

Cross-Dataset Component Ablation of a Hybrid Model for Robust Beat-Level ECG Arrhythmia Classification

Jonas Mindaugas Rimšelis* 0009-0009-4546-4339, Viktor Medvedev* 0000-0001-5813-8545,
Povilas Treigys* 0000-0002-6608-5508, Žygimantas Abramikas† 0000-0003-3579-7913,
Jurgita Markevičiūtė‡ 0000-0002-9653-7798 and Jolita Bernatavičienė* 0000-0001-5435-8348

* Vilnius University, Faculty of Mathematics and Informatics, Institute of Data Science and Digital Technologies
Akademijos g. 4 LT-04812, Vilnius, Lithuania

Email: jonas.rimselis@mif.vu.lt

† Clinic of Cardiovascular Diseases, Institute of Clinical Medicine, Faculty of Medicine
Vilnius University, LT-03101, Vilnius, Lithuania

‡ Vilnius University, Faculty of Mathematics and Informatics
Naugarduko g. 24, LT-03225, Vilnius, Lithuania

Abstract—This study presents a systematic ablation analysis of a hybrid model for beat-level electrocardiogram (ECG) arrhythmia classification under three different intensities of domain shift, which includes cross-dataset evaluation. The proposed multi-branch architecture combines a mixture-of-experts RR encoder with waveform branches that capture local QRS morphology, global beat context, and frequency-domain information. The full model and its ablated variants were trained over five random seeds on the MIT-BIH Arrhythmia Database under the inter-patient paradigm and evaluated on both internal test subsets and an external dataset. The results indicate that the RR encoder contributes more to robustness than the waveform encoder, especially under stronger domain shift. Furthermore, RR interval-based Feature-wise Linear Modulation (FiLM) and temporal attention have a limited impact on within-dataset evaluation but become more important in cross-dataset evaluation.

I. INTRODUCTION

AN electrocardiogram (ECG) records the electrical activity of the heart and is widely used to detect arrhythmias. Single-lead Holter devices allow continuous monitoring of cardiac activity during everyday tasks and identification of cardiovascular diseases. Due to the large amount of data generated by the monitoring devices, manual physician review of ECG recordings becomes impractical, thus motivating automated analysis approaches. In recent years, deep learning methods have become a standard approach for the detection and classification of abnormal beats [1]. Hybrid models have become a common approach for beat-level ECG arrhythmia classification, by jointly utilizing rhythm and morphology features of ECG beats [2], [3], [4], [5]. To assess each component's contribution effectively, ablation experiments are performed by measuring the decrease in model performance after its removal. Although many hybrid models have been proposed, their evaluation procedures lack rigorous assessment of robustness claims. Most studies evaluate classification models using the inter-patient paradigm, whereas cross-dataset

evaluation is less common [6]. This evaluation setting provides a stricter assessment of the generalization under domain shift [7]. Furthermore, existing studies rarely combine cross-dataset evaluation with systematic ablation, leaving unclear which hybrid-model components contribute most to robustness under domain shift. Thus, this study makes the following contributions:

- We propose a systematic cross-dataset ablation framework for evaluating the contribution of rhythm, morphology, and frequency-domain components to beat-level ECG arrhythmia classification under domain shift.
- We evaluate a multi-branch hybrid architecture that combines a mixture-of-experts RR encoder with waveform branches capturing local QRS morphology, wider beat context, and frequency-domain information.
- We assess the full model and its ablated variants over five random seeds using internal inter-patient test subsets from the MIT-BIH Arrhythmia Database and an external single-lead Holter dataset collected with a ZIVE ECG device.

II. RELATED WORK

A. Background

ECG waveforms consist of the following waves: P, Q, R, S, and T, whose shape and rhythm are disrupted in abnormal cardiac activity settings such as arrhythmia [8]. Premature arrhythmic beats such as supraventricular ectopic (S), ventricular ectopic (V) can be discerned from non-ectopic beats by rhythm features, while V-class beats also have a distinctive QRS complex, as illustrated in Fig. 1.

B. Deep Learning Methods for Beat-Level Arrhythmia Classification

Beat-level arrhythmia classification typically relies on extracting features from ECG signals. 1D CNNs are used to

