INSTITUTE OF INFORMATION AND COMMUNICATION TECHNOLOGIES
BULGARIAN ACADEMY OF SCIENCES
CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume **25**, No 4
Sofia • 2025 Print ISSN: 1311-9702; Online ISSN: 1314-4081
DOI: 10.2478/cait-2025-0042

A Semi-Supervised TCN-LSTM Model for Single Lead ECG Heartbeat Classification

Nitya N. Kulkarni[1], *Nagaraja G. S.*[1], *B. G. Sudarshan*[2], *Krishna M.*[3]

[1]*Department of Computer Science and Engineering, Center for Healthcare Technologies Research, RV College of Engineering, Visvesvaraya Technological University, Belagavi – 590018, India*
[2]*Department of Electronics and Instrumentation Engineering, Center for Healthcare Technologies Research, RV College of Engineering, Visvesvaraya Technological University, Belagavi – 590018, India*
[3]*Department of Mechanical Engineering, Center for Healthcare Technologies Research, RV College of Engineering, Visvesvaraya Technological University, Belagavi – 590018, India*
E-mails: *nityank.phdcs@rvce.edu.in nagarajags@rvce.edu.in sudharshanbg@rvce.edu.in krishnam@rvce.edu.in*

Abstract**:** *This research presents a semi-supervised hybrid Temporal Convolutional Network-Long Short-Term Memory* (*TCN-LSTM*) *model for interpretable and data-efficient ElectroCardioGram* (*ECG*) *heartbeat classification. ECG signals from the MIT-BIH and INCART databases were resampled at* 125 Hz*,* 4th *order Butterworth filtered* (*0.5-20 Hz*)*, and segmented into* 0.8 s (188-*sample*) *windows* (279,641 *beats*)*. The architecture integrates two Temporal Convolutional Network* (*TCN*) *blocks* (*kernel* = 3*, receptive field* = 63) *with parallel* 64*-unit Long Short-Term Memory* (*LSTM*) *layers fused via element-wise maximum to capture both local and global temporal dynamics. Data were split beat-wise* (60/20/20 *for SL;* 80/20 *for SSL*)*, with* 10-30% *labeled beats and pseudo-labels generated using adaptive thresholding. The model achieved* 0.980 *accuracy and an F1-score of* 0.870 *in supervised learning and* 0.979 *accuracy and an F1-score of* 0.850 *in semi-supervised mode using* 30% *labeled data, outperforming comparable deep learning architectures. Guided Grad-CAM visualizations highlighted activations over QRS and R-peak regions, validating the physiological interpretability and diagnostic potential.*

Keywords**:** *ElectroCardioGram* (*ECG*)*, Heartbeat classification, Temporal Convolutional Network* (*TCN*)*, Long Short-Term Memory* (*LSTM*)*, Semi-supervised learning.*

1. Introduction

An ElectroCardioGram (ECG) represents the electrical activity of the heart, depicting heart rate and rhythmicity of cardiac cycles. Accurate heartbeat classification from ECG signals plays a vital role in the early detection and monitoring of cardiovascular diseases. By analyzing the temporal variations in cardiac electrical activity, ECG analysis enables the identification of discrete cardiac events. PQRST complex alterations and other abnormalities in ECG morphology frequently point to

underlying cardiac conditions, which can range from benign ectopic beats to potentially fatal heartbeats. By reducing diagnostic workload and subjectivity, automated heartbeat classification systems enable physicians to diagnose such conditions quickly and reliably [1]. Some common Deep Learning (DL) methods used for ECG heartbeat classification are Artificial Neural Networks (ANN) [2], Convolutional Neural Networks (CNN) [3], Temporal Convolutional Neural Networks (TCN) [4], Recurrent Neural Networks (RNN) [5] and Long Short-Term Memory (LSTM) Neural Networks [6] although each one of these have relative merits and demerits. ANNs are computationally efficient yet fail to capture temporal dependencies. CNNs [7] are very efficient in extracting spatial features, while having a limited receptive field restricts long-term analysis. TCN [8] overcomes this drawback by using dilated convolutions to model temporal dependencies; however, this requires careful design and a substantial amount of data. RNNs can model a temporal sequence but suffer from vanishing gradients. This is overcome by LSTM [9] using their gated memory units. However, LSTM networks alone may also fail to capture the entire range of temporal patterns [10]. The limitations of individual models have prompted researchers to explore ensembles or hybrid DL models to increase heartbeat classification performance [11].

Recent works have underlined how DL methods for temporal modeling and optimization find applications beyond their original domains. Barot and Kapadia [12] have used LSTM with attention for air quality prediction, identifying the importance of learning long-term dependencies in sequential data. Alazzam, Alaboudi and Abubakar [13] employed a Binary Owl Optimizer in IoT forensics for efficient feature selection to improve classification accuracy, underlining the importance of optimized representation learning. Similarly, Karimunnisa, Ramya and Krishna [14] have applied deep learning to workload forecasting and optimization in cloud computing by integrating predictive modeling with adaptive scheduling. While these are domain-specific, they reinforce key principles-temporal dependency modeling, selective feature extraction, and optimization-which meet the goals of this work for designing an efficient and generalizable hybrid TCN-LSTM framework for ECG heartbeat classification.

Hybrid CNN-RNN architectures have yielded the state-of-the-art performance for AAMI EC57 [15] beat-level classification. On the MIT-BIH Arrhythmia Database (MIT-BIH) [16], CNN-GRU [17], CNN-BiLSTM [18], and CNN-LSTM-SE [19] hybrids reported accuracy of 98.98%, 99.51%, and 98.5% respectively. In addition, CAT-Net (CNN + Channel Attention + Transformer Encoder) [20] achieved 99.14% on MIT-BIH and 99.58% on the St. Petersburg INCART (INCART) dataset [21]. A lightweight CNN-BiLSTM with less than 1M parameters sustained 99.21%, 96.17%, and 97.20% on MIT-BIH, INCART, and MIT-BIH Supraventricular datasets [22]. In rhythm-level detection, the ensemble CNN-BiLSTM reached 98.5% accuracy in the PhysioNet Challenge 2021 dataset [23]. Overall, most of the hybrid models have been mainly explored on beat-level ECG classification. While hybrid CNN-RNN architectures have shown phenomenal performances on individual datasets, only a few studies have combined multiple datasets to classify arrhythmias at the beat [24] or rhythm levels [25]. Although the

high single-dataset accuracies were achieved, no previous hybrid deep learning study combined the MIT-BIH and INCART databases to test cross-demographic generalization at limited labeled data conditions.

Most of the ECG datasets are expert-annotated, which restricts the use of abundant unlabeled signals. Recent Supervised Learning (SL) models mimic diagnostic reasoning without manual labeling. K r a s t e v a et al. [26] used machine-generated reference labels on the CSE dataset to objectively benchmark encoder–decoder networks with residual and recurrent links. Their work highlighted that machine-generated annotations can effectively substitute expert labels and represent a critical enabler for scalable evaluation across large arrhythmia databases. Semi-Supervised Learning (SSL) relaxes this constraint and usually leverages a small amount of labeled data and a large pool of unlabeled data by means of pseudo-labeling and contrastive learning techniques [27]. R a s m u s s e n et al. [28] have reported 98.7% accuracy using 5% of labeled data on the MIT-BIH Atrial Fibrillation Database, while S h i et al. [29] reported an improvement in accuracy up to 13.6% on the CPSC2018, PTB-XL, and Chapman databases using hybrid pseudo-labeling and consistency regularization. Y i n g et al. [30] presented FedECG, a federated SSL framework on MIT-BIH, yielding 94.8% accuracy with 50% of data as unlabeled. Z h a i, Z h o u and T i n [31] presented a patient-specific 2D CNN SSL model on MIT-BIH with sensitivities of 91.8% and 75.6% for ventricular and supraventricular beats, respectively. S h i et al. [32] further improved the accuracy by 5.15% using adaptive threshold-based SSL on Chapman, PTB-XL, and SNPH datasets. However, most of the SSL studies process ECG as images and rely only on single DL architectures, which limits the temporal feature learning from the ECG signal.

