Brugada Detection from 12-Lead ECG Using a Multi-Scale Hybrid Transformer with Explainability

Junmo An[1], Ziyu Li[2], Dillon J Dzikowicz[1], Ben Bailey[1], Michael Silka[3]

[1]Philips, Cambridge, Massachusetts, USA
[2]Philips, Eindhoven, Netherlands
[3]Children's Hospital Los Angeles, Los Angeles, California, USA

Abstract

Brugada syndrome is associated with ventricular arrhythmias and sudden cardiac death, requiring reliable detection from 12-lead ECGs. We developed an end-to-end Multi-Scale Hybrid Transformer to classify Brugada-labeled versus control ECGs directly from raw waveforms and assessed its interpretability using gradient-weighted class activation mapping (Grad-CAM). The publicly available Brugada-HUCA database comprised 69 confirmed Brugada, 7 other/atypical, and 287 control ECG. Original 12-second recordings were cropped to 10-second segments and upsampled from 100 to 500 Hz. Randomized augmentation included amplitude scaling, baseline wander, Gaussian noise, and random point dropout. Performance was evaluated using a 10-loop rotating framework with separate training, validation, and held-out test folds. Across the test folds, the model achieved a mean sensitivity of 0.671, specificity of 0.918, F1-score of 0.663, AUROC of 0.867, and accuracy of 0.871. Beat-averaged Grad-CAM highlighted clinically relevant ST-segment morphology, particularly in leads V1-V2. These findings support proof-of-concept discrimination of Brugada-labeled/atypical ECGs from controls; however, the observed sensitivity is not adequate for clinical screening, and external validation is required.

1. Introduction

Brugada syndrome is an inherited arrhythmogenic disorder associated with polymorphic ventricular tachycardia, ventricular fibrillation, and sudden cardiac death in individuals without structural heart disease [1]. The diagnostic type 1 Brugada ECG pattern consists of coved ST-segment elevation ≥0.2 mV in at least one right precordial lead (V1 or V2, recorded in standard or high intercostal positions), followed by a negative T wave. However, Brugada manifestations can be dynamic and may vary with fever, sodium-channel-blocking drug exposure, autonomic state, electrode placement, and other physiological conditions. Consequently, a diagnostic pattern may not be evident on every routine ECG.

Recognition of Brugada patterns can also be complicated by right bundle branch block, early repolarization, electrode-position variability, and Brugada phenocopies. Although established diagnostic criteria provide a structured framework, interpretation can remain challenging when abnormalities are subtle or incomplete.

Traditional rule-based computerized ECG interpretation systems generally rely on explicitly programmed criteria and predefined measurements, such as J-point amplitude, ST-segment elevation and slope, T-wave polarity, and QRS morphology. A complementary quantitative study showed that automated measurements of ST-segment elevation, slope, and QRS and T-wave areas in leads V1-V2 differed significantly in Brugada syndrome [2]. However, these approaches depend on accurate waveform delineation and expert-defined features, making them vulnerable to fiducial-point and segmentation errors. Restricting analysis to a representative beat or single lead may also overlook temporal variability and complementary information across the full 12-lead ECG.

End-to-end deep learning (DL) learns discriminative features directly from ECGs. Convolutional neural networks (CNNs) capture local morphology, whereas Transformers model long-range temporal dependencies and interactions among signal regions [3]. Hybrid CNN-Transformer architectures combine these complementary capabilities [4], [5] and may be suitable for Brugada detection, which involves localized right-precordial patterns and broader beat- and lead-level context.

A limitation of DL is that high predictive performance does not identify which ECG findings drive model decisions, a key concern in rare disorders where errors may have serious consequences. Previous ECG studies used SHapley Additive exPlanations (SHAP) to identify influential waveform regions or features [6], [7]; here, we used beat-averaged Gradient-weighted Class Activation Mapping (Grad-CAM) to provide complementary, class-specific localization of salient ECG regions [8].

This study aimed to: 1) develop an end-to-end Multi-Scale Hybrid Transformer for Brugada detection from raw

10-second, 12-lead ECGs; 2) enhance robustness using randomized augmentation; 3) evaluate performance using 10-fold cross-validation; and 4) generate beat-averaged Grad-CAM visualizations to assess whether model attribution aligns with clinically relevant Brugada morphology.

2. Methods

2.1. Dataset and Preprocessing

The Brugada-HUCA database, publicly available through PhysioNet, was used in this study [9]. It contains one anonymized standard 12-lead ECG from each of 363 subjects evaluated at the Hospital Universitario Central de Asturias. Public metadata include 69 confirmed Brugada cases (code 1), 7 atypical or other cases (code 2), and 287 controls (code 0).