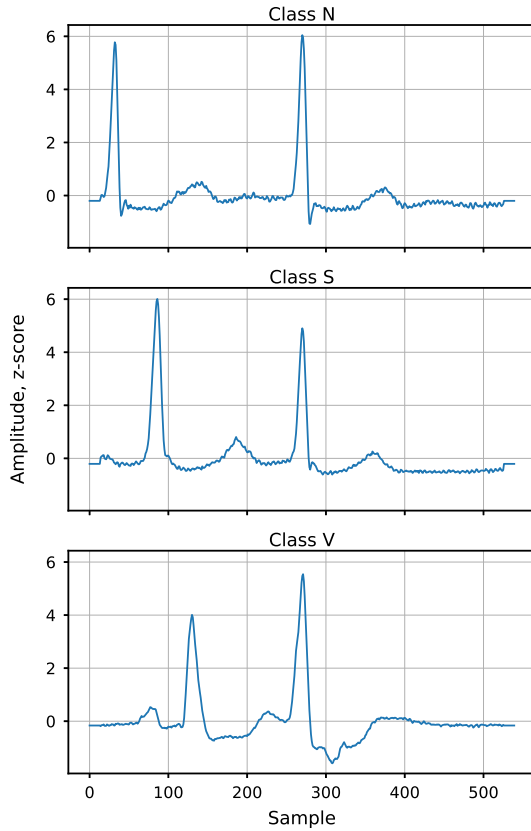


Fig. 1. Representative ECG segments containing classes N, S, and V.

extract morphology features from the waveform [9]. Additionally, frequency-domain features can be represented as a 1D signal and used as model input, which allows encoding of morphology [10]. To utilize deeper networks, residual networks are used, which also show robustness to noise [11]. Additionally, to improve classification performance for premature classes, BiLSTM and transformer modules to extract rhythm features are included [12], [13]. Thus, a combined architecture can use 1D CNNs with residual connections on both raw ECG signals and their frequency spectra to extract ECG morphological features, while BiLSTM and transformer modules capture rhythm features.

Hybrid model architectures have emerged as a leading approach because they allow a highly customizable and adaptable design whose component contributions can be evaluated through ablation studies. This can be exploited by utilizing multiscale 1D CNNs with different receptive fields to extract ECG morphological features at multiple temporal scales [2]. The input field could be controlled by passing different-length segments in separate CNN branches, allowing the model to focus on either local QRS complex information or beat-level context [3]. Additionally, temporal features extracted with the short-time Fourier transform are used to construct a time-frequency representation, which is then processed by a CNN-based branch [5]. Hybrid models also allow encoding of a sin-

gle ECG representation with different architectural branches, for instance, by passing the same input to parallel branches based on standard CNN modules and residual networks [4]. Notably, all of the cited hybrid studies also include RR features encoding with a separate branch. Thus, hybrid architectures allow the integration of multiscale morphological, contextual, rhythm, and time-frequency feature extraction within a single interpretable framework.

C. Evaluation Paradigms and Domain Shift

Evaluation of ECG arrhythmia classification on the beat level can be conducted on different levels of validity. Intra-patient evaluation can produce overly optimistic results because patient-specific patterns may be shared between the training and test sets, limiting the validity of the evaluation for previously unseen patients [1]. A stricter approach with an inter-patient paradigm provides a more realistic evaluation of model performance on unseen patients [6]. Cross-dataset evaluation assesses model generalization under domain shift arising from differences in acquisition devices, subject populations, class distributions, recording conditions, and noise characteristics between the training and external datasets [7]. Therefore, it is crucial to evaluate the model both in inter-patient and cross-dataset settings.

III. EXPERIMENT DESIGN

A. Datasets

The MIT-BIH Arrhythmia Database contains 48 two-lead ambulatory ECG recordings from 47 patients. The recordings were digitized at a 360 Hz sampling rate. For the single-lead ECG setting, only the ML II channel was used. Following recommendations of the Association for the Advancement of Medical Instrumentation (AAMI), patients (IDs: 102, 104, 107, and 217) with pacemakers were excluded because pacemaker beats introduce distinct patterns that could bias the dataset. The MIT-BIH Arrhythmia Database was split into proposed SET0, SET1, and SET2 (the composition is identical to the DS2 test set proposed by Chazal et al. [14]) subsets, which comply with the inter-patient paradigm. SET0 subset contains 17 patients (IDs: 101, 106, 108, 109, 112, 118, 115, 201, 119, 230, 124, 203, 208, 209, 215, 207, 223) and is used for training, while SET1 contains 5 patients (IDs: 116, 114, 205, 220, 122), and SET2 contains 22 patients (IDs: 100, 103, 105, 111, 113, 117, 121, 123, 200, 202, 210, 212, 213, 214, 219, 221, 222, 228, 231, 232, 233, 234) for testing.

Using the ZIVE ECG single-lead Holter monitoring device, cardiac activity was recorded for 103 patients with a 200 Hz sampling rate, resulting in 1092 10-minute recordings. Beat-level arrhythmia annotations were independently provided by two clinicians. Thus, the evaluation strategy covers three levels of generalization difficulty: SET1 as a mild internal domain shift, SET2 as a stronger internal domain shift, and SET3 as an external cross-dataset domain shift.

Datasets contain five superclasses based on the AAMI recommendations - Normal (N), Supraventricular Ectopic (S), Ventricular Ectopic (V), Fusions (F), and Unclassified (Q).

