0.067. Consistent with the results in Section 5.2.1, the D-Unet-L also exhibited the highest generalization performance with SDC of 89.1 % on the test set DST-M. Inactivating both optimized $\mathscr{L}_{FTL}$ and f_R in

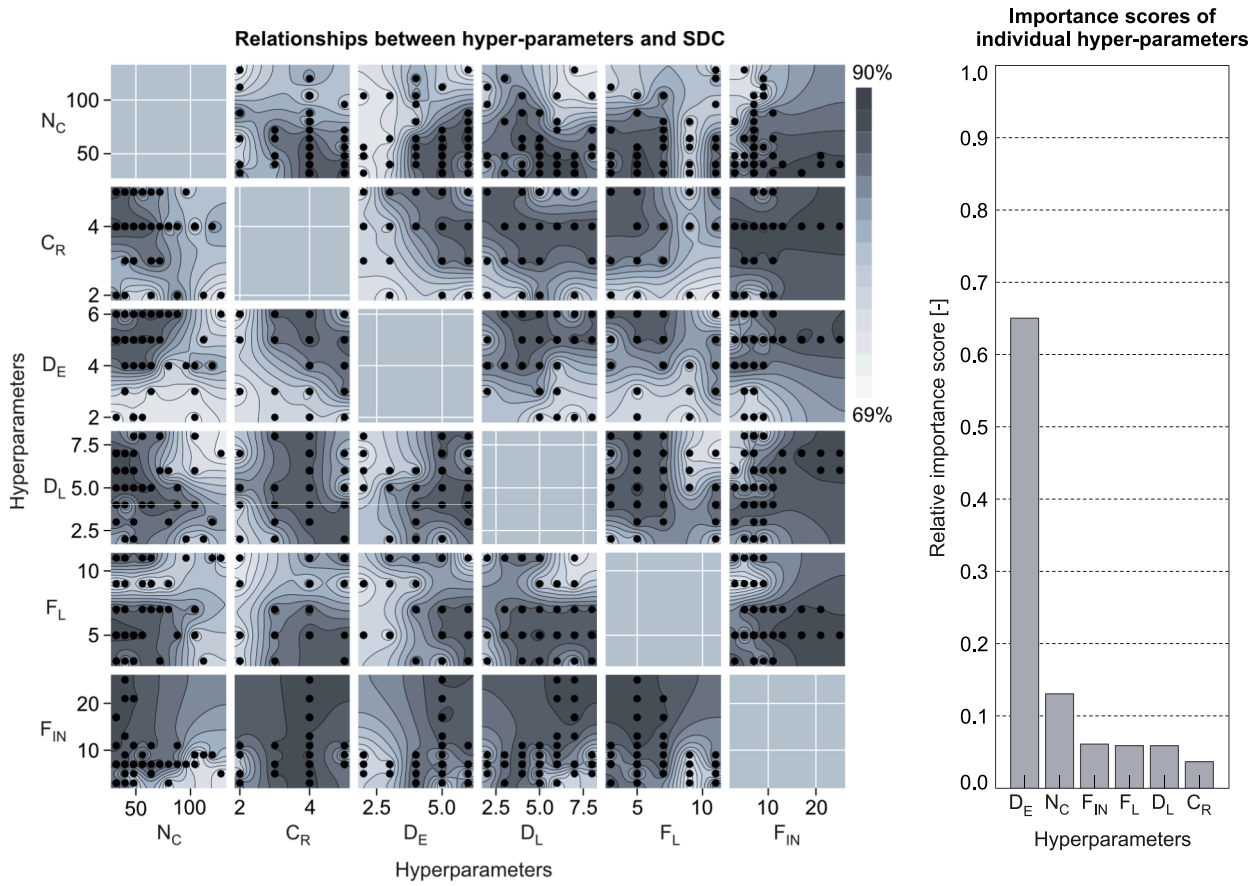


Fig. 5. Contour graphs (left panel) depicting the interactions between each pair of hyper-parameters and monitored Sørensen-Dice coefficient (SDC) as a results of Bayesian optimization (BO). Relative contribution of individual hyper-parameters to a BO optimization goal evaluated by fANOVA (right panel). The search focused on optimizing the most relevant hyperparameters, namely the depth of the encoder D_E – number of residual layers, D_L – number of residual blocks in the encoder residual layer, F_L – the size of the filters in the residual layers, F_{IN} – the size of the input convolution filters, N_C – the number of 1x1 expansion/compression filters, C_R – the cardinality of the receptive layer f_R.

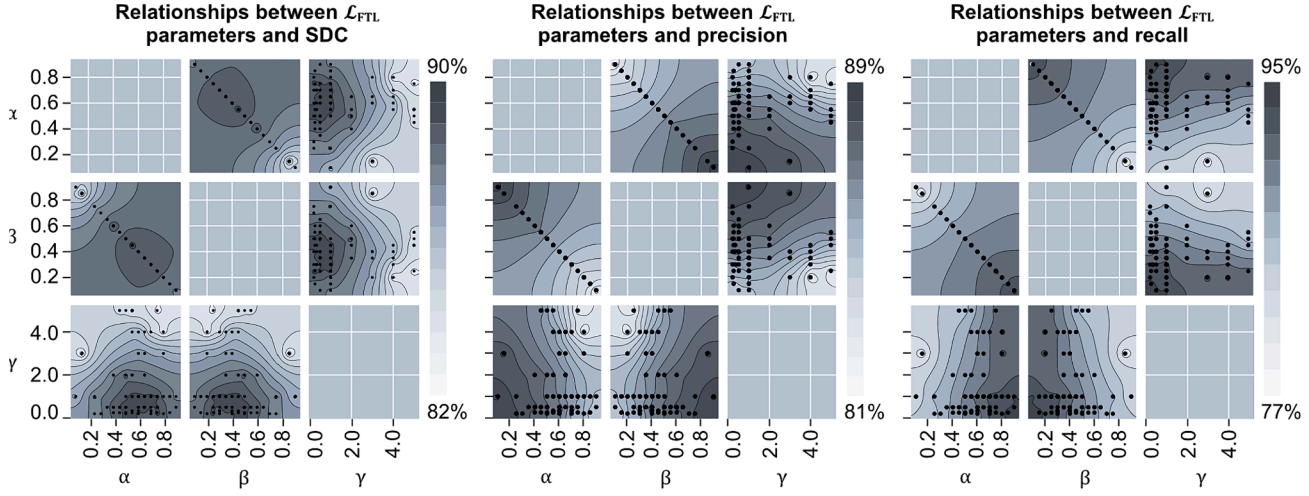


Fig. 6. Contour graphs depicting the interactions between $\mathscr{L}_{FTL}$ parameters α, β, γ and their contribution to the monitored Sørensen-Dice coefficient (SDC, left panel), precision (middle panel), and recall (right panel) as a results of Bayesian optimization.

the D-UNet and UNet-L resulted in a significant decline in the SDC to 79.6 % and 74.4 %, respectively.

While in the WOAF group (Section 5.2.1), all models except the D-UNet-F0 benefited from the presence of the Y^{+}_{AF}, the results of the D-UNet-F0 on the WAF group demonstrate better discrimination of the AF segments when assigned to the Y^{-}_{AF}. Among all models, the D-UNet-F0 achieved the highest overall accuracy (OAcc) of 97.9 ± 1.6, 99.8, and 92.0 on the DSCV-Y, DST-Y, and DST-M, respectively. Other key metrics could not be effectively determined for the D-UNet-F0 due to the absence of a positive class (AF) in the WAF group. It should be noted that

Table 2
Sample-level evaluation of semantic segmentation performance for B-EGMs containing isolated atrial beats without atrial fibrillation (WOAF). Cross-validation results for DSCV-Y expressed as mean ± standard deviation.

Model	Dataset	SDC (%)	Recall (%)	Precision (%)	MCC (-)	OAcc (%)
D-UNet	DSCV-Y	**87.0 ± 1.8**	89.9 ± 2.0	84.5 ± 3.6	**0.855 ± 0.021**	97.2 ± 0.6
	DST-Y	83.2	93.2	75.2	0.819	96.5
	DST-M	79.7	79.6	79.8	0.776	96.2
D-UNet-L	DSCV-Y	**87.0 ± 2.1**	87.0 ± 1.9	87.1 ± 4.3	**0.855 ± 0.025**	97.2 ± 0.6
	DST-Y	**88.0**	91.2	85.1	**0.868**	97.7
	DST-M	**81.8**	82.0	81.6	**0.799**	96.6
UNet-L	DSCV-Y	86.8 ± 2.9	86.0 ± 2.8	87.6 ± 4.1	0.852 ± 0.034	97.2 ± 0.8
	DST-Y	85.2	89.3	81.4	0.837	97.1
	DST-M	77.4	73.7	81.4	0.753	95.9
D-UNet-F0	DSCV-Y	85.1 ± 0.9	81.1 ± 2.2	89.7 ± 2.1	0.836 ± 0.010	97.1 ± 0.3
	DST-Y	79.9	80.0	79.9	0.779	96.3
	DST-M	75.0	64.7	89.4	0.740	95.9

Table 3
Sample-level evaluation of semantic segmentation performance for B-EGMs containing atrial fibrillation (WAF). Cross-validation results for DSCV-Y expressed as mean ± standard Deviation (NA: metric not applicable for unlabeled AF category).