While hybrid CNN-RNN models are dominating ECG classification [17-22], most of the existing architectures rely on a sequential pipeline and thereby may not effectively capture the full hierarchy of temporal features. Effective analysis needs modeling both local temporal features, for example, the shape and duration of P-wave, QRS complex, and T-wave, and global temporal features, such as the R-R interval trends and heart rate variability across the span of several beats. The integration of TCNs, which intrinsically model multi-scale patterns through the operation of dilated convolutions [33], with LSTMs in a hybrid architecture remains an important lacuna in ECG analytics. This paper proposes a hybrid TCN-LSTM to directly model these distinct temporal dynamics comprehensively for enhanced heartbeat classification.

Current SSL frameworks for ECG analysis are mostly architected around standalone feature extractors, which are predominantly CNNs [29-31] and Variational Auto-Encoders (VAEs) [28]. While these models leverage unlabeled data effectively through pseudo-labeling and consistency regularization [29, 30], their design depends on single-architecture backbones and lacks mechanisms to integrate multi-scale temporal dependencies critical for precise beat interpretation. In the realm of fully SL, on the other hand, it has been well established that hybrid models capture both local and global temporal features as discussed previously. The integration of such evidence-based hybrid temporal architectures within SSL paradigms, however, remains vastly unexplored.

The critical gap related to the usage of a hybrid model in the SSL framework to capture the temporal features of ECG for heartbeat classification is addressed in this research work. Specific research aims are to:

- Formally validate the hybrid TCN-LSTM architecture, trained within a semi-supervised framework, can achieve supervised-level accuracy in AAMI EC57 5-class single-lead heartbeat classification on combined datasets belonging to different demographic patients.
- Investigate the performance of the proposed framework under an extreme label scarcity of ≤30% labeled data by testing its robustness and generalizability.
- Provide interpretable visual explanations of model decisions for various classes of heartbeats, with confirmation that learned features match clinically relevant ECG morphology as a means of building diagnostic trust.

By advancing architectural design and data-efficient learning together, this work provides a key step towards scalable, interpretable, and clinically viable tools for heartbeat-based arrhythmia screening in resource-constrained environments, including ambulatory and wearable monitoring.

2. Materials and methods

2.1. Description of datasets

The two publicly available ECG databases with beat-level annotations were utilized to develop and evaluate the proposed model: the MIT-BIH Arrhythmia Database (MIT-BIH) [16] and the St. Petersburg Institute of Cardiological Technics (INCART) database [21]. The MIT-BIH database is a standard dataset used for benchmarking heartbeat classification algorithms. The INCART database serves as a complementary 12-lead data source with heartbeat annotations obtained from a different clinical context. Since both datasets have beat-level annotations, they were considered in this cross-dataset experimentation.

The MIT-BIH dataset consists of 48 half-hour ambulatory ECG recordings of two leads belonging to 47 subjects with a sampling frequency of 360 Hz. As mentioned by G o l d b e r g e r et al. [16], each beat of the ECG was labeled using the standard AAMI EC57 grouping them into 15 beat classes mapping to five main classes. MLII lead ECG, along with their five main beat classes (provided in the dataset), considered for this research, are mentioned in Table 1. The recordings of 102, 104, 107, 114, and 217 of the MIT-BIH dataset were not employed, as they have paced beats.

The INCART dataset has 75 half-hour ECG recordings of 12 leads of 32 subjects with a sampling frequency of 257 Hz. As mentioned in [21], beat level annotations for each ECG recording were grouped into 12 beat types. These 12 beat types were mapped to the five beat classes of AAMI in this study for lead II ECG, as mentioned in Table 1. The single lead of the ECG of both datasets was considered for classification to study the prospective usage of the model for wearable devices. Annotation inconsistencies between datasets were harmonized by mapping both to AAMI EC57 classes (Table 1), ensuring consistent 5-class labeling despite differing original schemes.

Table 1. Distribution of samples and annotations in datasets

AAMI EC57 class	MIT-BIH		INCART	
	Annotations	Samples	Annotations	Samples
Normal beat (N)	N, L, R, e, j	90,126	N, L, R	153,674
Supraventricular ectopic beat (S)	A, a, J, S	2781	A, S, J	1960
Ventricular ectopic beat (V)	V, E	7009	V, E	20,013
Fusion beat (F)	F	803	F	219
Unclassified beat (Q)	Q, f, /,?	3004	Q, /, U	52

2.2. Data preprocessing

ECG records obtained from the MIT-BIH and INCART databases were uniformly preprocessed with a pipeline that consisted of resampling, denoising, segmentation, normalization, and dataset splitting.

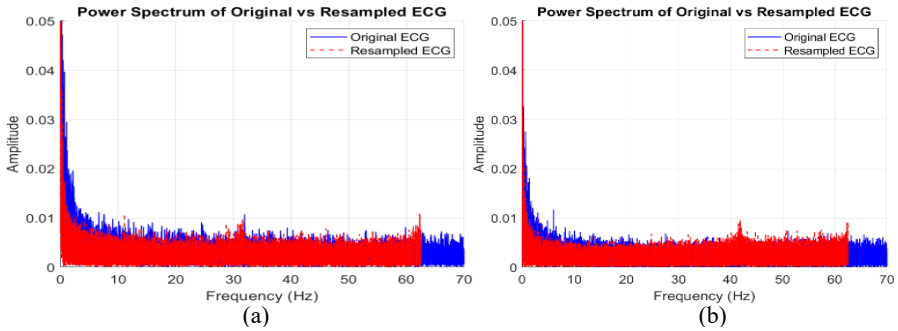

Fig. 1. Power spectrum of the MIT-BIH dataset (a); power spectrum of the INCART (b)

Signals were resampled to 125 Hz to unify the 360 Hz rate of the MIT-BIH and the 257 Hz rate of the INCART database while retaining diagnostic fidelity for beat-level classification. In order to ensure that key frequency components were not lost, the average Power Spectral Density (PSD) was computed before and after resampling, as shown in Fig. 1a and Fig. 1b for the two datasets. The analysis showed that over 97% of the total signal energy was below 20 Hz, thus validating that the down-sampling did not eliminate diagnostically relevant frequency content.

A 4th-order Butterworth band-pass filter (0.5-20 Hz) was applied to remove baseline wander and high-frequency noise. Though broader bandwidths are common in clinical ECG processing, literature supports the fact that the majority of heartbeat-morphology energy resides below 20 Hz [34-35]. Besides, K r a s t e v a et al. [36] in their encoder-decoder study on 12-lead ECG beats have shown the feasibility of low-pass filtering without a loss of diagnostic interval measurement accuracy. In this respect, it further justifies the use of a 20 Hz upper cutoff as sufficient for beat-level morphology in deep-learning contexts.

The window size was 0.4 s pre-R and 0.4 s post-R, providing a total 0.8 s window per segment to capture the full P-QRS-T complexes, including ectopic morphologies. This window length was determined by empirical tuning and also comes well within the range of various prior studies dealing with beat classification.

At 125 Hz, 0.8 s corresponds to 100 samples. To allow for the receptive field of the hybrid TCN-LSTM architecture (dilations {1, 2, 4, 8, 16}), each segment was fixed at 188 samples by zero-padding. Zero-padding preserves the temporal

alignment of events and prevents morphological distortion. Then, each segment was normalized by min-max normalization [0, 1] to standardize the amplitude across subjects and recordings.

For SL, the combined dataset was partitioned into 60% for training, 20% for validation, and 20% for testing. For SSL, the data was first split into 80% for training and 20% for testing. From the training portion, subsets of 10% and 30% were retained as labeled data, while the remainder was treated as unlabeled data, as shown in Fig. 2. The significant class imbalance was addressed by applying the Synthetic Minority Over-sampling Technique (SMOTE) [37] exclusively to the labeled training and validation splits. The model was always evaluated on the original, non-SMOTE test data. The final distribution of beats for these splits is detailed in Table 2.