Each ECG includes leads I, II, III, aVR, aVL, aVF, and V1-V6 and was recorded for 12 seconds at 100 Hz, with amplitudes in millivolts. Labels were assigned by expert electrophysiologists using accepted Brugada criteria, with consensus adjudication for ambiguous cases. Recordings with substantial noise, incomplete leads, or insufficient diagnostic information were excluded. In MATLAB, recordings were cropped to 10 seconds and upsampled to 500 Hz using finite impulse response (FIR)-based band-limited interpolation, yielding inputs of 12 leads x 5,000 samples. Lead order was standardized, and no handcrafted features, representative beats, or diagnostic criteria were provided to the model.

2.2. Data Augmentation and Automation

To improve model robustness and generalization, four signal augmentation techniques, including amplitude scaling, baseline wander, Gaussian noise, and random point dropout, were incorporated into a modified RandAugment method [6], [7], [10], [11]. During each training iteration, one technique was selected with equal probability of 0.25 and applied to the input ECG. Scaling uniformly adjusted amplitudes across all leads while preserving relative morphology between leads. Baseline wander added a randomized low frequency sinusoidal component to simulate respiratory or electrode related drift, whereas Gaussian noise introduced random acquisition noise with zero mean and variable magnitude. Random point dropout set selected signal samples to zero to simulate transient signal loss. Augmentation types and parameters were randomized across epochs to increase exposure to diverse conditions during repeated model optimization steps and applied only during training, while validation and test ECGs were evaluated without augmentation. Figure 1 illustrates the effect of each technique on representative ECG waveforms in lead V1.

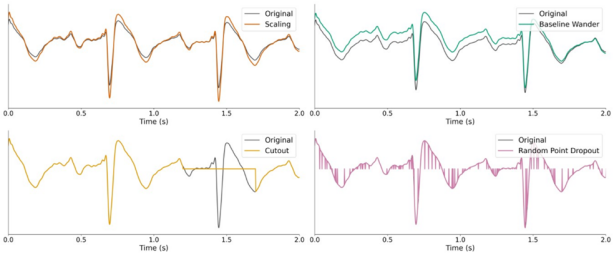

Figure 1. Data augmentation techniques incorporating amplitude scaling, baseline wander, and random point dropout, illustrated in lead V1.

2.3. Proposed Model Architecture

An end-to-end Multi-Scale Hybrid (MS-Hybrid) Transformer was developed to classify ECGs in the positive versus control dataset classes directly from raw 12-lead signals. The model first processed the ECG, a 10-second signal at 500 Hz (5,000 samples), using two parallel one-dimensional convolutional branches. The wide branch used a kernel size of 31 to capture broader waveform morphology, whereas the narrow branch used a kernel size of 7 to capture localized signal characteristics. Each branch generated 32 feature maps using a stride of 2, followed by batch normalization and rectified linear unit activation. The outputs were concatenated into 64 feature maps and downsampled using max pooling.

A convolutional reducer subsequently transformed the feature maps into 128-dimensional temporal embeddings while further reducing the sequence length. It consisted of a convolution with a kernel size of 15 and stride of 4, followed by group normalization and Gaussian error linear unit activation. A depthwise convolution with a kernel size of 3 refined channel-specific temporal features. For additional downsampling, a final convolution with a kernel size of 7 and stride of 2 was applied. For the 5,000-sample input, this process generated exactly 157 temporal tokens. Learnable positional embeddings were added to the tokens and are linearly interpolated when necessary to support variable input lengths.

The token sequence was processed by six pre-normalized Transformer encoder layers (indicated by 'norm_first=True' in the code) with an embedding dimension of 128, eight attention heads, a feed-forward dimension of 256, GELU activation, and a dropout rate of 0.3. Unlike architectures using a dedicated classification token, the proposed model applied attention pooling, which learned a normalized importance weight for each temporal token and generated a 128-dimensional weighted global representation. The pooled features were passed through an MLP head comprising layer normalization, a 128-to-64 linear layer, GELU activation, dropout (0.3), and a final linear layer for binary classification. The model produced a single logit representing the probability of the positive dataset class. To facilitate the entire development

lifecycle from data preprocessing and architectural design to model training, evaluation, and pipeline orchestration, we utilized a robust ecosystem of Python libraries including PyTorch, scikit-learn, SciPy, and the Waveform Database toolkit.