Model	Dataset	SDC (%)	Recall (%)	Precision (%)	MCC (-)	OAcc (%)
D-UNet	DSCV-Y	92.9 ± 3.2	95.8 ± 2.9	90.6 ± 6.5	0.713 ± 0.028	89.5 ± 3.9
	DST-Y	98.5	99.8	97.2	0.795	97.2
	DST-M	79.6	66.5	99.2	0.072	66.5
D-UNet-L	DSCV-Y	**93.5 ± 3.3**	98.2 ± 1.4	89.6 ± 6.4	**0.725 ± 0.067**	90.3 ± 4.1
	DST-Y	98.5	99.9	97.1	0.796	97.2
	DST-M	**89.1**	81.0	99.0	**0.078**	80.5
UNet-L	DSCV-Y	90.0 ± 4.7	89.2 ± 8.7	91.8 ± 6.0	0.656 ± 0.058	86.0 ± 5.8
	DST-Y	**98.6**	99.6	97.6	**0.813**	97.4
	DST-M	74.4	59.7	98.8	0.024	59.5
D-UNet-F0	DSCV-Y	NA	NA	NA	NA	**97.9 ± 1.6**
	DST-Y	NA	NA	NA	NA	**99.8**
	DST-M	NA	NA	NA	NA	**92.0**

some of the key results are likely inconclusive. On DST-Y, only a limited number of WAF signals were available, introducing highly biased results. Moreover, a presence of a small amount of mislabeled true positives because of manual annotation in the DST-M caused a significant drop in the MCC.

5.2.3. Cross-signal variations in MCC and Acc

Fig. 7 depicts the cross-signal variations in segmentation performance across different model designs. In the WAF group D-UNet-L overperforms both D-UNet and UNet-L in term of the MCC, however the metric can be only considered valid for the cross-validation dataset DSCV-Y. Here, D-UNet-L has the highest median MCCs 92.9 % (IQR 81.0–95.5 %) compared to 84.8 % (IQR 31.7–92.9 %) and 82.4 % (IQR 19.2–93.2 %) achieved by the D-UNet and UNet-L, indicating large variations across different WAF B-EGMs. While D-Unet-F0 systematically underperforms in the WOAF group, it is highly capable in correctly classifying AF segments into negative class Y^-, as demonstrated by very low dispersion in Acc with IQR concentrated between 96.6 and 100 % and 88.3–95.2 % on DSCV-Y and DSCV-M, respectively.

In the WAF group, D-UNet-L outperformed both D-UNet and UNet-L in terms of MCC. However, this metric is only valid for the cross-validation dataset DSCV-Y. Here, D-UNet-L achieved the highest median MCC of 92.9 % with an IQR of 81.0–95.5 %, compared to 84.8 % for D-UNet (IQR 31.7–92.9 %) and 82.4 % for UNet-L (IQR 19.2–93.2 %), indicating large variations across different WAF recordings. While D-Unet-F0 systematically underperformed in the WOAF group, it exhibited exceptional capability in correctly classifying AF segments into the negative class Y^-, as evidenced by very low dispersion in Acc with IQRs concentrated between 96.6 and 100 % and 88.3–95.2 % on DSCV-Y and DSCV-M, respectively.

5.3. Comparison of instance-level performance

5.3.1. Per-beat delineation errors

Table 4 summarizes the delineation errors of the atrial B-EGMs boundaries. After excluding detections outside the tolerance interval (see Section 4.5.2), the results demonstrate relatively consistent errors across all models. The D-UNet-L exhibited a marginally lower overall RMSE on datasets comprising the Y cohort, with values of 7.4 ± 0.5 ms and 7.2 ms on the DSCV-Y and DST-Y, respectively. In individual categories, the D-UNet-L and its non-dilated variant, UNet-L, consistently provided slightly more accurate results, particularly on the test sets DST-Y and DST-M. For instance, the D-UNet-L exhibited exceptional capability in estimating the onset of abnormal beats (AB), with the lowest RMSE of 4.5 ms on DST-Y and 6.7 ms on DST-M.

Both D-UNet-L and UNet-L also demonstrated improved delineation errors for the B-EGMs offsets. In this instance, the UNet-L without receptive field achieved a slightly superior overall RMSE of 7.1 ± 0.1 on DSCV-Y compared to the D-UNet-L's 7.8 ± 0.4. This trend reversed on DST-Y, where the D-UNet-L outperformed UNet-L with an overall RMSE of 5.7 ms compared to 6.3 ms. On DST-M, which consisted of adult patients, the results were comparable across all models.

5.3.2. The delineation accuracy grades

Fig. 8 summarizes the distribution of accuracy grades across all models. The results indicate that between 60 % and 80 % of the boundaries can be delineated with a deviation of ≤ 10 ms (highly accurate or accurate) in most cases, across all datasets and models, excluding stimulated (ST) atrial B-EGMs. The left panel of Fig. 8 highlights a significant drop in delineation accuracy in DST-Y. Upon visual inspection, we discovered significant systematic overestimation of the ST annotations, suggesting high intraindividual variability in localizing B-EGMs onsets in the presence of pacing artifacts.

Another substantial decrease in measurement accuracy was observed

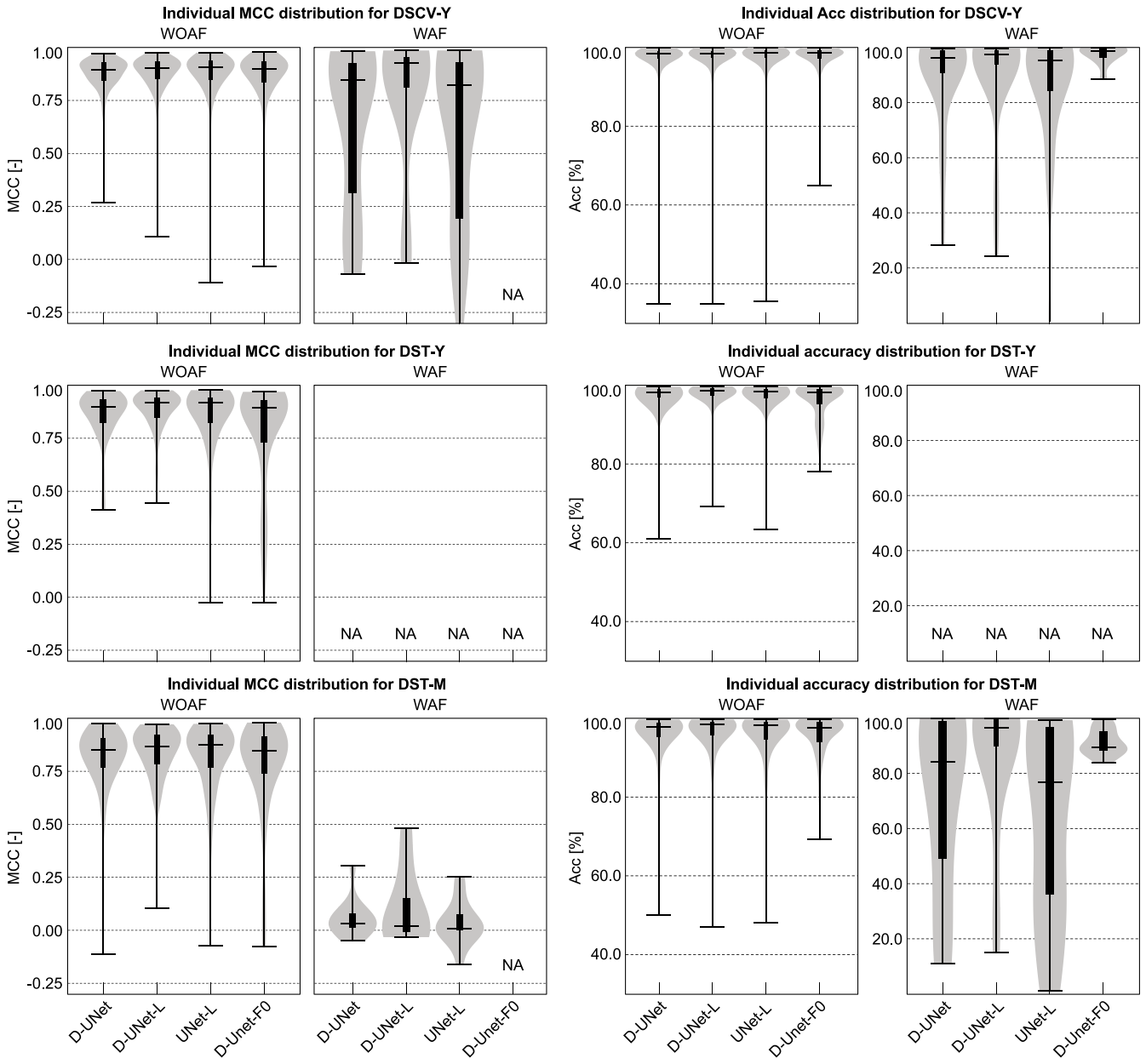


Fig. 7. Violin graphs representing the distribution of the Matthew's correlation coefficient (MCC, left panels) and accuracy (Acc, right panel) across individual B-EGM recordings. Both, MCC and Acc are depicted independently for DSCV-Y (top row), DST-Y (middle row), and DST-M (bottom row) datasets. Whiskers represents minimum and maximum value of monitored metrics, middle horizontal line and black rectangle denotes the median and inter-quartile ranges, respectively. NA: metric not available for unlabeled AF category (MCC for D-UNet-F0 on DST-M) or limited number of WAF signal available (MCC and Acc on DST-Y WAF group).

between the non-dilated UNet-L and models with an increased receptive field. The combined highly accurate and accurate delineation of ST offsets dropped from 72.8 % on DST-M to 44.7 %, 49.4 %, and 41.4 % for D-UNet, D-UNet-L, and D-UNet-F0, respectively. Furthermore, D-UNet-F0 demonstrated higher delineation errors, with a combined score of 34.4 %, on DST-U as well. This further supports the hypothesis of possible overfitting of the model in specific cases if not subjected to recognizing WAF B-EGMs as atrial activity during training.

All models performed well in the case of the AB B-EGMs onset, which is crucial for the extraction of clinically important CLs during pathological rhythms. In this instance, the non-dilated UNet-L and the reference D-UNet-L exhibited the best performance, with average delineations below 10 ms for 90.4 % and 84.5 % of onsets, respectively. Overall, the reference model D-UNet-L achieved 61.5 % highly accurate delineations, 20.8 % accurate delineations, and 17.7 % inaccurate delineations across all datasets, B-EGM groups, and boundaries.