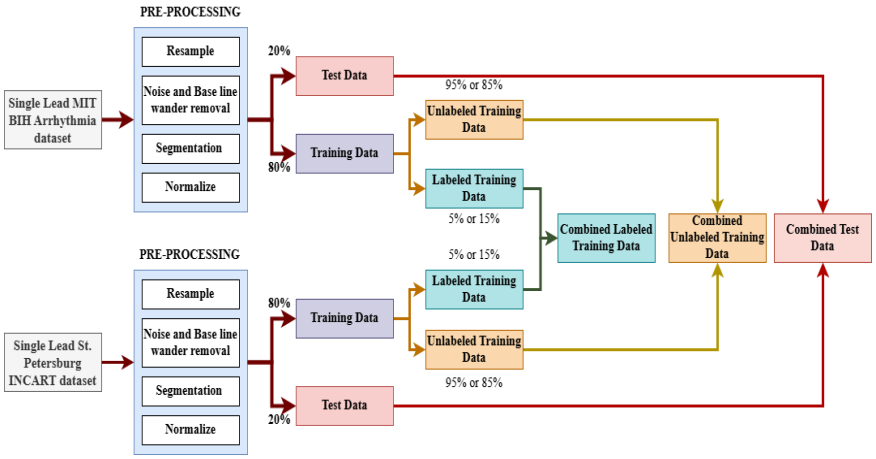

Fig. 2. Preprocessing and Data bifurcation of the ECG data in SSL framework

Table 2. Data distribution for SSL

Classes of heartbeat	10% labeled dataset		30% labeled dataset		Original test data
	Labeled training data	Unlabeled training data	Labeled training data	Unlabeled training data	
N	9752	185,310	29,259	165,803	48,738
S	192	3653	576	3269	896
V	1080	20,511	3239	18,352	5431
F	39	759	120	678	224
Q	121	2295	362	2054	640
Total (before SMOTE)	11,184	212,528	33,556	190,156	55,929
Total (after SMOTE)	48,760	926,550	146,295	829,015	

2.3. Hybrid TCN-LSTM model

The proposed hybrid TCN-LSTM architecture (Fig. 3) processes ECG segments of length 188 samples, corresponding to one cardiac cycle window after preprocessing. Each input sample has a single channel [188×1], and the model outputs a five-dimensional softmax probability vector corresponding to the five heartbeat classes – N, S, V, F, and Q. The architecture, as illustrated in Fig. 3, integrates TCN for

extracting short-term local temporal dependencies and LSTM layers for capturing long-term temporal dependencies.

The model employs two TCN blocks that contain five 1D separable convolutions with dilation rates $\{1, 2, 4, 8, 16\}$ and kernel size $k = 3$. This configuration expands the Receptive Field (RF) to 63 samples, effectively spanning multiple cardiac cycles. The RF was empirically tuned to ensure coverage of P-QRS-T waveform morphology while avoiding over-smoothening of local signal variations as validated for ECG analysis in recent TCN-based arrhythmia studies [38-40]. Residual connections and layer normalization enhance gradient stability and mitigate vanishing effects during deep temporal modeling [38].

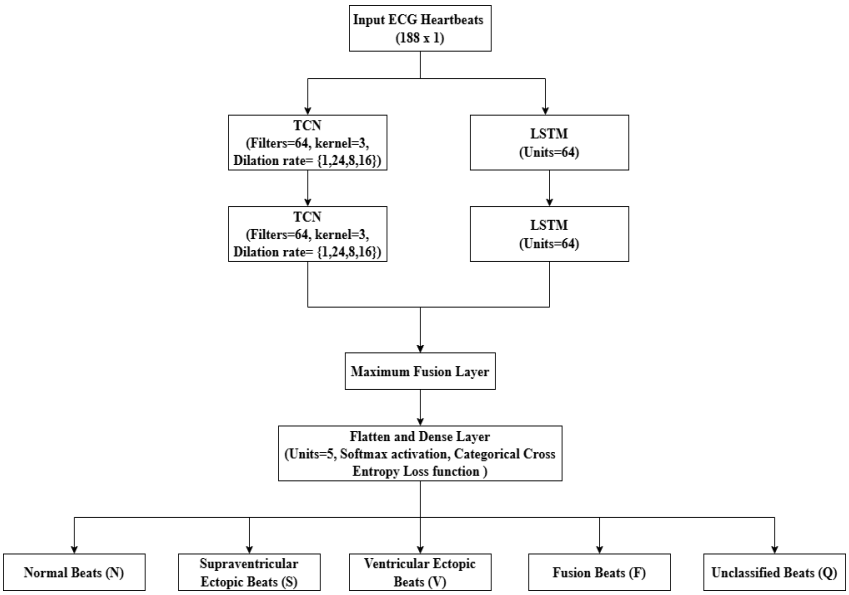

Fig. 3. TCN-LSTM model architecture diagram

The preprocessed ECG segments were provided to the LSTM layers of 64 units in parallel to extract temporal features at a long-term scale. The number of units for TCN and LSTM was determined through iterative experimentation with 16, 32, 64, and 128 units.

These experiments revealed that both TCN and LSTM layers performed better, with 64 units giving the best trade-off between accuracy and convergence speed [19]. Other hyperparameters, such as the number of layers, kernel size of 3 for the TCN, batch size of 128, 40 epochs, and learning rate of 0.0001, were determined on the basis of prior literature and preliminary experiments [41]. Dominant features from the TCN and LSTM layers were extracted by the maximum operation. Other approaches of merging layers of two DL models, such as concatenation, averaging, and multiplication methods, add complexity and smooth the features, reducing the performance. Further, the maximum layer was flattened and was passed to a dense layer with softmax activation to classify the heartbeats into five classes: N, S, V, F, and Q. The Flattened layer reshapes the multi-dimensional temporal feature maps to

a single-dimensional vector essential for the classification layer. The model was compiled using the Adam optimizer and categorical cross-entropy loss function, which is appropriate for multi-class classification tasks such as heartbeat classification.

2.4 Semi-supervised learning

SSL is advantageous as it reduces dependence on expert annotation costs, has domain adaptation potential, and is scalable to real-world unlabeled data. Hence, the pseudo-labeling technique of SSL was explored to train the model. Pseudo-labeling [42] is a popular SSL method that creates artificial labels (pseudo-labels) for unlabeled samples based on predictions made by the model. It employs only high-confidence predictions to be used as training samples, thus increasing the effective training set with self-labeled instances. The process employed in this study follows a structured five-step pipeline as shown in Fig. 4: (1) Train the TCN-LSTM model using a small labeled subset of ECG beats; (2) Use the trained model to generate pseudo-labels for the unlabeled data (Unlabeled training data); (3) Apply a thresholding mechanism (confidence, FlexMatch, adaptive, or entropy-based) to filter high-confidence pseudo-labels (pseudolabeling); (4) Combine the labeled and pseudo-labeled samples to form an expanded training set (combined training data); and (5) Retrain the model on this combined training data and evaluate its performance on the independent combined test data (20% of MIT-BIH and INCART datasets each).

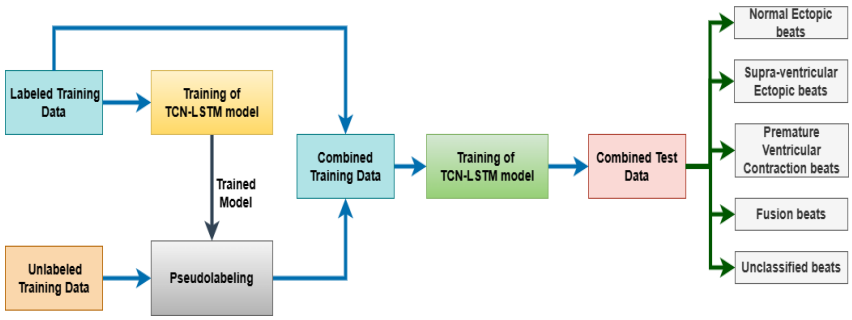

Fig. 4. TCN-LSTM model in the SSL method

Pseudo-labeling techniques, such as confidence thresholding [43], FlexMatch [43, 44], adaptive thresholding [32], and entropy-based adaptive thresholding [45, 46], were experimented with because they represent different and popular paradigms in pseudo-label generation, ranging from static to dynamically adaptive and uncertainty-aware approaches. These techniques ensure well-balanced coverage of confidence-based and entropy-driven learning mechanisms, offering robustness against label noise and class imbalance. More complex alternatives, including graph-based or generative SSL methods, were excluded due to higher computational cost and reduced interpretability.