Due to low prevalence, F and Q beats are excluded from the classification task but retained in the data as a beat context. In this study, the original annotations N, L, R, e, and j were mapped to class N; A, a, J, and S to class S; and V and E to class V. The summary statistics of beats are presented in Table I.

TABLE I
CLASS DISTRIBUTION ACROSS DATASETS

	N	S	V
SET0	34,743 (88.8%)	834 (2.1%)	3,565 (9.1%)
SET1	11,103 (97.1%)	110 (1.0%)	223 (1.9%)
SET2	44,177 (89.7%)	1,834 (3.7%)	3,218 (6.5%)
SET3	732,852 (96.8%)	6,536 (0.9%)	17,646 (2.3%)

B. Preprocessing

Raw ECG recordings were resampled to 257 Hz to ensure a common sampling frequency and consistent beat-segment representations across datasets. Resampled signals were filtered with a fourth-order Butterworth filter with cutoff frequencies of 0.5 Hz and 40 Hz to remove baseline wander and high-frequency noise.

C. RR Features

RR information was calculated in two stages. For beat i , 14 base rhythm measurements were first extracted. Three RR intervals were treated as augmentable measurements: the preceding RR interval ($R_i - R_{i-1}$), the current RR interval ($R_{i+1} - R_i$), and the following RR interval ($R_{i+2} - R_{i+1}$). During training, these three measurements were augmented using multiplicative scaling and additive Gaussian noise.

The remaining 11 base measurements were not directly augmented and represented short-term (5 beats), short-term (50 beats), long-term (150 beats), and recording-level rhythm context. They comprised the local RR mean and standard deviation, short-term and long-term RR means, long-term RR standard deviation, recording-level mean RR, detrended preceding, current, and following RR intervals, and mean RR intervals in the preceding and following local contexts.

The 14 base measurements were subsequently transformed into a 35-dimensional RR feature vector containing raw, normalized, standardized, detrended, ratio-based, variability-based, and contextual rhythm descriptors. The resulting feature vectors were processed by the RR encoder.

D. Segmentation

For each beat, two R-peak-centered segments of 0.4 s (short) and 0.8 s (medium) were extracted at a sampling rate of 257 Hz. Each segment was normalized using segment-level z-score standardization.

E. Frequency Feature Extraction

A fast Fourier transform was applied to the medium segment to extract the magnitude spectrum, which was used as model input.

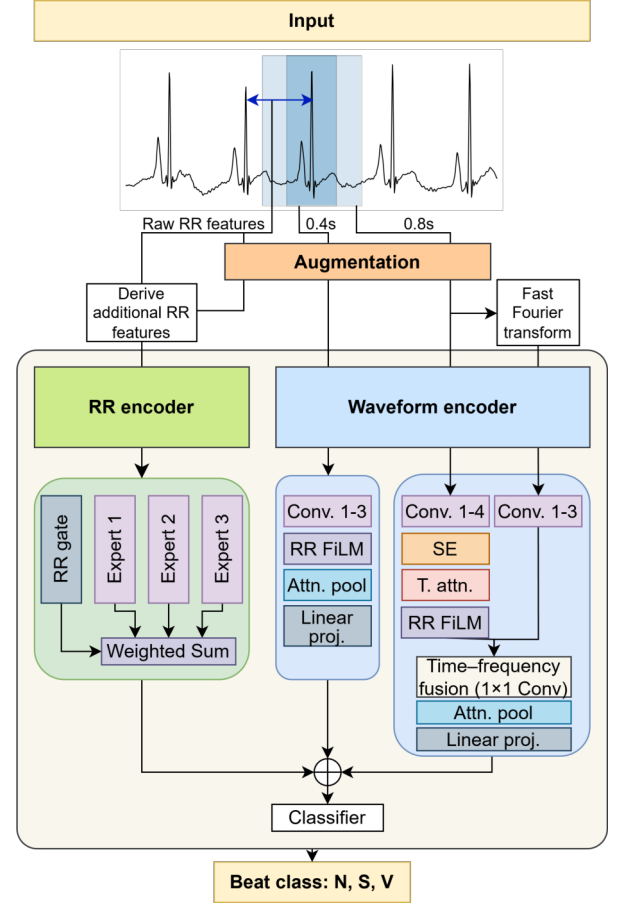


Fig. 2. Architecture of the proposed ECG beat-level arrhythmia classification model.

F. Model Architecture

The model architecture follows a multi-branch design to capture different types of features – morphology and rhythm with a waveform encoder and RR feature encoder, respectively. The waveform encoder contains separate subbranches for encoding short and medium segments. Architectural hyperparameters were selected based on pilot experiments on the training data to capture the ECG beat features for each arrhythmia class, and reduce the risk of overfitting.