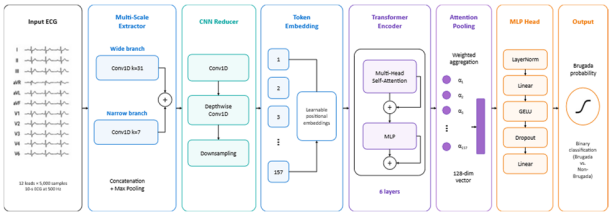

Figure 2. The proposed architecture of the end-to-end Multi-Scale Hybrid Transformer model.

2.4. Training and Performance Evaluation

Model performance was assessed using a 10-loop rotating fold evaluation, with six folds for training, two for validation, and two for testing in each loop. Data augmentation was applied only to the training set. The model was trained for up to 150 epochs with a batch size of 64 using AdamW, an initial learning rate of 0.00005, weight decay of 0.05, and a cosine-annealing warm-restart scheduler. Class imbalance was addressed using binary cross-entropy with logits and a positive-class weight based on the control-to-positive class ratio in each training set. Early stopping was applied with a patience of 15 epochs.

The optimal probability threshold was selected on the validation set from candidate cutoffs between 0.30 and 0.80 by maximizing the F1-score and was applied unchanged to the corresponding test set. Performance across the 10 loops was summarized using sensitivity, specificity, positive and negative predictive values, F1-score, AUROC, AUPRC, and accuracy, reported as mean ± standard deviation. Experiments were conducted on a workstation equipped with an Intel® Core™ i7-7700 CPU, 64 GB of RAM, and an NVIDIA GeForce RTX 3080 Ti graphics processing unit (GPU).

2.5. Explainability

Grad-CAM was used to generate class-specific, temporally localized attribution maps for ECG-based predictions. Grad-CAM was applied to the final Conv1D layer of the CNN reducer, immediately before token embedding and Transformer encoding. Gradients of the Brugada-class output were used to weight the convolutional feature maps, which were then interpolated to the temporal resolution of the original ECG. Higher activation indicated signal regions contributing more strongly to the prediction.

To reduce beat-to-beat variation and transient noise and to visualize the relative importance of individual waveform components within a representative cardiac cycle, activation maps were aligned to detected beats and averaged within each recording. The resulting beat-averaged Grad-CAM maps were examined for emphasis on clinically relevant regions, particularly those corresponding to the J-point and ST-T morphology in leads V1-V2. Grad-CAM was used to characterize temporal attribution rather than establish lead-specific importance or causal physiological relationships.

3. Results

Table 1 summarizes the performance of the MS-Hybrid Transformer across 10 rotating evaluation loops. The model achieved an AUROC of 0.867, AUPRC of 0.698, accuracy of 0.871, and specificity of 0.918, indicating good discrimination and high specificity within this predominantly control cohort. Sensitivity was 0.671, showing that approximately two-thirds of positive-class ECGs were detected.

The F1-score was 0.663 and precision was 0.668, reflecting moderately balanced performance for the positive class. Precision varied more across loops than accuracy or specificity, likely because of the limited number of Brugada cases in each test split. Overall, performance was characterized by high specificity but substantially lower sensitivity.

Table 1. Performance of the Multi-Scale Hybrid Transformer across 10 rotating evaluation loops.

Metric	Mean ± SD
F1-score	0.663 ± 0.056
Sensitivity	0.671 ± 0.072
Specificity	0.918 ± 0.039
Precision (PPV)	0.668 ± 0.100
Accuracy	0.871 ± 0.027
AUROC	0.867 ± 0.050
AUPRC	0.698 ± 0.093

As illustrated in Figure 3, Grad-CAM visualizations for two representative Brugada ECGs highlighted diagnostically relevant regions in leads V1 and V2. Higher Grad-CAM importance, indicating a stronger contribution to the model prediction, is shown in red, whereas lower importance is indicated in blue. The activation maps primarily emphasized the J point, ST segment, and early T wave regions, which are central to the electrocardiographic assessment of Brugada patterns. Similar attention distributions across the two samples suggest that the MS-Hybrid Transformer captures reproducible temporal features rather than isolated artifacts or nonspecific signal variation. The consistent emphasis on right precordial morphology supports the physiological plausibility of the model predictions and improves the interpretability of the proposed framework.