Fig. 9 demonstrates the detection and delineation capability of D-UNet-L with challenging B-EGMs involving ongoing arrhythmias. These arrhythmias include AV dissociation, transition from sinus rhythm to rapid atrial fibrillation, and coinciding AV beat during AV node reentry tachycardia. The fusion of atrial complexes with ventricular complexes can be observed in each case.

5.4. Data augmentations

The impact of data augmentation (DA) transformations was assessed through an ablation study using the first two subsets of DSCV-Y for training and the third subset for validating the D-UNet-L model. Comprehensive findings of the DA ablation study are presented in the Supplementary material. Most notably, the difference between the

Table 4
Instance-level evaluation of atrial complex delineation performance for SB, AB, and ST B-EGMs groups. Cross-validation results for DSCV-Y expressed as mean ± standard deviation (RMSE – root mean squared error).

Model	Dataset	Onset RMSE (ms)				Offset RMSE (ms)			
		Overall	SB	AB	ST	Overall	SB	AB	ST
D-UNet	DSCV-Y	8.0 ± 0.2	8.1 ± 0.1	8.1 ± 0.2	5.6 ± 1.8	8.3 ± 0.4	6.6 ± 0.6	9.5 ± 0.7	**8.7 ± 1.4**
	DST-Y	9.2	6.2	6.3	19.2	6.1	**4.5**	6.4	8.7
	DST-M	10.5	10.1	11.3	9.8	**11.6**	**10.4**	12.5	13.3
D-UNet-L	DSCV-Y	**7.4 ± 0.5**	**7.9 ± 0.5**	**7.2 ± 0.6**	**5.6 ± 1.0**	7.8 ± 0.4	6.1 ± 0.5	8.5 ± 0.9	11.1 ± 0.8
	DST-Y	**7.2**	6.1	**4.5**	**14.2**	**5.7**	4.7	**4.7**	**7.9**
	DST-M	9.3	**9.5**	9.5	8.0	12.2	10.6	13.5	14.1
UNet-L	DSCV-Y	7.6 ± 0.2	8.0 ± 0.4	7.4 ± 0.5	5.6 ± 1.6	**7.1 ± 0.1**	**5.7 ± 0.5**	**7.8 ± 0.2**	9.2 ± 1.2
	DST-Y	7.5	**5.6**	5.3	16.3	6.3	4.7	6.0	11.3
	DST-M	9.4	11.0	**6.7**	**6.7**	11.9	10.8	13.3	**12.6**
D-UNet-F0	DSCV-Y	8.3 ± 0.6	8.4 ± 0.7	8.2 ± 0.2	6.5 ± 2.4	7.5 ± 0.2	**5.7 ± 0.5**	8.5 ± 0.5	9.4 ± 0.5
	DST-Y	7.7	6.5	6.1	24.0	9.4	5.1	7.3	21.7
	DST-M	**8.9**	10.2	7.0	7.4	12.3	11.8	**12.3**	14.1

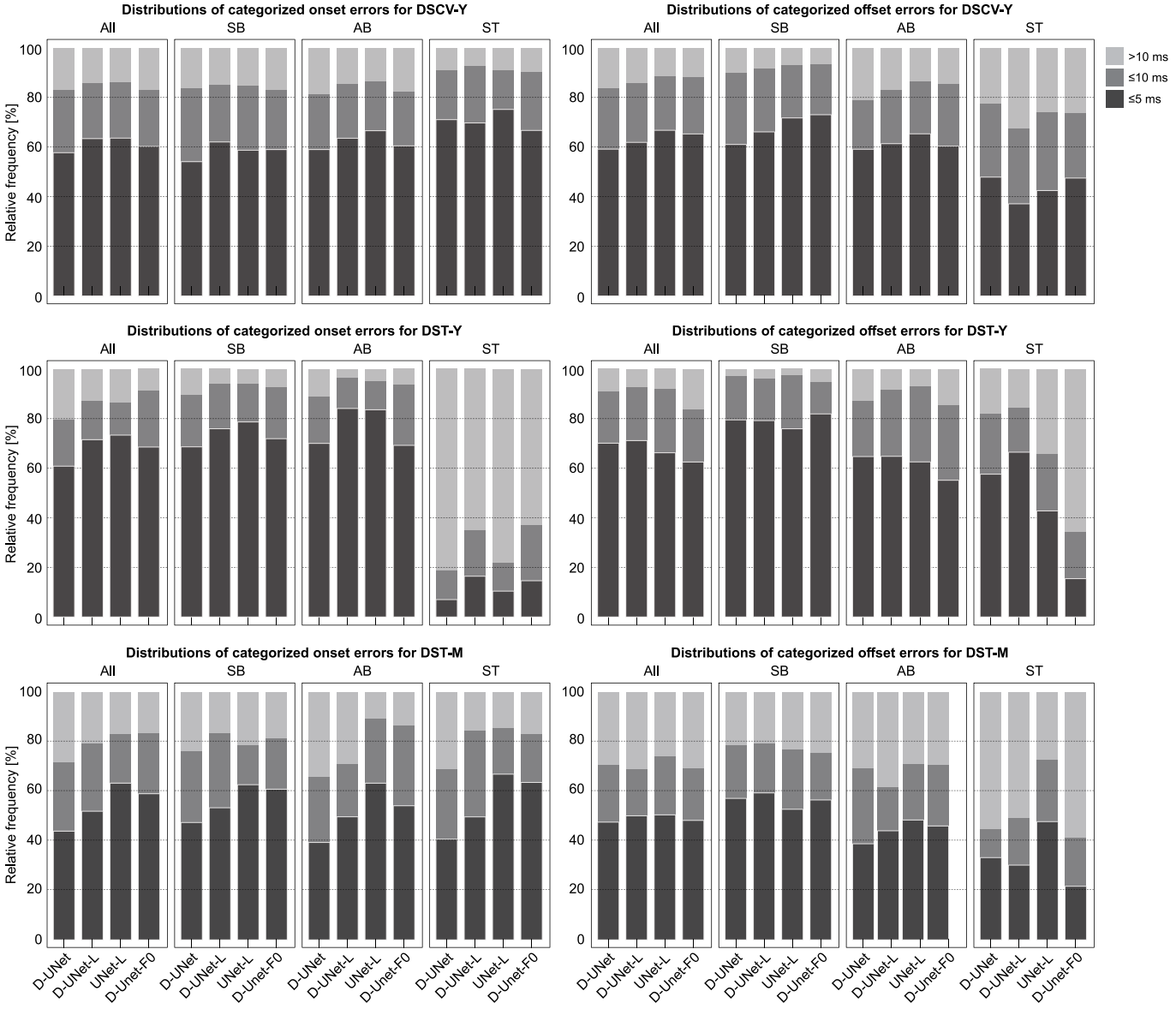


Fig. 8. Stacked bar charts summarizing the differences in the relative distributions of highly accurate (≤5 ms, dark gray rectangle), accurate (≤10 ms, gray rectangle), and inaccurate (>10 ms, light gray rectangle) delineation errors for B-EGMs onsets (left panel) and offsets (right panel) across the DSCV-Y, DST-Y, and DST-M datasets (top, middle, and bottom rows, respectively).