Confidence thresholding is a simple pseudo-labeling method in which the pretrained model generates predictions on unlabeled data. These predictions represent the model's confidence in the possible classes. The model selects

predictions with a confidence score above a predefined threshold (τ= 0.95) to ensure that only reliable pseudo-labels are used, as shown in the equation

(1) $$\text{mask} = \text{max\_probs} \geq \tau,$$

where max_probs is the highest Softmax probability for each sample. The model generates pseudo-labels $\hat{y}$, for the model's raw output $f(x)$ applied for input x by selecting the class with the highest probability, as shown in the equation

(2) $$\hat{y} = \text{argmax}(f(x)).$$

Confidence thresholding is sensitive to the threshold set, unlike the FlexMatch method, which can dynamically adjust the confidence threshold per class. More frequent classes were assigned higher thresholds (up to 0.95), making the creation of new pseudo-labels difficult, whereas minority classes were assigned lower thresholds (down to 0.7), creating pseudo-labels more easily. This class-wise thresholding helps minority classes contribute more towards the training process and avoids model bias towards the frequent classes. FlexMatch also uses Exponential Moving Average (EMA) to smooth threshold updates over time with α set to 0.99. Dynamically balancing class representations improves generalizability and robustness in cases with imbalanced data or when constrained to work with limited labeled data. Pseudo-label is assigned per class with the highest probability as in the Equation (2). The threshold for each class was considered to be $\tau_{\hat{y}(x)}$, which was applied by the model to determine the confidence of the pseudo-label, as shown in the equation

(3) $$\text{mask}(x) = 1(\text{max\_prob}(x) \geq \tau_{\hat{y}(x)}).$$

The class-specific thresholds were updated via EMA smoothing, as

(4) $$\tau_c = \alpha\tau_c + (1-\alpha)(1-r_c),$$

where τ_c is the updated confidence threshold for class c, α is the EMA smoothing factor set to 0.99, and r_c is the class ratio given by

(5) $$r_c = \frac{n_c}{\sum_{c=1}^{C} n_c + \epsilon},$$

where n_c is the number of pseudo-labels for class c in the batch given by

(6) $$n_c = \sum_{i=1}^{N} 1(\hat{y}(x_i) = c).$$

where $\in$ is a small smoothing constant that ensures stable threshold updates, and N is the total number of pseudo-labels in the batch;

Adaptive thresholding uses a class-specific dynamic thresholding mechanism with the memory of recent confidence scores (M_c=1000). The class-specific threshold was dynamically updated using the 90th percentile of the stored confidence, as

(7) $$T_{90}(c) = \text{percentile}_{90}(M_c).$$

The threshold for each class (τ_c) was determined through EMA smoothing, as

(8) $$\tau_c = \alpha \times \tau_c + (1-\alpha) \times T_{90}(c),$$

where α is the smoothing factor set to 0.99. The min-max constraints given in the next equations:

(9) $$\tau_c = \text{clip}(\tau_c, \tau_{\min}, \tau_{\max}),$$

(10) $$\tau_{\min} \epsilon \{0.5, 0.75, 0.6, 0.85, 0.8\},$$

$$\tau_{\max} \epsilon \{0.98, 0.9, 0.9, 0.93, 0.92\} \text{ for N, S, V, F, and Q,}$$

were employed on the threshold to prevent it from being too strict or lenient. Finally, the model applies class-specific thresholding to filter out low-confidence pseudo-

labels, as shown in the Equation (4). Thus, threshold updates were carried out on the basis of recent predictions of the Equation (5).

Entropy-based adaptive thresholding works similarly to adaptive thresholding but uses the entropy distribution of the model's prediction to dynamically update thresholds. Entropy computation is performed as shown in

(11) $$H(p) = -\sum_{i=1}^{C} p_i . \log(p_i),$$

where p_i is the predicted probability for class i. Lower entropy indicates higher confidence predictions. For each class, thresholds were dynamically updated on the basis of the 90th percentile of past entropy values, as shown in

(12) $$T_{90}(c) = \text{percentile}_{90}(\varepsilon_c).$$

Threshold updation was performed via (8) followed by class-specific min-max bounds, as indicated in (9) and (10) corresponding to classes {0: N, 1: S, 2: V, 3: F, 4: Q}. Pseudo-labels were selected on the basis of entropy filtering, as depicted in

(13) $$\text{mask}(x) = 1(H(p(x)) \leq \tau_{\hat{y}(x)},$$

where $\hat{y}(x)$ is the pseudo-labels as shown in (2).

2.5 Training configuration and Benchmark setup

The hybrid model was implemented in SL and SSL using Google Colaboratory with a Tesla T4 GPU. MATLAB R2024a with WFDB Toolbox was used to preprocess ECG data from the two datasets. The TCN-LSTM model was trained using the Adam optimizer with a learning rate of 0.0001, batch size of 128, and a maximum of 40 epochs. Early stopping was applied based on validation loss with a patience value of 5, preventing overfitting and ensuring optimal model generalization. Performance metrics-accuracy, macro-average precision, macro-average recall, macro-average F1-score, and AUROC were computed on the test data using the best-performing checkpoint obtained during training.

For comparative analysis, five benchmark architectures were implemented under identical experimental conditions: CNN, LSTM, TCN, CNN-LSTM hybrid, and CNN-BiLSTM hybrid (CNN-Bidirectional LSTM). The CNN was comprised of four Conv1D layers (64 filters, kernel = 5), each followed by max-pooling to extract local beat morphology. The LSTM model consisted of four LSTM layers with 64 units for long-range temporal modeling. The TCN model consisted of four TCN layers with separable conv1d layers (64 filters, kernel = 3) having dilation rates {1, 2, 4, 8, 16}. The hybrid model CNN-LSTM consisted of two layers of CNN (64 filters, kernel = 5) followed by max-pooling and two LSTM layers (64 units). The CNN and LSTM layers were concatenated, followed by a dense layer (64 units). The CNN-BiLSTM layers also utilize the same hyperparameters and layers as CNN-LSTM, except that the LSTM used is bidirectional. To benchmark the proposed hybrid TCN–LSTM, all the architectures were trained under identical pre-processed data, input shape (188 × 1), and optimization settings (Adam optimizer with learning rate of 1×10^{-3}, batch size of 128, categorical cross-entropy).

2.6. Model interpretability

To enhance clinical transparency and address the limitations of black-box DL models in ECG analysis, we employed Guided Grad-CAM [47] for visualizing the

TCN-LSTM hybrid's decision-making process. While originally designed for 2D image data, we adapted this technique for 1D ECG signals by modifying the gradient computation and visualization approach, as validated in recent ECG interpretability studies [48].

For a given input ECG segment x of length 188 samples, we computed the gradients of the predicted class score y^c with respect to the feature maps A^k of the maximum fusion layer (used for combining TCN and LSTM outputs) as indicated in the equation

$$\alpha_k^c = \frac{1}{z}\sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k}, \tag{14}$$

where α_k^c represents the importance weights for feature map k, and z is the normalization factor. We implement guided backpropagation using a custom ReLU function that only propagates positive gradients from positive activations, as depicted in the equation

$$R_i = \begin{cases} \frac{\partial y^c}{\partial x_i} & \text{if } x_i > 0 \text{ and } \frac{\partial y^c}{\partial x_i} > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{15}$$

The Class Activation Map (CAM) is obtained by linearly combining the feature maps with their importance weights as indicated in the equation

$$L_{\text{Grad-CAM}}^c = \text{ReLU}(\sum_k \alpha_k^c A^k). \tag{16}$$

The final saliency map is generated by element-wise multiplication of the Grad-CAM output with the guided backpropagation signals as

$$L_{\text{Guided Grad-CAM}}^c = L_{\text{Grad-CAM}}^c \odot R. \tag{17}$$

The resulting 1D saliency map is upsampled via linear interpolation to match the original input resolution (188 samples) and normalized to a [0, 1] scale.

The input ECG segments were constructed with 100 non-zero samples within a 188-sample window, with 88 zero-padded samples. This padding strategy was essential for allowing the TCN to accommodate its dilated convolutional receptive field while preserving the chronological ordering of cardiac events. All critical diagnostic information (P-QRS-T complexes) remains within the 100-sample segment, ensuring that no essential morphological features are lost in the zero-padded portions.

All visualizations were generated on held-out test samples from the combined MIT-BIH and INCART datasets. A sampling rate of 125 Hz yields a temporal resolution of 8ms, which is adequate to capture critical ECG morphological features, including P-wave duration (80-120 ms), QRS complex duration (80-100 ms), and T-wave characteristics, according to clinical literature [49].

3. Results

3.1. Supervised learning

The performance of the TCN-LSTM model was evaluated on the test set (*n*=55,929 beats). The confusion matrix for the model on combined test data in SL (Fig. 5) implies that the heartbeats of the Normal (N), Premature Ventricular contraction (V), and unclassified (Q) classes were correctly classified. Due to the similarities in the

morphology of the segments, 13% of Supraventricular ectopic (S), and 17% of Fusion beats (F) were misclassified as Normal (N), which is negligible. These percentages represent the proportion of misclassified beats within Each class.