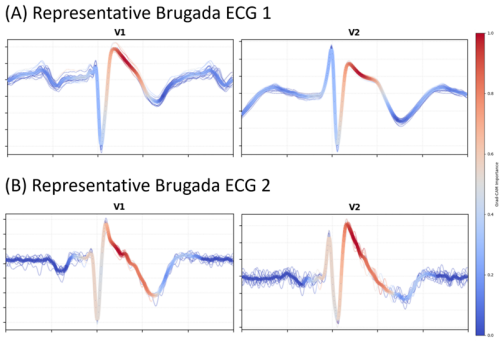

Figure 3. Beat averaged Grad-CAM visualizations of two representative Brugada ECGs in leads V1 and V2.

4. Discussion and Future Work

The proposed MS-Hybrid Transformer provides a proof of concept for Brugada ECG classification by combining multiscale convolutional features with Transformer modeling of full 10 second, 12 lead ECGs. It captures localized right precordial morphology and broader temporal relationships, although subtle and heterogeneous patterns remain challenging.

Grad-CAM maps showed clinically plausible temporal localization, emphasizing the terminal QRS, J point, ST segment, and early T wave in V1 and V2. These exploratory findings require validation through lead specific occlusion, lead ablation, quantitative attribution analysis, and blinded expert assessment.

Several limitations should be noted. The Brugada-HUCA database includes a relatively small number of Brugada cases from a single center, and the control group does not adequately represent the full spectrum of clinically challenging Brugada mimics, limiting generalizability. Future work should therefore include external validation in larger multicenter cohorts and evaluation against clinically relevant mimics and phenocopies, such as right bundle branch block, early repolarization, and other causes of right-precordial ST elevation, to improve the robustness and specificity of Brugada detection. Future analyses should separately evaluate confirmed Brugada (code 1), other/atypical cases (code 2), pathological basal-pattern subgroups, and clinically relevant mimics. Sensitivity-oriented threshold optimization, uncertainty estimation, and integration of additional clinical information may be required before evaluation as a screening or clinical decision-support tool.

References

[1] C. Antzelevitch *et al.*, "Brugada syndrome: report of the second consensus conference: endorsed by the Heart Rhythm Society and the European Heart Rhythm Association," *Circulation*, vol. 111, no. 5, pp. 659-670, 2005.

[2] Z. Li, D. J. Dzikowicz, B. Bailey, M. Silka, and J. An, "Multi-Domain Quantitative ECG Analysis Reveals Distinct Morphological Patterns in Brugada Syndrome," in *2026 Computing in Cardiology (CinC)*, 2026, pp. 1-4.

[3] A. Vaswani *et al.*, "Attention is all you need," *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.

[4] H. Lee and J. An, "Hybrid Convolution-Transformer Network Integrating Time and Frequency Domains from 12-Lead ECG for Chagas Disease Detection," in *2025 Computing in Cardiology (CinC)*, 2025, pp. 1-4.

[5] J. An, R. E. Gregg, H. Lee, D. J. Dzikowicz, and B. Bailey, "Multi-Scale Hybrid Transformer for Detecting Transient Myocardial Ischemia Using 10-Second 12-Lead STAFF III ECGs," in *2026 48th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE, 2026, pp. 4436-4441.

[6] J. An *et al.*, "Automated Pediatric Ventricular Pre-Excitation (Wolff-Parkinson-White Syndrome) Detection in 12-Lead ECG Using Deep Learning and SHAP-Based Interpretability," in *2025 Computing in Cardiology (CinC)*, 2025, pp. 1-4.

[7] J. An, B. Bailey, Y.-H. Zhang, and D. J. Dzikowicz, "Left bundle branch block detection in 12-Lead ECG using End-to-End deep learning with explainability," in *2024 Computing in Cardiology (CinC)*, 2024, pp. 1-4.

[8] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618-626.

[9] C. Cortez *et al.*, "Brugada-HUCA: 12-Lead ECG Recordings for the Study of Brugada Syndrome," *PhysioNet*, 2026.

[10] J. An, R. E. Gregg, and S. Borhani, "Effective data augmentation, filters, and automation techniques for automatic 12-lead ECG classification using deep residual neural networks," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE, 2022, pp. 1283-1287.

[11] J. An, B. Bailey, and R. E. Gregg, "Paradigm Shift from Feature-Based Machine Learning to End-to-End Deep Residual Neural Networks for Pediatric Age Classification from 12-Lead ECG," in *2023 Computing in Cardiology (CinC)*, IEEE, 2023, pp. 1-4.

Address for correspondence:

Junmo An, PhD
222 Jacobs St, Cambridge, MA 02141, USA
junmo.an@philips.com