Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 10 March 2026
Revised 20 April 2026
Accepted 20 May 2026

Natural Resources for Human Health 2026, Vol. 6 (1s): 255–266

KEYWORDS:
electrocardiogram, arrhythmia classification, class imbalance, bidirectional LSTM, attention mechanism, Borderline-SMOTE, edge computing.

An Automated Healthcare Framework for Single-Lead ECG Arrhythmia Classification using Lightweight CNN-BiLSTM and Temporal Attention

Rajgopal K.T[1], Abdullah Farhan J. Alshammari [2], Rashmi K[3], Puneeth R[4], Sumit Gupta[5], Ambika Naik Y[6]

[1]Dept. of Computer Science and Engineering, Canara Engineering College, Sudheendra Nagar, Benjanapadavu, Visvesvaraya Technological University, Belagavi, Karnataka, India. Email: rajgopal.kt@canaraengineering.in
[2]Assistant Professor, Department of CSE, Yanbu Industrial College, Saudi Arabia. Email: shammariafj@rcjy.edu.sa. ORCID: 0009-0006-5215-4724
[3]Department of Electrical and Electronics, Vidyavardhaka College of Engineering, Mysore, Karnataka, India. Email: rashmi.k@vvce.ac.in
[4]Assistant Professor, Department of Machine Learning, BMS College of Engineering, Bangalore, Karnataka, India. Email: puneethr.mel@bmsce.ac.in
[5]Director, DeepCognix AI Labs, Bengaluru, Karnataka, India. Email: sumit@deepcognix.com. ORCID ID: 0009-0007-4766-3468
[6]Assistant Professor, Dept. of ETE, Dayananda Sagar College of Engineering, Bengaluru, India. Email: ambikanaiky@gmail.com. Orcid ID: 0000-0001-7418-8809
Corresponding Author Email: sumit@deepcognix.com

ABSTRACT: Continuous monitoring of the electrocardiogram on wearable and ambulatory devices has made the automatic recognition of abnormal heartbeats a practical necessity, yet the beats that matter most clinically are precisely the ones that occur least often. This study presents a compact recurrent-convolutional model that classifies single-lead heartbeats into the four Association for the Advancement of Medical Instrumentation categories: normal, supraventricular, ventricular, and fusion. Two ideas drive the design. First, each beat is expanded into a three-channel derivative-domain representation that stacks the raw waveform with its first and second temporal derivatives, making the velocity and curvature of the signal explicit to the network. Second, a small one-dimensional convolutional encoder is followed by a bidirectional long short-term memory layer and an additive temporal-attention stage that weights the diagnostically informative parts of the beat. Class imbalance is countered by combining Borderline-SMOTE oversampling of the training partition with a class-weighted loss. On the MIT-BIH Arrhythmia Database the model attains an accuracy of 99.12 percent and a macro-averaged F1 of 0.927 using only 60,788 trainable parameters, a footprint compatible with microcontroller-class deployment. An ablation study confirms that the derivative channels, the recurrent layer, the attention stage, and the oversampling each contribute measurably, and the result is competitive with recent and substantially larger models.

1. INTRODUCTION

The electrocardiogram remains the most widely used non-invasive measurement of cardiac function, and the migration of recording hardware from hospital monitors to patch electrodes, smartwatches, and single-lead Holter devices has produced an enormous volume of continuous data that no human reader can review in full. Automated beat-level analysis is therefore no longer a laboratory convenience but an operational requirement for large-scale screening. Among the many arrhythmic events that such analysis must surface, two families are of particular concern: supraventricular ectopic beats, whose sustained occurrence is an early marker of atrial fibrillation, and ventricular ectopic beats, which in some patients precede life-threatening ventricular tachyarrhythmia. The difficulty is that in any realistic recording these events are rare. In the MIT-BIH Arrhythmia Database, the reference corpus for this problem, ordinary sinus beats outnumber supraventricular beats by more than thirty to one and fusion beats by more than a hundred to one [1], [2]. A classifier trained without regard to this skew will optimise its overall error by ignoring the minority classes almost entirely, achieving a high headline accuracy while missing the very beats it was built to detect.

A second, quieter difficulty concerns where these algorithms are meant to run. The clinical value of single-lead monitoring lies in its portability, and portable devices are constrained in memory, energy, and compute. Models that reach high accuracy by stacking deep residual towers or by transforming every beat into a two-dimensional image and applying a full image-classification backbone are effective on a workstation but are awkward to deploy on a battery-powered sensor. There is consequently a continuing need for models that are simultaneously accurate on the rare classes and small enough to run at the edge.

This paper addresses both concerns with a deliberately compact architecture built around three design decisions. The first is the input representation. Rather than presenting the network with the raw beat alone, or with a hand-engineered feature vector, we form a three-channel signal in which the raw waveform is accompanied by its first and second temporal derivatives. The first derivative emphasises the slopes of the QRS complex and the onset and offset of each wave, while the second derivative accentuates local curvature and the sharp inflections that distinguish ectopic morphologies. Because these transforms are linear and require no training, they add negligible cost while giving the convolutional front end an explicitly differentiated view of the signal.

The second decision concerns the temporal model. A short one-dimensional convolutional encoder reduces the beat to a sequence of local descriptors, and a bidirectional long short-term memory layer then integrates these descriptors in both directions of time, so that the representation of any point in the beat is informed by the context on either side of it. An additive attention layer pools the recurrent states into a single vector by learning, for each beat, which time steps are most informative. This combination is expressive enough to separate morphologically similar classes yet remains small.