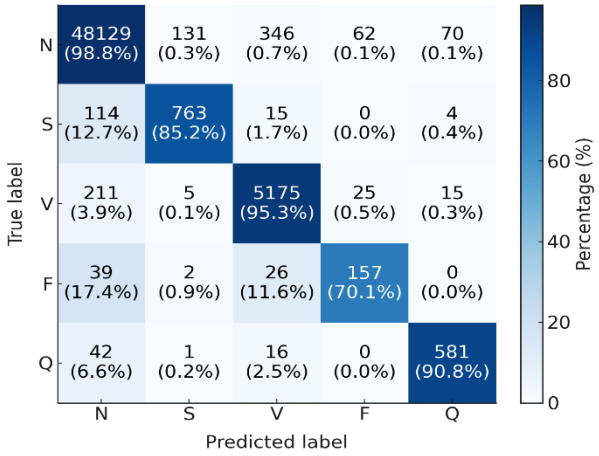

Fig. 5. Confusion matrix of the TCN-LSTM model on combined test data

The model was compared with other DL models trained on combined training data by keeping the same hyperparameters and tested on the combined test data, as shown in Table 3. The comparison shows that the TCN and TCN-LSTM models perform better than the other models in terms of accuracy and F1-score. The TCN-LSTM model achieves the highest precision of 0.860, implying its ability to recognize positive cases with fewer false positives. Even though the CNN model achieves the highest recall, indicating that it is better at identifying true positive cases, its low precision suggests a higher false positive rate. Comparatively, CNN-BiLSTM also shows better performance. Overall, the TCN-LSTM model achieves the best balance of accuracy, precision, and recall, making it a robust model for real-world ECG-based heartbeat classification.

Table 3. Comparison of the TCN-LSTM model with other DL models on the combined test data

DL Model	Accuracy	Precision	Recall	F1-score	AUROC
TCN	0.980	0.840	0.890	0.860	0.990
LSTM	0.969	0.770	0.890	0.820	0.990
CNN	0.972	0.760	0.910	0.820	0.990
CNN-LSTM	0.973	0.780	0.900	0.830	0.990
CNN-BiLSTM	0.979	0.810	0.900	0.850	0.990
TCN-LSTM	0.980	0.860	0.880	0.870	0.990

3.2. Semi-supervised learning

The TCN-LSTM model was trained and tested using different SSL methods such as confidence thresholding, FlexMatch, adaptive thresholding, and entropy-based adaptive thresholding, as indicated in Table 4. The number of unlabeled samples for the 10% and 30% labeled data was 926,550 and 829,015, respectively, after the application of SMOTE for the samples mentioned in Table 2. Compared with other

methods, the adaptive thresholding method yields better accuracy and F1-score with 10% labeled data.

Table 4. Pseudo-labeling methods using the TCN-LSTM model on the combined dataset

Pseudo-labeling method	Labeled data	Accuracy	Precision	Recall	F1-score	AUROC	Pseudo-labels
Confidence-thresholding (Th=0.95)	10%	0.968	0.780	0.810	0.800	0.960	753,917
	30%	0.979	0.890	0.830	0.860	0.970	759,459
FlexMatch (Th=0.95)	10%	0.972	0.830	0.810	0.820	0.960	753,917
	30%	0.976	0.860	0.840	0.850	0.970	759,459
Adaptive thresholding	10%	0.975	0.850	0.840	0.840	0.970	832,574
	30%	0.979	0.880	0.830	0.850	0.970	737,582
Entropy-based adaptive thresholding	10%	0.968	0.780	0.800	0.790	0.960	876,839
	30%	0.977	0.860	0.840	0.850	0.970	809,036

The performance of entropy-based adaptive thresholding is similar to that of adaptive thresholding, but it generates the highest number of pseudo-labels, 876,839, for 10% labeled data, indicating an aggressive pseudo-labeling strategy. The number of pseudo-labels decreases in the case of adaptive thresholding for 30% labeled data, implying that accurate pseudo-labels were generated with increasing labeled data. Overall, adaptive thresholding maintains a balance between accuracy, precision, and recall and generates more robust pseudo-labels.

Table 5. Comparison of DL models with TCN-LSTM using the adaptive thresholding-based SSL method with 30% labeled data

DL model	Accuracy	Precision	Recall	F1-score	AUROC	Pseudo-labels
TCN	0.970	0.790	0.840	0.810	0.970	737,062
LSTM	0.945	0.650	0.850	0.720	0.970	596,542
CNN	0.967	0.760	0.860	0.800	0.970	693,465
CNN-LSTM	0.969	0.790	0.850	0.820	0.970	700,422
CNN-BiLSTM	0.970	0.780	0.860	0.810	0.980	680,797
TCN-LSTM	0.979	0.880	0.830	0.850	0.970	737,582

Since the adaptive thresholding method showed better performance at both 10% and 30% labeled data for the TCN-LSTM model, it was further tested with other DL models. A comparison of different DL models with the TCN-LSTM model in the adaptive thresholding-based SSL method trained on 30% labeled data is shown in Table 5. The results indicate that the TCN-LSTM model outperforms other architectures when the same hyperparameters are used. Conversely, the TCN model has an accuracy of 0.970 and an F1-score of 0.810, which indicates that the inclusion of the LSTM component enhances the capacity of the model to learn temporal dependencies. The TCN-LSTM model produces the maximum number of pseudo-labels (737,582), indicating that its strong feature extraction ability and adaptive thresholding can provide more confident pseudo-labeling, as indicated in Table 5.

The radar plots shown in Fig. 6 reveal the performance of the TCN-LSTM model with SL and SSL for five types of ECG beats (N, S, V, F, and Q) in terms of precision, recall, and F1-score. The SSL model exhibited greater precision, recall, and F1-scores for V and N beats. However, SL seemed to outperform in terms of the number of S, F, and Q beats, especially in terms of the recall and F1-score. The results appear to

indicate that SSL enhances performance where beat types occur more frequently, and SL remains a stronger choice for more rare beats. It can also be inferred that the performance of the SSL method is similar to that of the SL method, indicating the effectiveness of the pseudo-labeling technique and DL model.

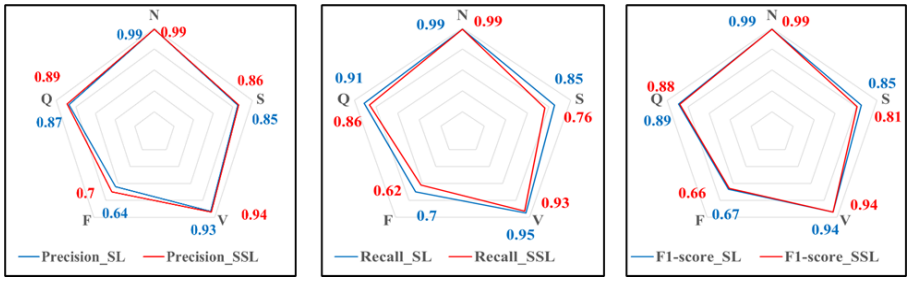

Fig. 6. Class-wise Precision, Recall and F1-score of TCN-LSTM model in SL and SSL method (30% labeled data), respectively

4. Discussion

4.1. Comparison with existing methods

Performance of the proposed model was compared with that reported in the literature, as shown in Table 6. This indicates that the proposed method, which uses a hybrid DL model in the SSL setting, is proposed for the first time on multiple datasets.