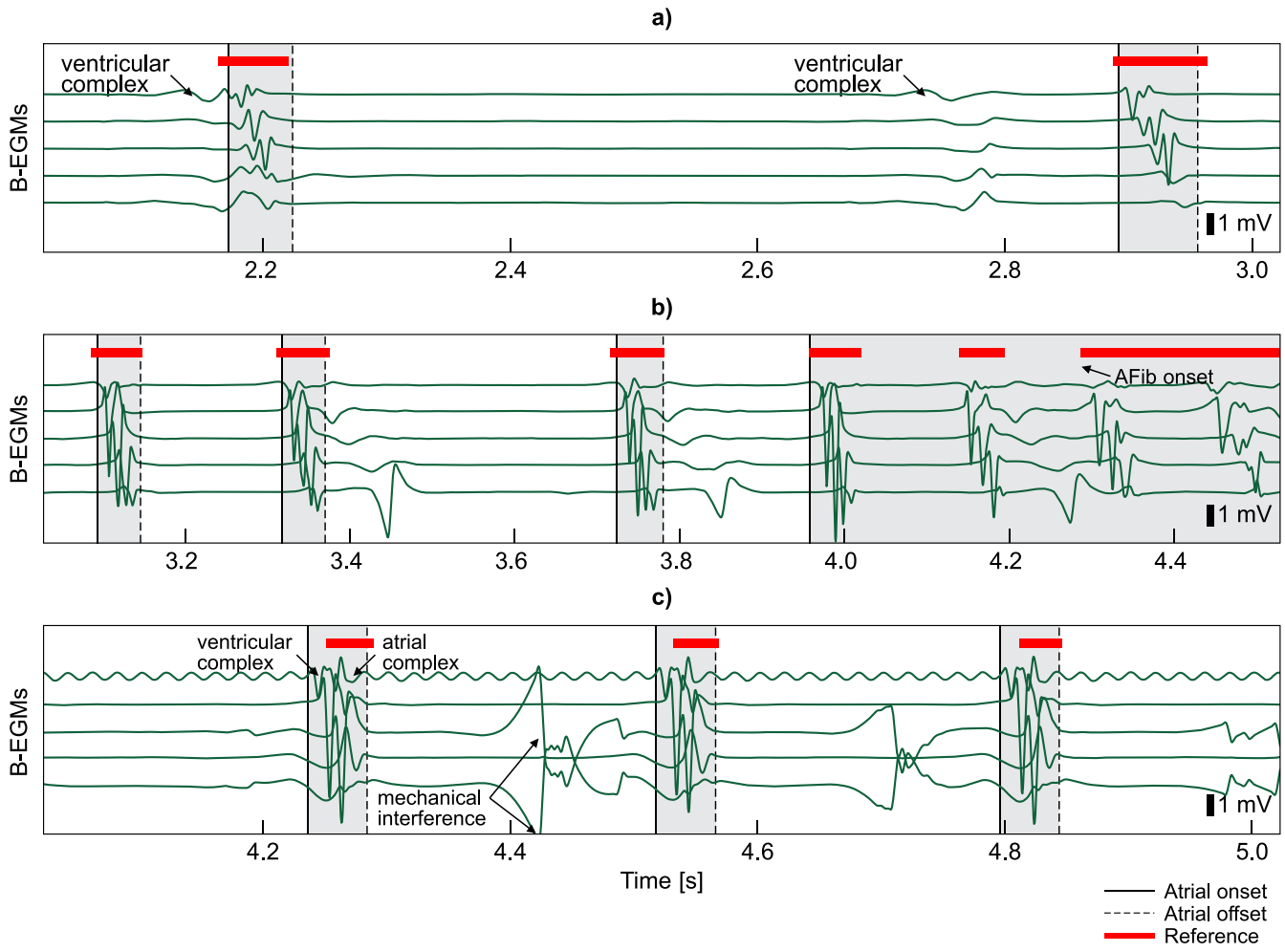


Fig. 9. Detection and delineation performance of D-UNet-L in challenging B-EGMs comprising of ongoing arrhythmias with fused atrial (red marked segments) and ventricular complexes (low frequency complexes in top and middle panel; high frequency complexes preceding atrial activation in the bottom pane). From top to bottom: AV dissociation (a); transition from sinus rhythm with premature atrial beats to atrial fibrillation (b); and a coinciding AV beat during AV node reentry tachycardia (c). In all cases properly detected interatrial CLs may help raise notification during the procedure for possible intervention. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

training and validation SDCs was −0.3 % with the DA pipeline and 7.7 % after complete ablation. Disabling DAs resulted in a 2.7 % decrease in overall validation SDC and a 5.3 % increase in training SDC, highlighting the significance of DAs in this work.

5.5. Robustness of artificial noise

The robustness of the method model was evaluated for a reference model D-UNet-L under varying noise conditions by independently introducing synthesized Gaussian white noise (GWN) and 50 Hz power line interference (PLI) to each signal at decreasing SNR levels ranging from 20 dB to 0 dB. PLI is a prevalent interference commonly observed in ECG/EGM signals. While it can be effectively mitigated using band-stop filters, it can induce distortions in the B-EGM morphology. GWN, on the other hand, poses a unique challenge by resembling the random nature of AF B-EGMs and slow-propagating waveforms in scar tissue. The noise was added to all accessible leads, and its intensity was adjusted according to the equivalent SNR in each channel. For this specific testing scenario, the D-UNet-L model was trained on the first two subsets of the DSCV-Y dataset, while the third subset was utilized to assess its noise robustness.

5.5.1. The effect of noise on a sample-level performance

Fig. 10 shows sample-level performance of D-UNet-L with SDC gradually degrading when signal quality decreases up to a critical level of 5 dB for both PLI, and particularly GWN. In the latter case the SDC was substantially reduced from 88.2 % to 83.5 % after reaching SNR of 0 dB. This was primarily due to elevated frequency of false positives represented by a drop in precision by 9.1 % and 1.5 % for GWI and PLI, respectively.

5.5.2. The effect of noise on the delineation errors

In correctly identified atrial complexes, no statistically significant change was observed in delineation errors (represented by ME and RMSE) compared to errors without artificial noise, up to an SNR of 0 dB, as depicted in Fig. 11. A 5 dB noise level was identified as a critical threshold, particularly in determining atrial offset. A further reduction in SNR (to 0 dB) resulted in a systematic deviation from reference values, with ME increasing from 1.52 to 2.61, and 1.39 to 2.17 ms for GWN and PLI, respectively. Simultaneously, RMSE worsened for both GWN and PLI, reaching 9.60 ms and 8.73 ms, respectively, compared to 8.28 ms and 7.96 ms at 5 dB. The only significant difference in the variance of residuals tested ($\alpha = 0.01$ after Bonferroni correction) against the baseline state (NA) was found in the determination of atrial offset for both GWN ($p < 0.0001$) and PLI ($p < 0.001$).

Figs. 12 and 13 showcase the delineation performance in high- and low-quality B-EGMs. In Fig. 12, the D-UNet-L model demonstrates remarkable performance in heavily contaminated B-EGMs. In contrast,

$

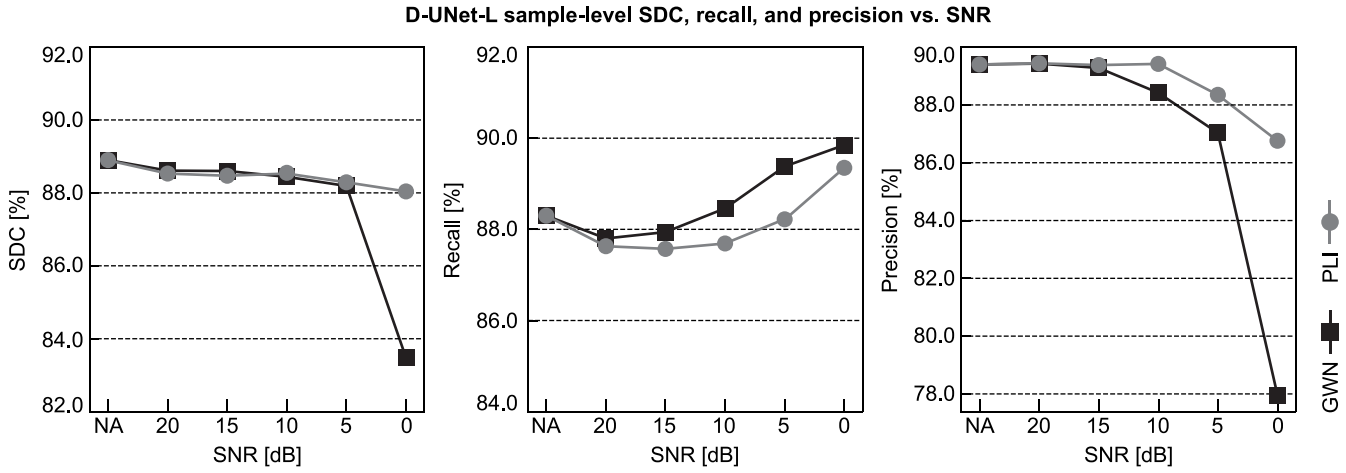


Fig. 10. Sample-level semantic segmentation performance analysis of the D-UNet-L model under varying signal-to-noise ratio (SNR) conditions: The D-UNet-L model's performance is evaluated using Dice score (left panel), recall (middle panel), and precision (right panel) metrics, with Gaussian white noise (GWN, black rectangle) and power line interference (PLI, grey circle) introduced independently at decreasing SNR levels from 20 dB to 0 dB (NA – default conditions without artificial noise).

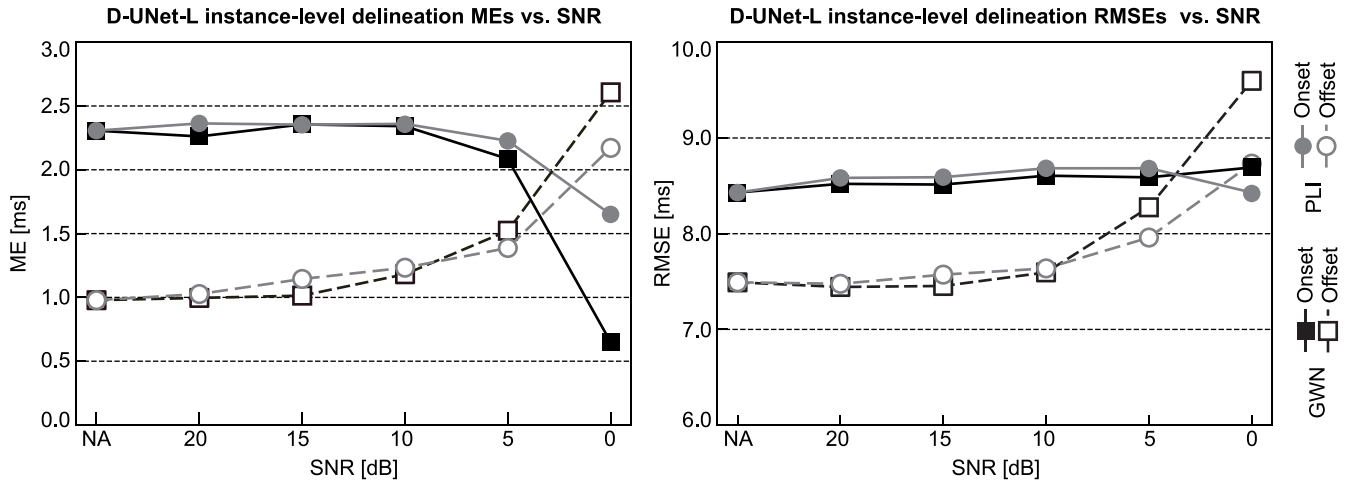


Fig. 11. Instance-level delineation performance analysis of the D-UNet-L model under varying signal-to-noise ratio (SNR) conditions: The delineation performance is evaluated independently for atrial complex onset (black rectangles) and offset (grey circles) using ME (left panel) and RMSE (right panel) metrics, with Gaussian white noise (GWN) and power line interference (PLI) introduced independently at decreasing SNR levels from 20 dB to 0 dB (NA – default conditions without artificial noise).

Fig. 13 presents a purposefully selected low-voltage atrial B-EGMs polluted with GWN and two distinct types of failures associated with the D-UNet-L and D-UNet-F0 models.