1) *Waveform Encoder*: The short-segment encoder consisted of three sequential 1D convolution blocks. Each block used 96 output channels, kernel size 5, stride 1, dilation 1, and padding 2. Each convolution was followed by group normalization, LeakyReLU activation, and dropout. After feature extraction, the feature maps were modulated by RR interval features based on feature-wise linear modulation (FiLM). Attention pooling was then applied using a 1×1 convolution to weight features temporally.

The medium segment branch consists of temporal and frequency subbranches. The temporal subbranch consisted of four sequential 1D convolution blocks, followed by squeeze-and-excitation (SE) channel recalibration, temporal attention,

RR interval features based FiLM, and attention pooling. The frequency subbranch consisted of three 1D convolution blocks followed by attention pooling. Subbranch representations were concatenated and fused using a 1×1 convolution.

2) *RR Encoder*: The RR encoder was implemented as a mixture-of-experts module with three expert multilayer perceptrons and a learned gating network. Each RR expert consisted of two fully connected blocks, which consisted of the projection of concatenated RR input to 128 and 64 dimensions in the first and second block, respectively, followed by layer normalization, SiLU activation, and dropout. A final linear layer projected to the shared RR embedding dimension. A gating network mapped the same RR input to expert weights through a 64-dimensional hidden layer with layer normalization, SiLU activation, and dropout, followed by a linear layer and softmax. The final RR representation was obtained as the weighted sum of the three expert outputs, followed by layer normalization.

3) *Classifier*: The waveform, both short and medium segments, and RR embeddings were linearly projected to have identical dimensions and fused by simple concatenation and then passed to the classifier. The classifier consisted of two multilayer perceptron blocks followed by a final output layer. The first block projected the fused feature vector to 256 dimensions, followed by layer normalization, LeakyReLU activation, and dropout. The second block projected the features to 128 dimensions, followed by layer normalization, LeakyReLU activation, and dropout. A final linear layer mapped the 128-dimensional representation to the three output classes.

G. Augmentation

For waveform augmentation, each ECG segment was augmented by adding Gaussian noise with standard deviation 0.003 in standardized signal units, applying global amplitude scaling by a factor sampled uniformly from 0.95 to 1.05, and applying a random temporal shift of between -3 and 3 samples. For RR augmentation, only the raw augmentable features were modified using global amplitude scaling by a factor sampled uniformly from 0.97 to 1.03, and with additive Gaussian noise with standard deviation 0.01. Transformations were applied online during training to improve robustness to morphological, rhythm, and alignment variations.

H. Training

Model training was performed using five fixed random seeds (seeds = 2, 5, 7, 8, 9) for Python, NumPy, and PyTorch, while deterministic cuDNN execution was enabled to improve reproducibility. Training was conducted using the Adam optimizer with weight decay of 0.01. The loss function was focal loss with weights: N:1, S:3, V:1. Learning-rate adaptation was performed using the OneCycleLR scheduler for 15 epochs with a learning rate of 0.0005.

I. Systematic Ablation Experiment

The ablation experiment was conducted by training the full model and its ablated variants on the SET0 training set and

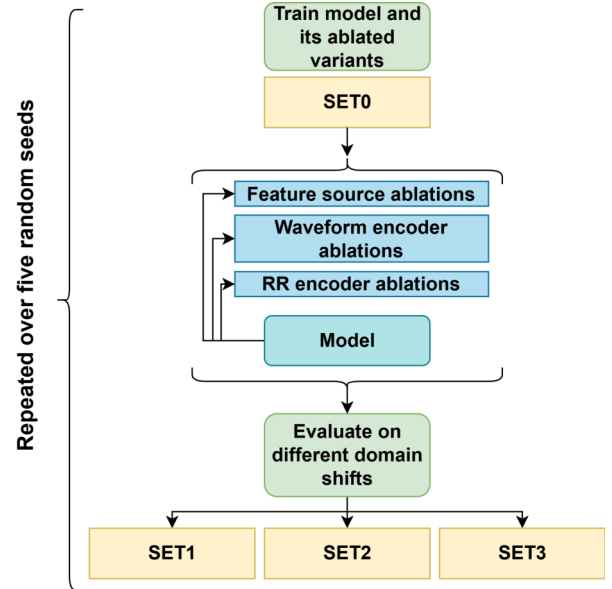


Fig. 3. Proposed ablation experiment.

evaluating them on the SET1, SET2, and SET3 datasets (Fig. 3). Each ablation experiment was repeated five times using different random seeds. The same set of seeds was used across all ablation settings. The ablations were divided into three groups (Table II):

- **Feature source ablations** – major ablations assessing the contribution of different feature types.
- **Waveform encoder structural ablations** – module and subbranch level ablations within the waveform encoder.
- **RR encoder structural ablations** – comparisons between different RR expert configurations.

TABLE II
ABLATION VARIANTS.