Table 6. Comparison of the proposed method with previous methods from the literature

Paper	Learning	DL model	Dataset used	Performance measures			
				Accuracy	Precision	Recall	F1-score
[50]	SL	HADLN with ResNet and BiLSTM	PhysioNet 2017 challenge	0.867	0.866	0.859	0.88
[52]	SL	Hybrid 2D CNN-LSTM	MIT-BIH in images	Average: 0.998	-	Average: 0.998	-
[53]	SL	Hybrid CNN-BiLSTM	MIT-BIH	0.9836	0.894	0.944	0.917
[41]	SL	1D CNN-LSTM	MIT-BIH & LTAF	0.982	-	0.861	-
[51]	SL	HMFF ConVexNet	MIT-BIH	0.993	0.993	0.993	0.993
Proposed method	SL	Hybrid TCN-LSTM model	MIT-BIH and INCART	0.980	0.860	0.880	0.870
[31]	SSL	2D CNN	MIT-BIH	SVEB: 0.974	0.590	SVEB: 0.938	0.725
				VEB: 0.986	VEB: 0.909	VEB: 0.875	VEB: 0.892
[30]	Federated SSL	ResNet-9 (50% labeled data)	MIT-BIH	0.9480	-	-	-
[29]	SSL	VGGNet (10% labeled data)	CPSC2018	0.7085			0.6451
			PTB-XL	0.7609	-	-	0.5871
			Chapman	0.9608	-	-	0.9569
[32]	SSL	Adaptive threshold-based SSL model (10% labeled)	Chapman	0.9613	-	-	0.9570
			PTB-XL	0.7682	-	-	0.5942
			SNPH	0.9443	-	-	0.9318
Proposed method	SSL	Hybrid TCN-LSTM model (10% labeled)	MIT-BIH and INCART	0.975	0.850	0.840	0.840
		Hybrid TCN-LSTM model (30% labeled)	MIT-BIH and INCART	0.979	0.880	0.830	0.850

The hybrid models proposed in the literature focus only on a single dataset, leading to less generalization of the model. In the SL environment, the proposed model performed with an accuracy of 0.980, precision of 0.860, recall of 0.880, F1-score of 0.870, and an AUROC of 0.990, indicating its ability to classify ECG heartbeats effectively. In the SSL configuration, the model was highly accurate even with sparse labeled data, reaching 0.975 accuracy using 10% labeled data and increasing to 0.979 accuracy with 30% labeled data, demonstrating the model's efficient use of unlabeled data with adaptive thresholding-based pseudo-labeling. Compared with previous methods, the new method outperforms many of the state-of-the-art methods, such as the HADLN [50] with ResNet and BiLSTM models (accuracy = 0.867) and the VGGNet-based SSL model [22], which has an accuracy of 0.7085 on CPSC2018. Although the HMFF ConVexNet model in SL [51] indicated a greater accuracy of 0.993 on the MIT-BIH dataset, its performance on merged datasets as a test of the generalizability is unknown. The model's balanced precision and recall in both SL and SSL ensure fewer diagnostic errors. The balanced accuracy, precision, recall, and F1-score performance, combined with the identical AUROC values, highlight the ability of the proposed model to process varied ECG data and thus be appropriately applied in real-world clinical environments.

4.2. Model interpretability

Performance of the model was interpreted by using Guided GRAD-CAM visuals for the maximum layer in SL and SSL, as shown in Fig. 7 and Fig. 8, respectively. Activation intensity from 0 (purple, low relevance) to 1 (yellow, high relevance), overlaid on the ECG signal (black line) and time axis in milliseconds (ms) relative to the R-peak (vertical line at 0 ms), is shown in these Guided GRAD-CAM.

The SL approach reveals distinct class-specific activation patterns that align with known cardiac electrical activity. Class N (Fig. 7a) exhibits strong early-phase activation around the R-peak, corresponding to normal ventricular depolarization patterns. Class S (Fig. 7b) exhibits broader attention over the QRS complex and early ST segment, probably capturing the aberrant conductive pathways characteristic of these beats. Ventricular ectopic beats (Fig. 7c) show unique dispersed activation patterns over several phases of the cardiac cycle, reflecting their distinctive morphological characteristics and sites of origin. The model highlights clinically relevant regions consistently across all classes, with particular emphasis on the QRS complex morphology. At the 125 Hz sample rate, temporal resolution is at 8ms intervals, a sampling which captures these important morphological features with computational efficiency in mind for real-time applications.

The SSL approach with adaptive thresholding at 30% labeled data (Fig. 8) retains physiologically meaningful attention patterns while showing some adaptation due to limited supervision. Class N (Fig. 8a) maintains strong QRS-focused activation, Class S (Fig. 8b) shows broader mid-cycle attention, and Class F (Fig. 8d) maintains well-localized late-phase activation. Class V (Fig. 8c) displays more distributed attention patterns compared to SL, potentially reflecting the model's adaptation to learn from confident pseudo-labels rather than full supervision. Comparative analysis confirms that both learning paradigms successfully identify

critical ECG components, with the SSL approach achieving comparable interpretability despite using substantially less labeled data.

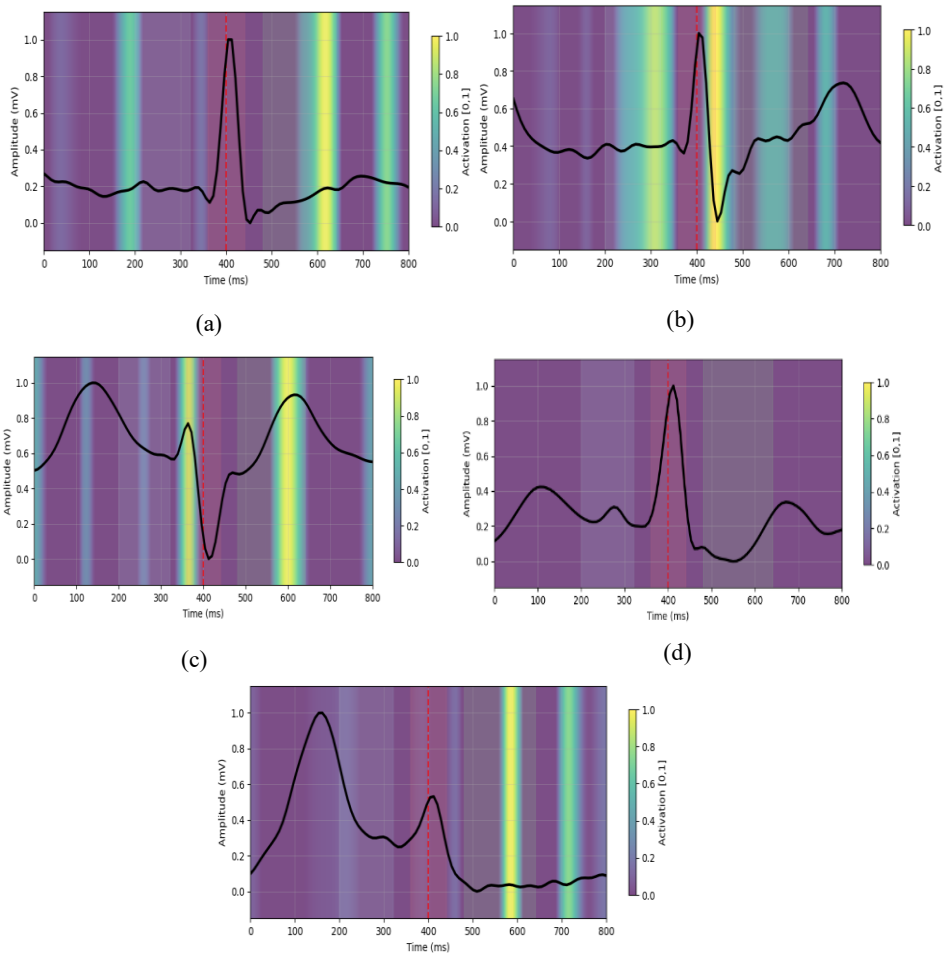

Fig. 7. Guided Grad CAM visuals for SL method: Normal beat (a); Supraventricular ectopic beat (b); Ventricular beat (c); Fusion beat (d); Unclassified beat (e)

Preservation of key diagnostic features across both learning paradigms confirms that the SSL approach learns medically relevant patterns rather than artifact-based correlations and engenders greater trust in the model's decision-making process for clinical applications.