6. Discussion

In complex mapping procedures, numerous individual beats may be omitted from the annotation process due to predefined validation criteria and a narrow user-defined window of interest. The available filtering and detection techniques offered by modern electrophysiological acquisition systems often prove inadequate in handling partially fused beats, thus, fully manual measurements still prevail as a result. It can be attributed to several factors, including a lack of confidence in the measurement reproducibility of automated methods, the computational complexity involved, and the requirement for certification processes.

Deep learning techniques have exhibited noteworthy advancements in performance for various challenging segmentation tasks within the medical domain, thereby aiding decision-making and improving time efficiency [63]. However, they are prone to overfitting when provided with insufficient amount of variable data. By employing intrinsic regularization techniques and architectural optimization adjustments, a lightweight deep neural model for B-EGMs segmentation with robust generalization capability and resilience to typical EP laboratory interference can be established even with a relatively constrained dataset.

6.1. Performance analysis

The D-UNet-L model demonstrates robust detection and delineation performance comparable to manual annotation, while offering distinct advantages over existing techniques due to its ability to handle diverse fusion beat types and extended sequences with irregular F-waves. The model exhibits significantly superior generalization capability across diverse patient cohorts and subsets of the data compared to other explored variants. This is evidenced not only by overall evaluation metrics but also by reduced performance variability across individual B-EGM recordings, despite a slightly higher RMSE in estimating

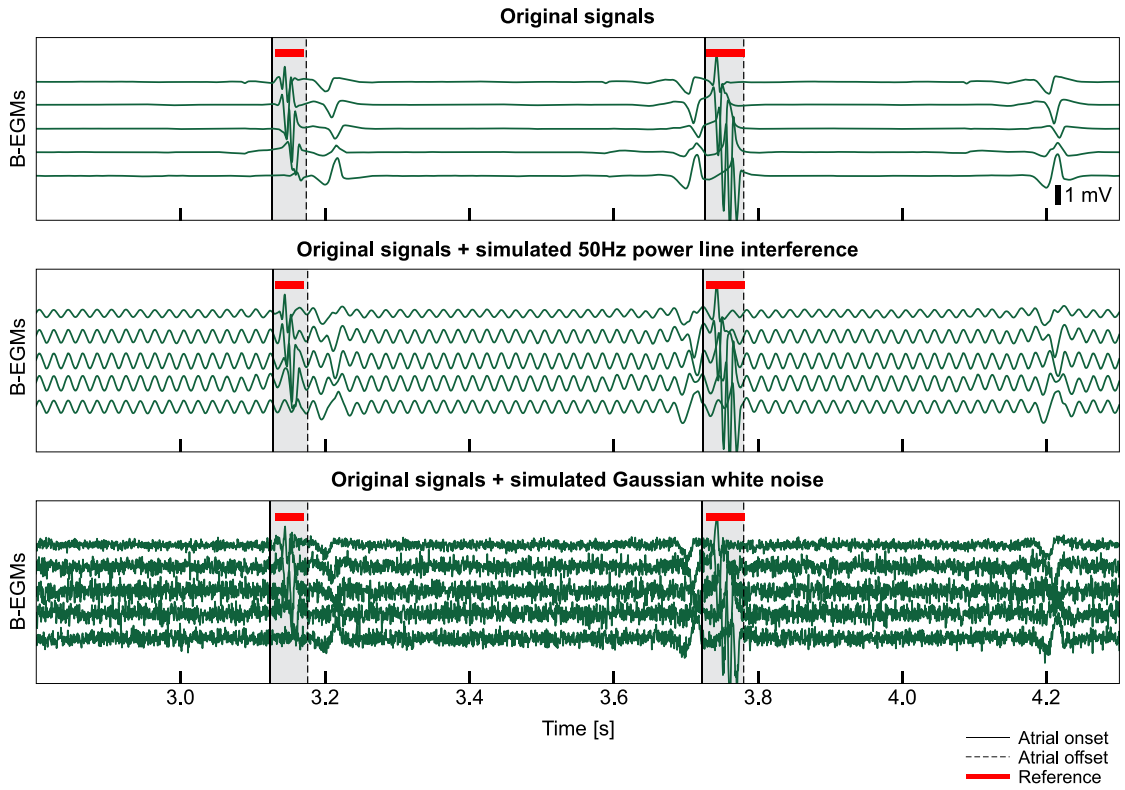


Fig. 12. A comparison of the detection and segmentation performance of the D-UNet-L model is presented for both high (top panel) and low-quality signals (middle and bottom panel). B-EGMs were corrupted with 50 Hz power line interference (middle panel) and additive Gaussian white noise (bottom panel) to achieve a signal-to-noise ratio of 0 dB. Minimal deviation from the reference can be observed for both cases.

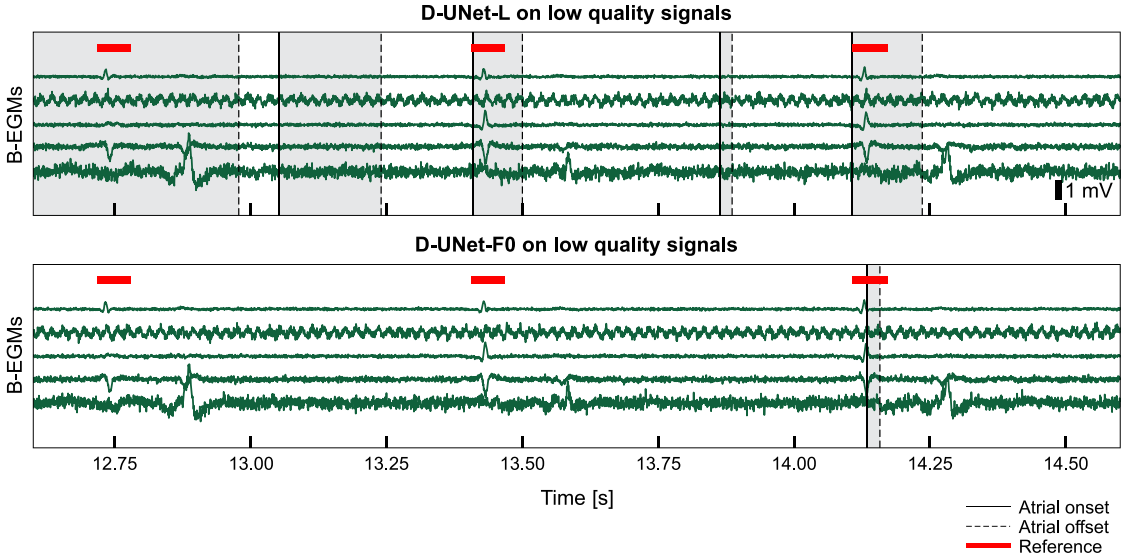


Fig. 13. The segmentation performances of two different models, D-UNet-L (top panel) and D-UNet-f0, are compared on low-quality B-EGMs. D-UNet-L, trained with AF B-EGMs labeled as the positive class, exhibited misclassification of isoelectric segments subjected to Gaussian white noise. In contrast, D-UNet-f0, trained with AF B-EGMs marked as the negative class, exhibited enhanced resistance to random noise but still failed in atrial complex detection. Note the peak-to-peak amplitude of B-EGMs in this case is substantially lower (below 1 mV) compared to normal B-EGMs. The SNR for both signals was 0 dB.

electrogram offsets. The most consistent delineation errors are observed across different types of atrial beats, which can be attributed to the increased morphological variations, rebalanced loss function, and more challenging optimization goals imposed by artificial label noise.

6.1.1. Wide receptive layer

The incorporation of a wide receptive layer with exponentially increasing dilation factor in convolution kernels enhanced the semantic analysis of extended sequences with atrial fibrillation in the WAF group, as demonstrated by a decline in performance of the ablated UNet-L model on DSCV-Y and DST-M, the datasets that primarily contained WAF B-EGMs. Surprisingly, higher segmentation scores were also observed for dilated models in groups of regular atrial B-EGMs. Nevertheless, further analysis revealed an increase in onset and offset

measurement errors compared to the non-dilated UNet-L, which achieved a slightly larger proportion of errors below 5 ms in multiple B-EGM types. This may imply improved discrimination of isolated B-EGMs from other complexes and interference, albeit at the cost of measurement accuracy.

6.1.2. Exponentially weighted Tversky loss

The imbalance in the label distribution can lead to a higher mis-segmentation rate for objects of interest that are inherently small or have intrinsic features that make them more difficult to segment. Tversky loss encourages the model to produce spatially coherent segmentations by focusing on overlap rather than sample-wise modeling of conditional probabilities, similarly to other geometrical loss functions using Sørensen-Dice or Jaccard metrics. By penalizing false positives more than false negatives with a β of 0.6, the trade-off between precision and recall was significantly improved compared to the widely used Sørensen-Dice loss ($\alpha = 0.5$, $\beta = 0.5$, $\gamma = 1$) for all models trained with the rebalanced Tversky loss in the segmentation of WOAF B-EGMs. Notably, the trade-off remained practically the same across test sets DST-Y and DST-M. No comparable generalization was achieved through manual optimization of the decision threshold, a traditional technique for improving model performance without directly affecting trained features. Thus, we hypothesize that, specifically for our dataset, utilizing the rebalanced loss function during training enables the model to learn a better latent representation and generate well-calibrated conditional probabilities.