Ablation setting	Disabled component(s)
Feature source ablations	
Full model	None
No RR encoder	RR branch, RR FiLM
No waveform encoder	All waveform branches
Waveform encoder ablations	
Full model	None
No SE block	Mid-scale branch SE block
No temporal attention	Mid-scale branch temporal attention
No RR FiLM	RR FiLM
No short-scale branch	Short-scale waveform branch
No mid-scale branch	Mid-scale waveform branch and its frequency subbranch
No frequency subbranch	Mid-scale frequency subbranch
RR encoder ablations	
Three RR experts	Mixture-of-experts RR encoder with three experts
Single RR expert	Single-expert RR encoder

J. Evaluation

To evaluate model performance on test sets, class-level Precision, Recall, and F1 score are used. For aggregated

TABLE III

PERFORMANCE COMPARISON OF THE PROPOSED METHOD AND RECENTLY PUBLISHED METHODS FOR THREE-CLASS N, S, AND V CLASSIFICATION ON THE MIT-BIH ARRHYTHMIA DATABASE SET2 (DS2). FOR THE PROPOSED METHOD, RESULTS ARE REPORTED FOR THE BEST-PERFORMING MODEL AMONG FIVE MODELS TRAINED WITH DIFFERENT RANDOM SEEDS. BOLD VALUES INDICATE THE BEST RESULT FOR EACH CLASS AND METRIC.

Method source	Metrics	N	S	V
The proposed method	Precision (%)	99.20	76.80	87.86
	Recall (%)	98.29	84.30	94.00
	F1 score (%)	98.74	80.40	90.83
Z. Qiu et al. (2025) [4]	Precision (%)	98.21	87.34	96.73
	Recall (%)	99.43	83.69	83.89
	F1 score (%)	98.42	85.51	89.82
Z. Ran et al. (2024) [15]	Precision (%)	99.29	36.48	68.91
	Recall (%)	91.24	82.88	95.62
	F1 score (%)	95.09	50.66	80.10
M. Farag (2023) [16]	Precision (%)	99.00	82.68	95.63
	Recall (%)	99.10	81.60	95.00
	F1 score (%)	99.05	82.14	95.31

evaluation, the macro F1 score is reported [4].

IV. RESULTS

A. Comparison of the Proposed Model with Existing Methods

Compared with recently published studies on beat-level arrhythmia classification in the three-class setting (Table III), the proposed model achieves competitive performance on the SET2 (MIT-BIH Arrhythmia Database DS2 [14]) under the inter-patient paradigm. It reaches second-best performance for the classification of V beats and third-best performance for the classification of S beats.

B. Feature Source Ablations

TABLE IV

MACRO F1 SCORE FOR THE FEATURE SOURCE ENCODER ABLATIONS ON THE SET1, SET2 AND SET3. VALUES ARE REPORTED AS MEAN $\pm$ STANDARD DEVIATION.

Ablation setting	Macro F1 score, %		
	SET1	SET2	SET3
Full model	93.09 $\pm$ 2.79	89.18 $\pm$ 1.09	80.09 $\pm$ 2.44
No RR encoder	76.36 $\pm$ 5.53	58.63 $\pm$ 1.81	42.88 $\pm$ 5.02
No waveform encoder	82.00 $\pm$ 4.02	63.05 $\pm$ 3.16	68.96 $\pm$ 0.97

Table IV shows a decrease in performance when either the RR or waveform encoder is ablated from the model across all datasets. For the SET1 and SET2, the decline in performance observed for no RR encoder and no waveform encoder settings is of comparable magnitude. In contrast, on the SET3, ablation of the RR encoder resulted in a moderately larger decrease in model performance ($42.88 \pm 5.02\%$) than removing the waveform encoder ($68.96 \pm 0.97\%$).

Analysis of S-class F1 scores indicates that the feature-source ablations differ in impact magnitude: the RR encoder consistently has a larger effect on performance than the waveform encoder across all datasets (Fig. 4). Furthermore, the RR encoder is critical for classifying S beats, as its ablation

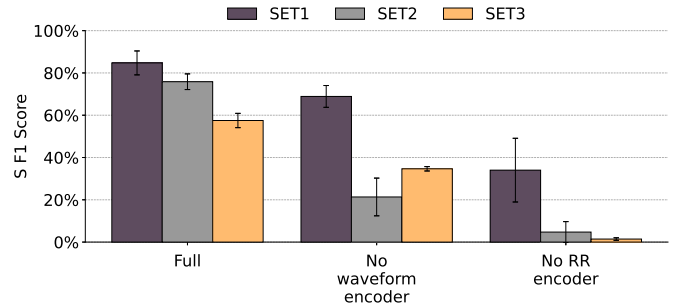


Fig. 4. S-class F1 scores for the full model and feature source ablations across the SET1, SET2, and SET3 datasets. Error bars indicate standard deviation.

on SET2 and SET3 yields minimal F1 scores of $4.74 \pm 4.95\%$ and $1.43 \pm 0.61\%$, respectively.