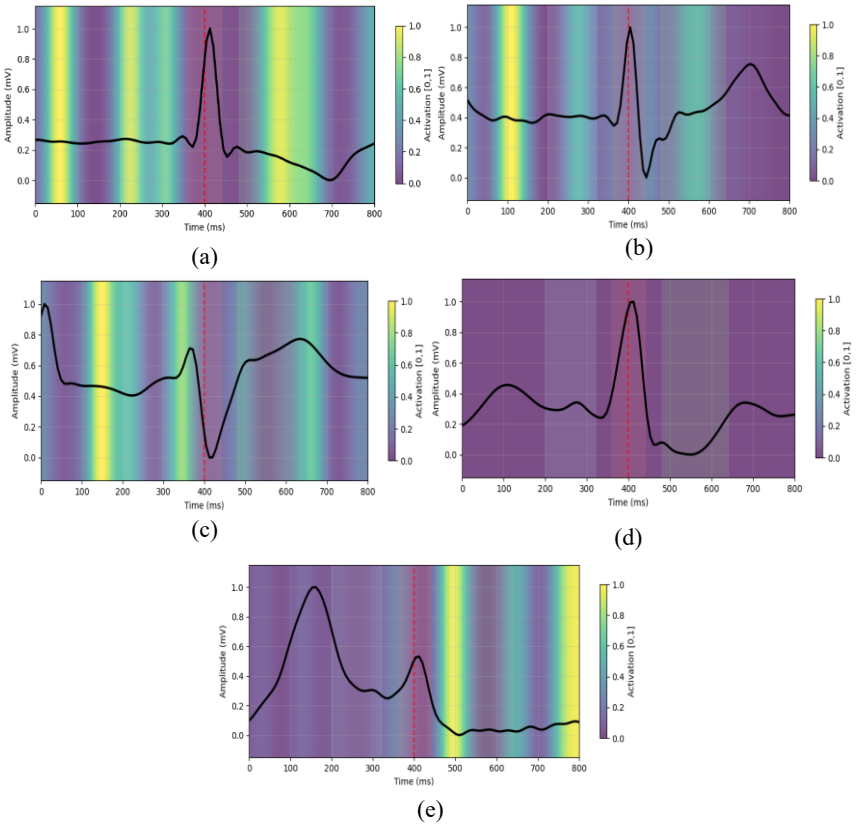

Fig. 8. Guided Grad CAM visuals for SSL method: Normal beat (a); Supraventricular ectopic beat (b); Ventricular beat (c); Fusion beat (d); Unknown beat (e)

5. Conclusion

This paper proposes a hybrid TCN-LSTM model combined with an SSL framework using adaptive threshold-based pseudo-labeling to effectively classify heartbeats. The generalization capability of the developed model across different ECG datasets was evaluated by training and validating using the combined dataset of the MIT-BIH and INCART databases. The proposed hybrid architecture achieved accurate, consistent performance in classification by capturing both long-term and local temporal dependencies within the ECG signals. In an SSL setup, it was able to use the unlabeled data efficiently by achieving 0.979 accuracy and an F1-score of 0.850 with just 30% labeled data, whereas its SL model accuracy reached up to 0.980 with a precision and F1-score of 0.860 and 0.870, respectively. Guided Grad-CAM visualizations further confirmed that SSL-based training enhances feature localization and morphological interpretability in ECG waveform analysis. Overall, the proposed TCN-LSTM model retained computational efficiency with architectural simplicity while competing or outperforming other architectures like CNN-LSTM, CNN-BiLSTM, and HMFF-ConvexNet.

These results demonstrate that SSL SSL-based hybrid architecture can retain diagnostic reliability while drastically reducing dependency on manual annotation. However, this research is constrained to benchmark datasets and single-lead ECG analysis. Future work will include extending the model to multi-lead ECG signals, exploring transformer-based fusion mechanisms, and testing the system in either wearable or real-time clinical ECG monitoring applications.

References

1. Rajpurkar, P., A. Y. Hannun, M. Haghpanahi, C. Bourn, A. Y. Ng. Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks. – Nature Medicine, Vol. **25**, 2019, No 1, pp. 65-69.
2. Savalia, S., V. Emamian. Cardiac Arrhythmia Classification by Multi-Layer Perceptron and Convolution Neural Networks. – Bioengineering, Vol. **5**, 2018, No 2, 35.
3. Tesfai, H., H. Saleh, M. Al-Qutayri, M. B. Mohammad, T. Tekeste, A. Khandoker, B. Mohammad. Lightweight Shufflenet-Based CNN for Arrhythmia Classification. – IEEE Access, Vol. **10**, 2022, pp. 111842-111854.
4. Zhao, Y., J. Ren, B. Zhang, J. Wu, Y. Lyu. An Explainable Attention-Based TCN Heartbeats Classification Model for Arrhythmia Detection. – Biomedical Signal Processing and Control**,** Vol. **80**, 2023, 104337.
5. Singh, S., S. K. Pandey, U. Pawar, R. R. Janghel. Classification of ECG Arrhythmia Using Recurrent Neural Networks. – Procedia Computer Science, Vol. **132**, 2018, pp. 1290-1297.
6. Boda, S., M. Mahadevappa, P. K. Dutta. An Automated Patient-Specific ECG Beat Classification Using LSTM-Based Recurrent Neural Networks. – Biomedical Signal Processing and Control, Vol. **84**, 2023, 104756.
7. Kıymaç, E., Y. Kaya. A Novel Automated CNN Arrhythmia Classifier with Memory-Enhanced Artificial Hummingbird Algorithm. – Expert Systems with Applications, Vol. **213**, 2023, 119162.
8. Lea, C., R. Vidal, A. Reiter, G. D. Hager. Temporal Convolutional Networks: A Unified Approach to Action Segmentation. – In: B. Leibe, J. Matas, N. Sebe, M. Welling, Eds. Computer Vision – ECCV 2016 Workshops, Proceedings, Part III, Springer, 2016, pp. 47-54.
9. Hochreiter, S., J. Schmidhuber. Long Short-Term Memory. – Neural Computation, Vol. **9**, 1997, No 8, pp. 1735-1780.
10. Xiao, Q., K. Lee, S. A. Mokhtar, I. Ismail, A. L. B. M. Pauzi, Q. Zhang, P. Y. Lim. Deep Learning-Based ECG Arrhythmia Classification: A Systematic Review. – Applied Sciences, Vol. **13**, 2023, No 8, 4964.
11. Ansari, Y., O. Mourad, K. Qaraqe, E. Serpedin. Deep Learning for ECG Arrhythmia Detection and Classification: An Overview of Progress for the Period 2017-2023. – Frontiers in Physiology, Vol. **14**, 2023, 1246746.
12. Barot, R., H. Kapadia. Air Quality Prediction Using Long Short-Term Memory with Attention Mechanism. – Cybernetics and Information Technologies, Vol. **22**, 2022, No 1, pp. 125-139.
13. Alazzam, M., A. A. Alaboudi, A. A. Abubakar. Deep Learning Framework for IoT Network Forensics with Feature Selection and Temporal Modeling. – Cybernetics and Information Technologies, Vol. **22**, 2022, No 3, pp. 94-108.
14. Karimunnisa, S., K. R. Ramya, P. V. Krishna. Deep Learning-Driven Workload Prediction and Optimization for Load Balancing in a Cloud Computing Environment. – Cybernetics and Information Technologies, Vol. **24**, 2024, No 3, pp. 17-28.
15. ANSI/AAMI EC57:2012/(R)2020. Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms. Association for the Advancement of Medical Instrumentation, 2013.