Further improvement in discriminating isolated beats within the WOAF group was attained by optimizing the trade-off between penalties for challenging positive B-EGM samples. For an exponential weight γ less than 1, the Tversky loss diminishes the contribution of easy to segment B-EGMs on the overall penalization. This step proved to be beneficial for highly imbalanced data and has been previously reported in [52,53]. Exploratory analysis provided in this study revealed enhanced delineation accuracy across multiple datasets. The lowest delineation RMSE values for both overall and individual beats (SB, AB, and ST) were achieved for two models optimized with rebalanced Tversky loss, the D-UNet-L and UNet-L. Both models also demonstrated a greater proportion of highly accurate and accurate delineation grades. Nonetheless, there remains substantial room for improvement, particularly in ST beats and the ambiguous B-EGMs offsets, which poses a challenge even for expert annotation. Inaccurate boundary adjustments could potentially be addressed by incorporating region proposal networks [64,65] and extending the optimization objective towards bounding box regression with penalization based on more sophisticated geometric measures.

6.1.3. Labeling of atrial fibrillation

Annotating AF signals as a positive class enhanced the model's regularization and its capacity to recognize regular beats. This can be attributed to the heightened morphological variations and more demanding optimization objectives presented by the artificial label noise introduced by this labeling strategy. The F-waves may exhibit either randomly distributed electrical activity or more structured complexes with fluctuating amplitude and duration. By treating these occurrences as a continuous segment irrespective of signal complexity, it is plausible that overconfidence bias was attenuated. However, this also resulted in a significant decrease in accurately classified signal samples within the WAF group compared to the D-UNet-F0 model trained with inverse labeling.

Exploring alternative labeling strategies, such as a multi-class approach for differentiating regular B-EGM complexes from irregular AF segments, may provide fresh perspectives on the regularization effect. However, we faced difficulties in achieving convergence for this scenario owing to the inadequate number of WAF segments available.

6.2. Generalization in low quality signals

El Haddad et al. [24] determined a signal-to-noise ratio (SNR) cutoff of 12.8 ± 1.3 dB to be critical for distinguishing between low and high-quality signals using traditional thresholding methods. Our study examined the behavior of the D-UNet-L reference model in the presence of artificial white noise (GWN) and power-line interference (PLI).

The D-UNet-L maintained satisfactory performance down to an SNR of 5 dB in detecting the onset and offset of B-EGM complexes. However, below this threshold, the correct identification of isolated atrial beats significantly deteriorated, as did the error between detected boundaries and reference annotations. It is worth noting that noise-aware DA was employed during the training, setting the SNR limit to 5 dB for both GWN and PLI.

Exploring different DA parameters or implementing adaptive training schedules could potentially enhance the model's robustness in specific scenarios. However, conducting an extensive study in this area would require a substantial number of additional experiments, which exceeded the scope of this work. Thus, a constant configuration for DA pipeline was maintained during all experiments.

Although assigning WAF signals to the positive class improved the overall SDC, it made the model more vulnerable to the tested types of interference, resulting in an increased number of false positive detections as the signal quality degraded. Further analysis revealed that the D-UNet-L model often confused GWN for F-waves, particularly in the presence of low SNR. This issue was mitigated in the D-UNet-F0 model, which was trained without annotated WAF segments and prioritized the accurate identification of discrete atrial complexes.

6.3. Evaluating the performance relative to other studies

Several challenges hinder the fair comparison of automated EGM detection and delineation algorithms. Up to this point, most datasets employed by other researchers have consisted of highly selective patient cohorts comprising only a limited number of individuals, typically in the tens. Moreover, a significant portion of these datasets remain inaccessible to the public, precluding a comprehensive performance evaluation. The restricted number of patients often necessitates the fine-tuning of the algorithm on the same dataset utilized for testing, introducing potential biases.

Existing works have predominantly focused on specific procedural aspects of the EP study. For instance, Nothstein et al. [14] primarily concentrated on analysing restitution curves during programmed stimulation protocols, while Hajimolahoseini et al. [28] evaluated their algorithm on single-lead B-EGMs acquired during atrial fibrillation. Alcaine et al. [22] employed ventricular B-EGMs from patients to design an automated algorithm for mapping idiopathic outflow tract ventricular arrhythmias, Treo et al. [18] utilized electrograms from canine hearts, and Haddad et al. [23] exclusively focused on patients with atrial tachycardia in their study.

Hajimolahoseini et al. [28] achieved overall standard deviations (SDs) of 9.0 ms and 12.6 ms, respectively, for the delineation of B-EGM onset and offset. Alcaine et al., in two separate studies, obtained SDs of 12.3 ms [29] and 10.9 ms [22] for LAT annotations in ventricular B-EGMs. While both methods demonstrate comparable performance, delineation of B-EGMs with irregular F-waves in [28] can be regarded as more challenging compared to ventricular B-EGMs with significantly higher signal-to-noise ratio (SNR) as explored in [29]. In contrast, our referential model, D-UNet-L, achieved average root mean square errors (RMSEs) of 8.0 ms and 8.6 ms for B-EGM onset and offset delineation, respectively, across all datasets. This suggests significantly lower delineation errors, albeit for different case scenarios and utilizing advanced ML models together with information from all available leads.

Nohstein et al. [14] reported absolute measurement errors between 3.5 and 7.5 ms in 50 % of cases, depending on electrode placement. Alcaine et al. [22] reported that 51 % of LAT detection errors

were ≤ 5 ms, 23 % were ≤ 10 ms, and 26 % were > 10 ms. Among the compared studies, the work of Nohstein et al. [14] exhibits the most similar objectives. In this regard, our D-UNet-L based method outperforms the previous method, achieving 61.5 % of onset and offset delineations combined with errors less than 5 ms, 20.8 % with errors ≤ 10 ms, and 17.7 % with errors > 10 ms. With a similar grading system, Haddad et al. [23] reported higher overall accuracy in both high- and low-quality B-EGMs, with 80 % of onset errors ≤ 5 ms, 15 % with errors ≤ 10 ms, and 5 % with errors greater than 10 ms. It is important to note, however, that these results are reported for a single type of regular arrhythmia.

The general lack of comprehensive evaluation across diverse datasets, patient cohorts, and procedural settings makes thorough assessment of the true capabilities of automated EGM detection and delineation algorithms challenging. To address these limitations, future research should focus on developing standardized datasets and evaluation protocols that encompass a wider range of patient populations.

6.4. Limitations

Our study has yielded valuable insights, but it is essential to address the limitations that may impede its overall generalizability. Although deep neural network (DNN) models demonstrated robust performance in handling fusion B-EGM beats involving far-field ventricular complexes, they often encountered difficulties in accurately identifying the atrial complex or its boundaries when the atrial and ventricular B-EGMs exhibited similar spectral characteristics.

The datasets inherently comprise a broad spectrum of morphological variations in both atrial and ventricular complexes from two different patient cohorts. This variability was further amplified by the utilization of DA techniques. However, it is essential to recognize that the recordings obtained from the coronary sinus exclusively represent a specific region within the heart. Consequently, undertaking a comprehensive assessment using B-EGMs from other anatomical structures, both healthy and pathologically altered, is indispensable to identify potential errors that may arise when extrapolating beyond the available data. Given the expenses associated with manual annotation, fine-tuning the hyperparameters of the DA pipeline and exploring more intricate generative techniques could prove instrumental in overcoming these limitations.

Alternative strategies for modelling long-term temporal context, such as recurrent neural networks [66], self-attention layers [67], or increasing the number of residual stages, may potentially enhance the model's performance. However, both recurrent neural networks and attention layers are computationally expensive, and the depth of the model is constrained by the Nyquist-Shannon theorem, particularly when dealing with objects of relatively short duration. Conversely, this method was found to be a source of performance degradation due to overparameterization. It should be emphasized that directly comparing these architectural blocks can be misleading, given the intricate relationship between the number of parameters, computational complexity, and information capacity. In the case of dilated kernels, the receptive field expands exponentially with the dilation factor, while the increase in the number of parameters remains linear. By simply setting a dilation factor of 1, we can evaluate its impact without making significant changes to the model's structure.

The primary limitation of the proposed method lies in its reliance on a fixed number of input channels and globaly estimated boundaries for the B-EGM complex. To address the fixed input size, an alternative approach is to incorporate multiple input heads that correspond to the required number of inputs [46]. Furthermore, to compress the spatio-temporal distribution of electrical wave propagation across multiple channels, learnable pooling functions such as RNN-based pooling [68] or self-attentive pooling [69] can be employed. When dealing with scenarios where only global labels are available, techniques like gradient-weighted class activation mapping [70] and multiple instance learning [71,72] could be used to capitalize on incomplete knowledge and highlight relevant regions and fine-grained channel-wise features for predicting the general concept.

While the dataset encompasses recordings with a diverse range of cardiac rhythm disorders and comprises a substantial number of individual subjects compared to other published studies, certain rhythm disorders do not originate from structural myocardial impairments but rather result from procedural interventions. Additionally, it is worth mentioning that the boundary annotations have not undergone cross-validation, which would aid in mitigating uncertainties related to onset and offset determination.

6.5. Clinical implications

Traditional manual assessment during EP studies is time-consuming and prone to inter-rater variability. Automated algorithms for the delineation of atrial B-EGMs and extraction of intracardiac CLs offer a promising solution to streamline the assessment process, ensuring reproducibility and reducing the reliance on human interpretation. By automatically identifying and extracting atrial B-EGMs, these algorithms can accelerate the acquisition process during mapping procedures, minimizing the number of rejected beats caused by noise or superimposed waveforms. A reliable measurement of intracardiac CLs through automated analysis can enhance real-time monitoring of the patient's rhythm during EP procedures, providing crucial early warnings or safety triggers for pacing and ablation systems. Furthermore, the automated segmentation pipeline can be effectively employed in post-procedural electrophysiology research and automated ECG waves labeling, further expanding its applications.

7. Conclusion

The method presented in this study facilitates automated extraction of fundamental inter-atrial cycle lengths from clinical data during interventional electrophysiology procedural stages including various atrial and ventricular pacing protocols, early activation mapping, and ongoing arrhythmias.