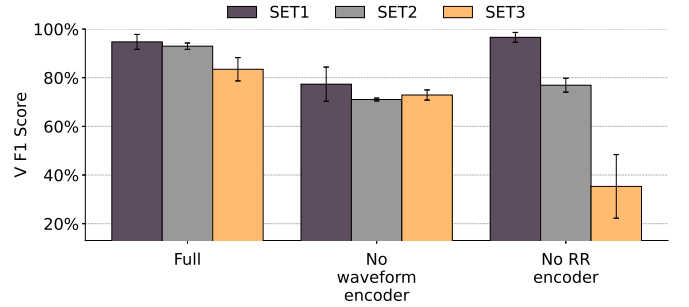


Fig. 5. V-class F1 scores for the full model and feature source ablations across the SET1, SET2, and SET3 datasets. Error bars indicate standard deviation.

In contrast, analysis of F1 scores for V-class (Fig. 5) indicates that ablations have distinct effects across datasets. For the SET1, removal of the RR encoder yields performance comparable to that of the full model, whereas the waveform encoder removal leads to a decrease in performance. On the SET2, both ablations lead to a decrease in performance, but the waveform encoder has a slightly larger impact. On the SET3, removal of the RR encoder leads to a substantial decline in performance.

Overall, the RR encoder has a greater impact than the waveform encoder on S-class classification across all the datasets, while for V-class beats, this pattern is observed only for the SET3.

C. Waveform Encoder Structural Ablations

Results of waveform structural ablations are presented in Table V and suggest that excluding modules from the encoder has only a limited effect on the performance of the model for SET1 and SET3. Conversely, on SET2, removing individual encoder modules generally had a greater effect on performance and increased variability across random seeds. This suggests that the ablated models were more sensitive to initialization under the more heterogeneous and challenging inter-patient evaluation setting, although dataset-specific characteristics may also contribute to the observed variability.

TABLE V
MACRO F1 SCORE FOR THE WAVEFORM ENCODER ABLATIONS ON THE SET1, SET2 AND SET3. VALUES ARE REPORTED AS MEAN $\pm$ STANDARD DEVIATION.

Ablation setting	Macro F1 score, %		
	SET1	SET2	SET3
Full model	93.09 $\pm$ 2.79	89.18 $\pm$ 1.09	80.09 $\pm$ 2.44
No RR FiLM	92.79 $\pm$ 3.03	80.85 $\pm$ 9.43	75.29 $\pm$ 4.11
No short-scale branch	89.92 $\pm$ 1.86	68.36 $\pm$ 1.96	76.99 $\pm$ 2.31
No mid-scale branch	91.91 $\pm$ 1.90	81.28 $\pm$ 8.05	78.75 $\pm$ 3.31
No SE	92.76 $\pm$ 1.89	86.76 $\pm$ 7.82	76.32 $\pm$ 5.08
No temporal attention	92.42 $\pm$ 0.62	78.97 $\pm$ 9.71	74.37 $\pm$ 6.19
No frequency subbranch	90.03 $\pm$ 4.82	86.01 $\pm$ 3.96	77.37 $\pm$ 3.93

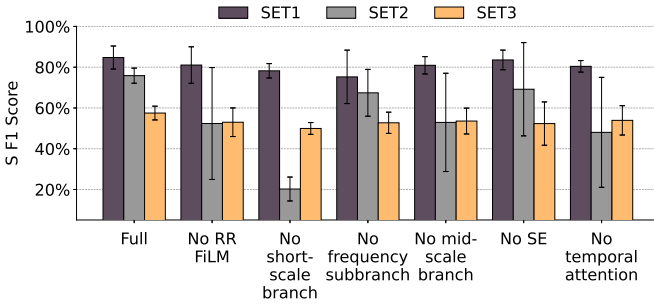


Fig. 6. S-class F1 scores for the full model and waveform encoder ablations across the SET1, SET2, and SET3. Error bars indicate standard deviation.

Analysis of S-class F1 scores shows that, for SET1 and SET3, ablations of waveform encoder modules have no major effect on performance (Fig. 6). In contrast, for the SET2, ablations of almost all modules, except for the short branch, reduce F1 scores and increase their variability, suggesting that these modules contribute to model stability. Additionally, ablating the short-scale branch substantially reduced the F1 score relative to the full model, suggesting that local waveform morphology, particularly QRS-related features, contributes to distinguishing class S.

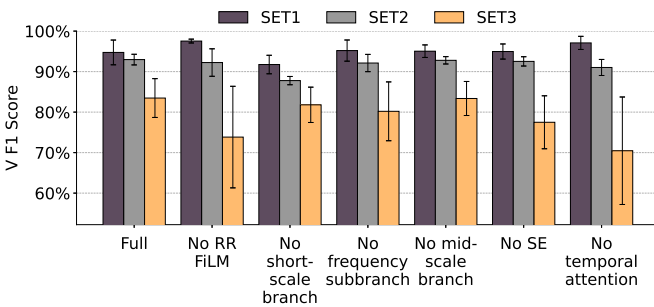


Fig. 7. V-class F1 scores for the full model and waveform encoder ablations across the SET1, SET2, and SET3. Error bars indicate standard deviation.