16. Goldberger, A. L., L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. 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, Vol. **101**, 2000, No 23, pp. e215-e220.
17. Almasoud, A. S., H. A. Mengash, M. M. Eltahir, A. Alabdulkarin, A. M. Mahmoud, O. I. Khalaf. Arrhythmia Classification Using the Farmland Fertility Algorithm and Hybrid Deep Learning. – Sensors, Vol. **23**, 2023, 8265.
18. Begum, S. G., E. Priyadarshi, S. Pratap, A. Kumar, A. Alshuhail. Automated Detection of Abnormalities in ECG Signals. – Biomedical Engineering Advances, Vol. **5**, 2023, 100066.
19. Sun, A., W. Hong, J. Li, J. Mao. An Arrhythmia Classification Model Based on the CNN-LSTM-SE Algorithm. – Sensors, Vol. **24**, 2024, 6306.
20. Islam, M. R., M. Qaraqe, K. Qaraqe, S. A. Al-Maadeed, E. Serpedin. CAT-Net: Convolution, Attention, and Transformer-Based Network for ECG Arrhythmia Classification. – Biomedical Signal Processing and Control, Vol. **93**, 2024, 106211.
21. Moody, G. B., I. Jekova, A. L. Goldberger. St. Petersburg INCART 12-lead Arrhythmia Database. – PhysioNet, 2008 (online).
https://physionet.org/content/incartdb/1.0.0/
22. Prakash, A. J., 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**, 2025, 110754.
23. Omarov, B., Z. Momynkulov. Hybrid Deep Learning Model for Heart Disease Detection on 12-Lead Electrocardiograms. – In: Procedia Computer Science. Elsevier B. V., 2024, pp. 439-444.
24. Qi, M., H. Shao, N. Shi, G. Wang, Y. Lv. Arrhythmia Classification Detection Based on Multiple Electrocardiogram Databases. – PLoS One, Vol. **18**, 2023, e0290995.
25. Liu, L.-R., M.-Y. Huang, S.-T. Huang, L.-C. Kung, C.-H. Lee, W.-T. Yao, M.-F. Tsai, C.-H. Yang, C.-H. Luo, Y.-Y. Lin. An Arrhythmia Classification Approach via Deep Learning Using Single-Lead ECG without QRS Wave Detection. – Heliyon, Vol. **10**, 2024, e26874.
26. Krasteva, V., T. Stoyanov, R. Schmid, I. Jekova. Delineation of 12-Lead ECG Representative Beats Using Convolutional Encoder-Decoders with Residual and Recurrent Connections. – Sensors, Vol. **24**, 2024, No 14, 4645.
27. Yang, X., Z. Song, I. King, Z. Xu. A Survey on Deep Semi-Supervised Learning. – IEEE Transactions on Knowledge and Data Engineering, Vol. **35**, 2022, No 9, pp. 8934-8954.
28. Rasmussen, S. M., M. E. K. Jensen, C. S. Meyhoff, E. K. Aasvang, H. B. D. Sørensen. Semi-Supervised Analysis of the Electrocardiogram Using Deep Generative Models. – In: Proc. of 43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'21), IEEE, 2021, pp. 1124-1127.
29. Shi, J., W. Liu, H. Zhang, S. Chang, H. Wang, J. He, Q. Huang. CPSS: Fusing Consistency Regularization and Pseudo-Labeling Techniques for Semi-Supervised Deep Cardiovascular Disease Detection Using All Unlabeled Electrocardiograms. – Computer Methods and Programs in Biomedicine, Vol. **254**, 2024, 108315.
30. Ying, Z., G. Zhang, Z. Pan, C. Chu, X. Liu. FedECG: A Federated Semi-Supervised Learning Framework for Electrocardiogram Abnormalities Prediction. – Journal of King Saud University – Computer and Information Sciences, Vol. **35**, 2023, No 6, 101568.
31. Zhai, X., Z. Zhou, C. Tin. Semi-Supervised Learning for ECG Classification Without Patient-Specific Labeled Data. – Expert Systems with Applications, Vol. **158**, 2020, 113411.
32. Shi, J., Z. Li, W. Liu, H. Zhang, D. Luo, Y. Ge, S. Chang, H. Wang, J. He, Q. Huang. An Adaptive Threshold-Based Semi-Supervised Learning Method for Cardiovascular Disease Detection. – Information Sciences, Vol. **677**, 2024, 120881.
33. Ingolfsson, T. M., X. Wang. ECG-TCN: Wearable Cardiac Arrhythmia Detection with a Temporal Convolutional Network. – arXiv Preprint, arXiv:2103.13740, 2021.
34. Lin, C.-H. Frequency-Domain Features for ECG Beat Discrimination Using Grey Relational Analysis-Based Classifier. – Computers & Mathematics with Applications, Vol. **55**, 2008, No 4, pp. 680-690.

35. S i n g h, A. K., S. K r i s h n a n. ECG Signal Feature Extraction Trends in Methods and Applications. – BioMedical Engineering OnLine, Vol. **22**, 2023, 22.
36. K r a s t e v a, V., I. J e k o v a, T. S t o y a n o v, R. S c h m i d. In Search of an Optimal FIR Filter for ECG Delineation. – In: Computing in Cardiology, Vol. **51**, 2024.
37. C h a w l a, N. V., K. W. B o w y e r, L. O. H a l l, W. P. K e g e l m e y e r. SMOTE: Synthetic Minority Over-Sampling Technique. – Journal of Artificial Intelligence Research, Vol. **16**, 2002, pp. 321-357.
38. B i, J., Y. Z h a n g, H. Y u a n, J. X u e, J. S h i, Y. Z h o u, J. X u, Y. S o n g, L. X i n g, Y. P a n g. Accurate Arrhythmia Classification with Multi-Branch Multi-Head Attention Temporal Convolutional Networks. – Frontiers in Physiology, Vol. **14**, 2023, 1167916.
39. S h a n, L., Y. L i, H. J i a n g, P. Z h o u, J. N i u, R. L i u, Y. W e i, B. L i u, J. Q i a o, Y. L v, L. W a n g, Y. L i, B. L i u. Abnormal ECG Detection Based on an Adversarial Autoencoder. – Frontiers in Physiology, Vol. **13**, 2022, 961724.
40. A l b a r r a k, A. M., R. A l h a r b i, I. A. I b r a h i m. Detection and Classification of Unhealthy Heartbeats Using Deep Learning Techniques. – Sensors (Basel), Vol. **25**, 26 September 2025, No 19, 5976. DOI: 10.3390/s25195976. PMID: 41094799. PMCID: PMC12527069.
41. A l a m a t s a z, N., L. T a b a t a b a e i, M. Y a z d c h i, H. P a y a n, N. A l a m a t s a z, F. N a s i m i. A Lightweight Hybrid CNN-LSTM Explainable Model for ECG-Based Arrhythmia Detection. – Biomedical Signal Processing and Control, Vol. **90**, 2024, 105884.
42. L e e, D.-H. Pseudo-Label: The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks. – In: Workshop on Challenges in Representation Learning, ICML, Vol. **3**, 2013, 896.
43. O l i v e i r a, L. C., Z. L a i, H. M. S i e f k e s, C.-N. C h u a h. Generalizable Semi-Supervised Learning Strategies for Multiple Learning Tasks Using 1-d Biomedical Signals. – In: NeurIPS 2022 Workshop on Learning from Time Series for Health, 2022.
44. Z h a n g, B., Y. W a n g, W. H o u, H. W u, J. W a n g, M. O k u m u r a, T. S h i n o z a k i. Flexmatch: Boosting Semi-Supervised Learning with Curriculum Pseudo Labeling. – In: Advances in Neural Information Processing Systems, Vol. **34**, 2021, pp. 18408-18419.
45. G r a n d v a l e t, Y., Y. B e n g i o. Semi-Supervised Learning by Entropy Minimization. – In: Advances in Neural Information Processing Systems, Vol. **17**, 2004.
46. G u o, L.-Z., Y.-F. L i. Class-Imbalanced Semi-Supervised Learning with Adaptive Thresholding. – In: Proc. of International Conference on Machine Learning (PMLR'22), 2022, pp. 8082-8094.
47. S e l v a r a j u, R. R., M. C o g s w e l l, A. D a s, R. V e d a n t a m, D. P a r i k h, D. B a t r a. Grad-Cam: Visual Explanations from Deep Networks via Gradient-Based Localization. – In: Proc. of IEEE International Conference on Computer Vision, 2017, pp. 618-626.
48. S u h, J., J. K i m, S. K w o n, E. J u n g, H.-J. A h n, K.-Y. L e e, E.-K. C h o i, W. R h e e. Visual Interpretation of Deep Learning Model in ECG Classification: A Comprehensive Evaluation of Feature Attribution Methods. – In: Computers in Biology and Medicine, Vol. **182**, 2024, 109088.
49. M a l m i v u o, J., R. P l o n s e y. Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields. New York, Oxford University Press, 1995.
50. J i a n g, M., J. G u, Y. L i, B. W e i, J. Z h a n g, Z. W a n g, L. X i a. HADLN: Hybrid Attention-Based Deep Learning Network for Automated Arrhythmia Classification. – Frontiers in Physiology, Vol. **12**, 2021, 683025.
51. T a l u k d e r, M. A. A Hybrid Multiscale Feature Fusion Model for Enhanced Cardiovascular Arrhythmia Detection. – Results in Engineering, 2025, 104244.
52. M a d a n, P., V. S i n g h, D. P. S i n g h, M. D i w a k a r, B. P a n t, A. K i s h o r. A Hybrid Deep Learning Approach for ECG-Based Arrhythmia Classification. – Bioengineering, Vol. **9**, 2022, No 4, 152.
53. B h a t i a, S., S. K. P a n d e y, A. K u m a r, A. A l s h u h a i l. Classification of Electrocardiogram Signals Based on Hybrid Deep Learning Models. – Sustainability, Vol. **14**, 2022, No 24, 16572.

Fast-track. Received: 21.09.2025, First revision: 11.11.2025, Second revision: 15.11.2025, Accepted: 17.11.2025