Exploratory analysis revealed exponentially weighted Tversky loss, when rebalanced to decrease the importance of easy to classify samples, significantly improved generalization capability across independent test sets. By extending a temporal context modeled by a receptive layer with widely dilated convolution kernels, the model improved its classification performance of longer sequences with atrial fibrillation. Surprisingly, it also slightly contributed to higher segmentation scores for regular atrial beats, yet with slight decrease in the delineation accuracy.

Annotating AF signals into a positive class enhanced regularization of the model and its ability to identify regular beats. However, this also resulted in a substantial decline in performance on AF rhythm when compared to model trained with inversed class assignment.

Basic noise-aware and morphology data augmentations demonstrated exceptional ability to mitigate potential overfitting, as evidenced by the virtually negligible difference in evaluation metrics between training and validation sets. However, in tasks involving analysis of F-Waves naturally exhibiting random characteristics, augmentation with Gaussian noise should be employed with caution, as it can lead to a model with heightened susceptibility to noise.

It is crucial to recognize that our proposed method exhibits certain limitations. Additional research is necessary to enable processing of an arbitrary number of input channels and accurately identify the boundaries of a single-channel electrogram. Addressing these limitations will contribute to the advancement of fully automated electrogram delineation in clinical applications.

CRediT authorship contribution statement

Jakub Hejc: Conceptualization, Investigation, Methodology,

Software, Validation, Writing – original draft, Writing – review & editing. **Richard Redina:** Data curation, Investigation, Software, Writing – original draft. **Jana Kolarova:** Resources, Supervision. **Zdenek Starek:** Resources, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Richard Redina reports financial support was provided by Brno City Municipality. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

Funding: Brno Ph.D. Talent Scholarship Holder (R.R.) funded by the Brno City Municipality.

Data and code availability

Datasets are available at request. The code used in this study is available at [73].

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bspc.2024.106274.

References

[1] D.G. Katritsis, et al., European heart rhythm association (EHRA) consensus document on the management of supraventricular arrhythmias, endorsed by heart rhythm society (HRS), asia-pacific heart rhythm society (APHRS), and sociedad latinoamericana de estimulación cardiaca y electrofisiologia (SOLAECE), EP Eur. 19 (2016) 465–511, https://doi.org/10.1093/europace/euw301.

[2] J.M. Miller, R. Jain, G. Dandamudi, S. Han, T.R. Kambur, Entrainment mapping in supraventricular tachycardias, In Cardiac Mapping 665–677 (2019), https://doi.org/10.1002/9781119152637.ch55 (John Wiley & Sons Ltd.

[3] Z.F. Issa, J.M. Miller, D.P. Zipes, Electrophysiological mechanisms of cardiac arrhythmias, In Clinical Arrhythmology and Electrophysiology 51–80 (2019), https://doi.org/10.1016/b978-0-323-52356-1.00003-7 (Elsevier.

[4] (a) C. Cantwell, et al. Techniques for automated local activation time annotation and conduction velocity estimation in cardiac mapping, Comput. Biol. Medicine 65, 229–242, DOI: 10.1016/j.compbiomed.2015.04.027 (2015). (b) D.W. Kaiser, 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, Hear. Rhythm. 13, 695–703, DOI: 10.1016/j.hrthm.2015.11.032 (2016).

[5] R.S. Damle, et al., Atrial and accessory pathway activation direction in patients with orthodromic supraventricular tachycardia: insights from vector mapping, J. Am. Coll. Cardiol. 23 (1994) 684–692, https://doi.org/10.1016/0735-1097(94)90755-2.

[6] D. Steven, J. Seiler, K.C. Roberts-Thomson, K. Inada, W.G. Stevenson, Mapping of atrial tachycardias after catheter ablation for atrial fibrillation: use of bi-atrial activation patterns to facilitate recognition of origin, Hear. Rhythm. 7 (2010) 664–672, https://doi.org/10.1016/j.hrthm.2010.01.009.

[7] P. Pascale, et al., Pattern and timing of the coronary sinus activation to guide rapid diagnosis of atrial tachycardia after atrial fibrillation ablation, Circ. Arrhythmia Electrophysiol. 6 (2013) 481–490, https://doi.org/10.1161/circep.113.000182.

[8] H. Miyazaki, W.G. Stevenson, K. Stephenson, K. Soejima, L.M. Epstein, Entrainment mapping for rapid distinction of left and right atrial tachycardias, Hear. Rhythm. 3 (2006) 516–523, https://doi.org/10.1016/j.hrthm.2006.01.014.

[9] E.G. Tolkacheva, D.G. Schaeffer, D.J. Gauthier, W. Krassowska, Condition for alternans and stability of the 1:1 response pattern in a "memory" model of paced cardiac dynamics, Phys. Rev. E 67 (2003), https://doi.org/10.1103/physreve.67.031904.

[10] D.G. Katritsis, F. Morady, Basic intervals and atrial and ventricular conduction curves, Clin. Cardiac Electrophysiol. (2022) 54–71.e1, https://doi.org/10.1016/b978-0-323-79338-4.00015-7.

[11] P.A. Santucci, N. Varma, J. Cytron, J.G. Akar, D.J. Wilber, M.O. Al Chekakie, et al., Electroanatomic mapping of postpacing intervals clarifies the complete active circuit and variants in atrial flutter, Heart Rhythm 6 (2009) 1586–1595, https://doi.org/10.1016/j.hrthm.2009.08.010.

[12] G.F. Michaud, H. Tada, S. Chough, R. Baker, K. Wasmer, C. Sticherling, et al., Differentiation of atypical atrioventricular node re-entrant tachycardia from orthodromic reciprocating tachycardia using a septal accessory pathway by the response to ventricular pacing, J. Am. Coll. Cardiol. 38 (2001) 1163–1167, https://doi.org/10.1016/s0735-1097(01)01480-2.

[13] K. Hashimoto, T. Kimura, Y. Seki, S. Ibe, T. Yamashita, H. Miyama, et al., Delineation of conduction gaps of linear lesions during atrial fibrillation ablation using ultra-high-density mapping, Europace 25 (2023), https://doi.org/10.1093/europace/euad188.

[14] M. Nothstein, et al., CVAR-seg: an automated signal segmentation pipeline for conduction velocity and amplitude restitution, Front. Physiol. 12 (2021), https://doi.org/10.3389/fphys.2021.673047.

[15] C. Schilling, M. Keller, D. Scherr, T. Oesterlein, M. Haïssaguerre, C. Schmitt, et al., Fuzzy decision tree to classify complex fractionated atrial electrograms, Biomed. Eng. / Biomed. Tech. 60 (2015), https://doi.org/10.1515/bmt-2014-0110.

[16] M. Muffoletto, A. Qureshi, A. Zeidan, L. Muizniece, X. Fu, J. Zhao, et al., Toward patient-specific prediction of ablation strategies for atrial fibrillation using deep learning, Front. Physiol. 12 (2021), https://doi.org/10.3389/fphys.2021.674106.

[17] C.D. Cantwell, C.H. Roney, F.S. Ng, J.H. Siggers, S.J. Sherwin, N.S. Peters, Techniques for automated local activation time annotation and conduction velocity estimation in cardiac mapping, Comput. Biol. Med. 65 (2015) 229–242, https://doi.org/10.1016/j.compbiomed.2015.04.027.

[18] E.F. Treo, D.O. Cervantes, E.J. Ciaccio, Automated detection and mapping of electrical activation when electrogram morphology is complex, Biomed. Signal Process. Control. 8, 41–49, DOI: 10.1016/j.bspc.2012.04.006 (2013). 13/16.

[19] C. Schilling, M.P. Nguyen, A. Luik, C. Schmitt, O. Dössel, Non-linear energy operator for the analysis of intracardial electrograms, In IFMBE Proceedings 872–875 (2009), https://doi.org/10.1007/978-3-642-03882-2_233 (Springer, Berlin Heidelberg.

[20] Y. Prudat, A. Luca, S. Yazdani, N. Derval, P. Jaïs, L. Roten, et al. Evaluation and optimization of novel extraction algorithms for the automatic detection of atrial activations recorded within the pulmonary veins during atrial fibrillation, BMC Med. Inform. Decis. Mak. 2022 Aug 28;22(1).

[21] T.G. Oesterlein, J. Schmid, S. Bauer, A. Jadidi, C. Schmitt, O. Dössel, et al., Analysis and visualization of intracardiac electrograms in diagnosis and research: concept and application of KaPAVIE, Comput. Methods Programs Biomed. 127 (2016 Apr) 165–173.

[22] A. Alcaine, et al., A wavelet-based electrogram onset delineator for automatic ventricular activation mapping, IEEE Trans. Biomed. Eng. 61 (12) (Dec. 2014) 2830–2839, https://doi.org/10.1109/TBME.2014.2330847.

[23] M.E. Haddad, et al., Algorithmic detection of the beginning and end of bipolar electrograms: implications for novel methods to assess local activation time during atrial tachycardia, Biomed. Signal Process. Control 8 (2013) 981–991, https://doi.org/10.1016/j.bspc.2012.11.005.

[24] J. Sánchez, et al., Using machine learning to characterize atrial fibrotic substrate from intracardiac signals with a hybrid in silico and in vivo dataset, Front. Physiol. 12 (2021), https://doi.org/10.3389/fphys.2021.699291.

[25] C. Li, C. Zheng, C. Tai, Detection of ECG characteristic points using wavelet transforms, IEEE Trans. Biomed. Eng. 42 (1) (1995) 21–28.

[26] Veshchezerova "Dinara, Bars C, Seitz" J. Cycle Length Estimation Using Accurate Adaptive Detection of Local Activations in Atrial Intracardiac Electrograms. Computing in Cardiology Conference (CinC) 2022. https://doi.org/10.22489/cinc.2022.142.

[27] J. Ng, V. Sehgal, J.K. Ng, D. Gordon, J.J. Goldberger, Iterative method to detect atrial activations and measure cycle length from electrograms during atrial fibrillation, IEEE Trans. Biomed. Eng. 61 (2) (Feb. 2014) 273–278, https://doi.org/10.1109/TBME.2013.2290003.