The F1 scores for V beats show an interesting pattern (Fig. 7). Although ablations do not lead to a substantial decrease in performance on the SET1 and SET2, RR FiLM and temporal attention appear to contribute moderately to SET3. Ablation of these modules leads not only to lower average V-class F1

scores but also to an increased variability, suggesting reduced model stability.

D. RR Encoder Structural Ablations

Table VI indicates that a mixture of three experts, the RR encoder, performs slightly better than a single expert variant across all datasets.

TABLE VI
MACRO F1 SCORE FOR THE RR ENCODER ABLATIONS ON THE SET1, SET2 AND SET3. VALUES ARE REPORTED AS MEAN $\pm$ STANDARD DEVIATION.

Ablation setting	Macro F1 score, %		
	SET1	SET2	SET3
Full model	93.09 $\pm$ 2.79	89.18 $\pm$ 1.09	80.09 $\pm$ 2.44
Single RR expert	90.73 $\pm$ 2.36	84.70 $\pm$ 5.89	73.86 $\pm$ 2.50

Analysis of S-class classification performance indicates that using only a single RR expert consistently decreases the mean F1 score and, for the SET2, increases result variability (Fig. 8).

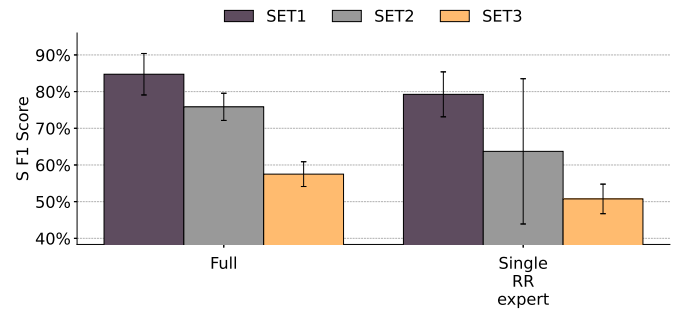


Fig. 8. S-class F1 scores for the full model and RR encoder ablations across the SET1, SET2, and SET3. Error bars indicate standard deviation.

Investigation of V-class classification performance suggests that the number of RR experts has little effect on SET1 and SET2 but moderately decreases the mean F1 score on SET3 (Fig. 9).

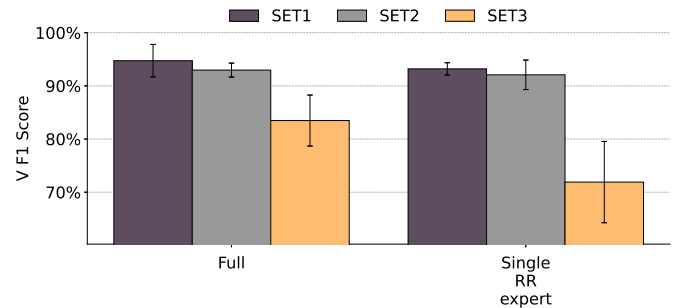


Fig. 9. V-class F1 scores for the full model and RR encoder ablations across the SET1, SET2, and SET3. Error bars indicate standard deviation.

V. CONCLUSION

This study investigated the robustness of the proposed hybrid model for beat-level ECG classification under inter-patient and cross-dataset domain shift through a systematic ablation framework. The results showed that RR features capturing rhythm information across multiple temporal contexts were critical for maintaining the performance of the proposed model in both the MIT-BIH Arrhythmia Database inter-patient evaluation and the cross-dataset evaluation, particularly for the S and V classes. The removal of RR FiLM, mid-branch, SE block, and temporal attention modules from the waveform encoder leads to higher variability of S-class classification results, indicating reduced model stability. The waveform encoder ablation results indicate that features extracted by individual components may provide limited discriminative value on one evaluation subset, such as SET2, yet become important under domain shift caused by differences in recording devices and patient populations, as observed for SET3.

To conclude, model evaluation on the inter-patient paradigm using a single test set may not fully reveal the behavior of model components under cross-dataset domain shift. Therefore, cross-dataset systematic ablation provides a more informative framework to assess how individual model components contribute to robustness.

This study is limited to a single proposed model architecture, a three-class classification setting, evaluation over only five random seeds, and one external dataset. Future work should concentrate on extending this framework to additional external datasets, alternative hybrid model architectures, and more comprehensive statistical comparison of ablation effects under different forms of domain shift.

ACKNOWLEDGMENT

The authors are thankful for the high performance computing resources provided by the Information Technology Research Center of Vilnius University. This project has received funding from the Research Council of Lithuania (LMTLT), agreement No. S-ITP-25-9.