[28] H. Hajimolahoseini, J. Hashemi, S. Gazor, D. Redfearn, Inflection point analysis: a machine learning approach for extraction of IEGM active intervals during atrial fibrillation, Artif. Intell. Med. 85 (2018) 7–15, https://doi.org/10.1016/j.artmed.2018.02.003.

[29] A. Alcaine, D. Soto-Iglesias, J. Acosta, V. Korshunov, D. Penela, M. Martínez, et al., Automatic activation mapping and origin identification of idiopathic outflow tract ventricular arrhythmias, J. Electrocardiol. 51 (2018) 239–246, https://doi.org/10.1016/j.jelectrocard.2017.10.015.

[30] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P.C. Ivanov, R. Mark, J.E. Mietus, G.B. Moody, C.K. Peng, H.E. Stanley, PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals, Circulation [Online]. 101 (23), pp. e215–e220.

[31] J.M. Jenkins, R.E. Jenkins, Arrhythmia database for algorithm testing: surface leads plus intracardiac leads for validation, J. Electrocardiol. 36 (2003) 157–161, https://doi.org/10.1016/j.jelectrocard.2003.09.041.

[32] B. Abdi, R.C. Hendriks, A.-J. van der Veen, N.M.S. de Groot, Improved local activation time annotation of fractionated atrial electrograms for atrial mapping, Comput. Biol. Med. 117 (2020) 103590, https://doi.org/10.1016/j.compbiomed.2019.103590.

[33] R. Redina, et al. Arrhythmia database with annotated intracardial atrial signals from pediatric patients undergoing catheter ablation. In 2022 Computing in Cardiology (CinC), vol. 498, 1–4, DOI: 10.22489/CinC.2022.282 (2022).

[34] S. Krishnan, Y. Athavale, Trends in biomedical signal feature extraction, Biomed. Signal Process. Control 43 (2018) 41–63, https://doi.org/10.1016/j.bspc.2018.02.008.

[35] D. Pathak, R. Kashyap, S. Rahamatkar, A study of deep learning approach for the classification of electroencephalogram (EEG) brain signals, Artificial Intelligence and Machine Learning for EDGE Computing (2022) 133–144, https://doi.org/10.1016/b978-0-12-824054-0.00009-5.
[36] A. Mehrish, N. Majumder, R. Bharadwaj, R. Mihalcea, S. Poria, A review of deep learning techniques for speech processing, Information Fusion 99 (2023) 101869, https://doi.org/10.1016/j.inffus.2023.101869.
[37] Y. Tian, Y. Zhang, A comprehensive survey on regularization strategies in machine learning, Information Fusion 80 (2022) 146–166, https://doi.org/10.1016/j.inffus.2021.11.005.
[38] C. Cabo, J.M. Wharton, R.E. Ideker, W.M. Smith, "Spatial filtering of unipolar electrograms," [1990] Proceedings computers in cardiology, Chicago, IL, USA (1990) 417–420, https://doi.org/10.1109/CIC.1990.144247.
[39] K. Venkatachalam, J.E. Herbrandson, S.J. Asirvatham, Signals and signal processing for the electrophysiologist, Circ. Arrhythmia Electrophysiol. 4 (2011) 965–973, https://doi.org/10.1161/circep.111.964304.
[40] A. Halevy, P. Norvig, F. Pereira, The unreasonable effectiveness of data, IEEE Intell. Syst. 24 (2009) 8–12, https://doi.org/10.1109/MIS.2009.36.
[41] C. Sun, A. Shrivastava, S. Singh, A.K. Gupta, Revisiting unreasonable effectiveness of data in deep learning era, 2017 IEEE Int. Conf. on Comput. vis. ICCV (2017) 843–852.
[42] E.D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, L.QV. AutoAugment, Learning Augmentation Policies from Data (2018). https://doi.org/10.48550/ARXIV.1805.09501.
[43] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, In Lecture Notes in Computer Science 234–241 (2015), https://doi.org/10.1007/978-3-319-24574-4_28 (Springer International Publishing.
[44] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition. CoRR abs/1512.03385 (2015). 1512.03385.
[45] T. Vicar, J. Hejc, P. Novotna, M. Ronzhina, O. Janousek, Ecg abnormalities recognition using convolutional network with global skip connections and custom loss function, In 2020 Computing in Cardiology, 1–4, DOI: 10.22489/CinC.2020.189 (2020).
[46] T. Vicar, P. Novotna, J. Hejc, O. Janousek, M. Ronzhina, Cardiac abnormalities recognition in ecg using a convolutional network with attention and input with an adaptable number of leads, In 2021 Computing in Cardiology (CinC), vol. 48, 1–4, DOI: 10.23919/CinC53138.2021.9662806 (2021).
[47] R. Zhang, Making Convolutional Networks Shift-Invariant Again (2019), https://doi.org/10.48550/ARXIV.1904.11486.
[48] T.-Y. Lin, et al., Feature Pyramid Networks for Object Detection (2016), https://doi.org/10.48550/ARXIV.1612.03144.
[49] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, Deeplab, Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, IEEE Transactions on Pattern Analysis & Mach. Intell. 40 (2018) 834–848, https://doi.org/10.1109/TPAMI.2017.2699184.
[50] S. Gong, et al. Dilated FCN: listening longer to hear better. CoRR abs/1907.11956 (2019). 1907.11956.
[51] A. Carass, et al., Evaluating white matter lesion segmentations with refined sørensen-dice analysis, Sci. Reports 10 (2020), https://doi.org/10.1038/s41598-020-64803-w.
[52] N. Abraham, N.M. Khan, A novel focal tversky loss function with improved attention u-net for lesion segmentation. CoRR abs/1810.07842 (2018). 1810.07842.
[53] R. Zhao, et al., Rethinking dice loss for medical image segmentation, In 2020 IEEE International Conference on Data Mining (ICDM), 851–860, DOI: 10.1109/ICDM50108.2020.00094 (2020).
[54] D. Karimi, H. Dou, S.K. Warfield, A. Gholipour, Deep learning with noisy labels: exploring techniques and remedies in medical image analysis, Med. Image Anal. 65 (2020) 101759, https://doi.org/10.1016/j.media.2020.101759.
[55] M. Butkuviene, et al., Considerations on performance evaluation of atrial fibrillation detectors, IEEE Transactions on Biomed. Eng. 68 (2021) 3250–3260, https://doi.org/10.1109/TBME.2021.3067698.
[56] F. Hutter, H. Hoos, K. Leyton-Brown, An efficient approach for assessing hyperparameter importance. In Xing, E. P. & Jebara, T. (eds.) Proceedings of the 31st International Conference on Machine Learning, vol. 32 of Proceedings of Machine Learning Research, 754–762 (PMLR, Bejing, China, 2014).
[57] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization. In Bengio, Y. & LeCun, Y. (eds.) 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings (2015).
[58] I. Loshchilov, F. Hutter, Decoupled weight decay regularization. In International Conference on Learning Representations, 2019.
[59] K. He, X. Zhang, S. Ren, J. Sun, Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification, In 2015 IEEE International Conference on Computer Vision (ICCV), 1026–1034, DOI: 10.1109/ICCV.2015.123 (2015).
[60] G. Van Rossum, F.L. Drake Jr, Python reference manual (Centrum voor Wiskunde en Informatica Amsterdam, 1995).
[61] A. Paszke, et al. Pytorch: An imperative style, high-performance deep learning library. In Advances in Neural Information Processing Systems 32, 8024–8035 (Curran Associates, Inc., 2019).
[62] T. Akiba, S. Sano, T. Yanase, T. Ohta, M.O. Koyama, A next-generation hyperparameter optimization framework. In Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2019.
[63] J.H. Chen, S.M. Asch, Machine learning and prediction in medicine — beyond the peak of inflated expectations, New Engl. J. Medicine 376 (2017) 2507–2509, https://doi.org/10.1056/nejmp1702071.
[64] R. Girshick, Fast r-cnn. In 2015 IEEE International Conference on Computer Vision (ICCV), 1440–1448, DOI: 10.1109/ ICCV.2015.169 (2015).
[65] J.R.R. Uijlings, K.E.A. van de Sande, T. Gevers, A.W.M. Smeulders, Selective search for object recognition, Int. J. Comput. Vis. 104 (2013) 154–171.
[66] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (1997) 1735–1780, https://doi.org/10.1162/neco.1997.9.8. 1735.
[67] A. Vaswani, et al., Attention Is All You Need (2017), https://doi.org/10.48550/ARXIV.1706.03762.
[68] S. Li, W. Li, C. Cook, C. Zhu, Y. Gao, A fully trainable network with RNN-based pooling, Neurocomputing 338 (2019) 72–82, https://doi.org/10.1016/j.neucom.2019.02.004.
[69] F. Chen, G. Datta, S. Kundu, P. Beerel, Self-Attentive Pooling for Efficient Deep Learning 2209 (2022) 07659.
[70] R.R. Selvaraju, et al. Grad-cam: Visual explanations from deep networks via gradient-based localization. In 2017 IEEE International Conference on Computer Vision (ICCV), 618–626, DOI: 10.1109/ICCV.2017.74 (2017).
[71] J. Ramon, L. De Raedt, Multi instance neural networks, in: In Proceedings of the ICML-2000 Workshop on Attribute-Value and Relational Learning, 2000, pp. 53–60.
[72] Z.-H. Zhou, M.-L. Zhang, Neural networks for multi-instance learning, In Proceedings of the International Conference on Intelligent Information Technology, Beijing, China, 455–459 (2002).
[73] J. Hejc, R. Redina, LaTeNet Version 1.2.3 [Online]. Zenodo, Available from: https://github.com/fnusa-icrc-ice/LaTeNet, DOI: 10.5281/zenodo.10427685, 2023.