During the preparation of this work, the author(s) used ChatGPT for code assistance and editorial corrections, and ScopusAI for literature exploration. The authors take full responsibility for the content of this publication.

REFERENCES

- [1] Q. Xiao, K. Lee, S. A. Mokhtar, I. Ismail, A. L. b. M. Pauzi, Q. Zhang, and P. Y. Lim, "Deep learning-based ECG arrhythmia classification: A systematic review," *Applied Sciences*, vol. 13, no. 8, 2023. doi: 10.3390/app13084964
- [2] F. yan Zhou, Y. hao Sun, and Y. wen Wang, "Inter-patient ECG arrhythmia heartbeat classification network based on multiscale convolution and FCBA," *Biomedical Signal Processing and Control*, vol. 90, p. 105789, 2024. doi: 10.1016/j.bspc.2023.105789. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1746809423012223>
- [3] J. Cai, J. Song, and B. Peng, "Enhancing ECG heartbeat classification with feature fusion neural networks and dynamic minority-biased batch weighting loss function," *Physiological Measurement*, vol. 45, no. 7, p. 075002, jul 2024. doi: 10.1088/1361-6579/ad5cc0
- [4] Z. Qiu, S. Tang, H. Liu, X. Zhao, and J. Lin, "Cdlr-net: a ecg classification network based on deep residual shrinkage networks and lstm," *Computer methods in biomechanics and biomedical engineering*, pp. 1–15, 09 2025. doi: 10.1080/10255842.2025.2554260
- [5] F. Zhou and J. Wang, "Heartbeat classification method combining multi-branch convolutional neural networks and transformer," *iScience*, vol. 27, no. 3, p. 109307, 2024. doi: 10.1016/j.isci.2024.109307
- [6] H. Di and X. Li, "Arrhythmia detection algorithm based on domain knowledge and spatio-temporal network model," in *2025 IEEE 5th International Conference on Electronic Technology, Communication and Information, ICETCI 2025*, 2025. doi: 10.1109/ICETCI64844.2025.11084081 p. 252 – 255.
- [7] R. Li, Y. Aierken, Y. Xu, J. Liu, and Y. Tang, "Research on cross-dataset cardiac signal domain generalization and feature interpretability," *Scientific Reports*, pp. 1–23, 2026. doi: 10.1038/s41598-025-33057-9
- [8] Z. Kotsialou, N. Makris, and S. Gall, "Fundamentals of the electrocardiogram and common cardiac arrhythmias," *Anaesthesia & Intensive Care Medicine*, vol. 25, no. 3, pp. 219–222, 2024. doi: 10.1016/j.mpaic.2023.11.014
- [9] H. Wang, H. Shi, X. Chen, L. Zhao, Y. Huang, and C. Liu, "An Improved Convolutional Neural Network Based Approach for Automated Heartbeat Classification," *Journal of medical systems*, vol. 44, no. 2, p. 35, 2019. doi: 10.1007/s10916-019-1511-2
- [10] T. Lan, Q. Hu, X. Liu, K. He, and C. Yang, "Arrhythmias classification using short-time fourier transform and GAN based data augmentation," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 2020-July, 2020, Conference paper. doi: 10.1109/EMBC44109.2020.9176733 p. 308 – 311.
- [11] J. Zang, J. Wang, Z. Zhang, Y. Zheng, and C. Xue, "Improving ECG classification performance by using an optimized one-dimensional residual network model," *Applied Sciences*, vol. 12, no. 24, 2022. doi: 10.3390/app122412957
- [12] A. J. Prakash and M. Atef, "A lightweight deep learning approach for patient-specific electrocardiogram beat classification using local and long-term dependencies," *Engineering Applications of Artificial Intelligence*, vol. 152, no. April, 2025. doi: 10.1016/j.engappai.2025.110754
- [13] S. Chon, K.-W. Ha, S. Park, and S. Jung, "An ECG beat classification method using multi-kernel resnet with transformer," in *Proceedings - 2023 IEEE International Conference on Big Data and Smart Computing, BigComp 2023*, 2023. doi: 10.1109/BigComp57234.2023.00031 p. 140 – 144.
- [14] P. de Chazal, M. O'Dwyer, and R. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 7, pp. 1196–1206, 2004. doi: 10.1109/TBME.2004.827359
- [15] Z. Ran, M. Jiang, Y. Li, Z. Wang, Y. Wu, and W. Ke, "Arrhythmia classification based on multi-feature multi-path parallel deep convolutional neural networks and improved focal loss," vol. 21, no. March, pp. 5521–5535, 2024. doi: 10.3934/mbe.2024243
- [16] M. M. Farag, "A tiny matched filter-based CNN for inter-patient ecg classification and arrhythmia detection at the edge," *Sensors*, vol. 23, no. 3, 2023. doi: 10.3390/s23031365