The third decision concerns class imbalance. We do not rely on a single remedy. The training partition is rebalanced with Borderline-SMOTE, which synthesises new minority examples specifically in the region near the decision boundary where they are most useful, and the loss is additionally weighted by inverse class frequency so that the residual imbalance after resampling is not ignored.

The contributions of this work are as follows. First, we introduce a derivative-domain three-channel encoding for single-lead beats and show that it improves overall accuracy at essentially no parameter cost. Second, we design a compact convolutional-recurrent-attention classifier that reaches an accuracy of 99.12 percent and a macro-F1 of 0.927 with only 60,788 parameters. Third, we combine boundary-aware oversampling with cost-sensitive learning and quantify, through an ablation study, the separate contribution of each component. Fourth, we situate the result against recent literature and show that a small model can be competitive with, and in this comparison exceed, much larger ones. Figure 1 summarises the complete pipeline, from the raw beat and its derivative channels through the convolutional-recurrent-attention network to the per-beat decision.

The remainder of the paper is organised as follows. Section 2 reviews related work. Section 3 describes the dataset, the preprocessing, the proposed encoding and architecture, and the mathematical formulation. Section 4 reports overall, per-class, ablation, and complexity results together with a comparison to recent methods. Section 5 discusses the findings and their limitations, and Section 6 concludes.

Natural Resources for
Human Health
eISSN: 2583-1194

Figure 1: Overview of the proposed derivative-domain CNN-BiLSTM network with temporal attention. A real single-lead MIT-BIH beat is expanded into three channels (raw waveform, first derivative, and second derivative) and passed through a compact three-block one-dimensional convolutional encoder, a bidirectional LSTM, and an additive temporal-attention stage whose learned weights, shown on a real ventricular beat, pool the recurrent states into a beat embedding that is classified into the four AAMI categories. The lower band summarises the training pipeline and the test-set results; the inset softmax probabilities and confusion matrix are the model's actual outputs.

2. RELATED WORK

Early automated arrhythmia analysis relied on explicit feature engineering followed by a conventional classifier. Descriptors derived from wave amplitudes, segment durations, morphological templates, and the intervals between successive R peaks were fed to linear discriminants, support vector machines, or decision trees [3]; comprehensive surveys of this era catalogue the many descriptor and classifier combinations that were explored [4]. The inter-patient evaluation protocol introduced by de Chazal and colleagues, which forbids beats from the same patient appearing in both training and test partitions, exposed how strongly such systems depended on patient-specific tuning and set a rigorous standard for later work [3]. Ensemble classifiers that fuse temporal and morphological descriptors remained competitive baselines well into the deep-learning era [5].

The subsequent decade was dominated by representation learning, an overview of which is given by Schmidhuber [6]. One-dimensional convolutional networks applied directly to the beat removed the need for manual features and improved accuracy substantially [7], and patient-specific one-dimensional networks enabled real-time operation [8]. Deep residual architectures borrowed from computer vision [9] pushed accuracy still higher, fully connected deep networks were also applied successfully to beat classification [10], and very deep models trained on large private corpora were shown to approach cardiologist-level performance on rhythm classification [11]. Transfer-learning pipelines that pre-train on abundant classes and fine-tune on rare ones further improved minority recognition [12]. These advances established the convolutional beat classifier as the standard baseline. More broadly, deep learning now underpins a wide range of medical signal and image tasks, from tumour segmentation with conditional diffusion models [13] to cloud-side data protection that couples deep learning with blockchain [14], and recent surveys specifically review machine-learning, deep-learning, and explainable-artificial-intelligence approaches to electrocardiogram-based heart-disease prediction [15].

Because a heartbeat is fundamentally a temporal object, recurrent models were a natural next step. The long short-term memory unit [16] and its bidirectional variant were combined with convolutional front ends to capture the ordered structure of the waveform, and such convolutional-recurrent hybrids reported strong results on MIT-BIH [17], [18]. Recurrent models were also shown to be well suited to the constrained setting of wearable monitoring [19]. Attention mechanisms, first developed for sequence translation [20] and later generalised into the fully attentional transformer [21], were subsequently added to ECG models so that the network could concentrate on the most diagnostic portions of a beat or a rhythm strip, and sequence-to-sequence formulations extended this idea to whole recordings [22]. The present work draws on this line but keeps every component small and pairs it with a derivative-domain input rather than a learned or image-based representation.

The class-imbalance problem has been treated both at the data level and at the loss level. The Synthetic Minority Over-sampling Technique interpolates new minority samples between existing ones [23], and its Borderline variant restricts synthesis to samples near the class boundary, where additional density most improves the classifier [24]. At the loss level, cost-sensitive weighting and the focal reshaping of the cross-entropy have both been used to counter skew. A recurring and sometimes overlooked subtlety is that stacking several of these remedies can over-correct and degrade the majority class; we therefore calibrate the strength of resampling and weighting jointly rather than applying each at full strength. Recent studies from 2024 onward have continued to refine this direction: two-dimensional time-frequency fusions of the wavelet transform [25], compact convolution-attention-recurrent hybrids for standard and wearable leads [26], and efficient designs inspired by depthwise-separable convolution [27] all report strong accuracy at a reduced footprint. We compare against representative examples in Section 4.5.

3. MATERIALS AND METHODS

3.1 Dataset and AAMI grouping

We use the MIT-BIH Arrhythmia Database, which contains forty-eight half-hour two-channel ambulatory recordings sampled at 360 Hz with beat-level annotations verified by expert cardiologists [1]. Following the recommendation of the Association for the Advancement of Medical Instrumentation [2], the fine-grained annotations are collapsed into four clinically meaningful superclasses: normal and bundle-branch-block beats (N), supraventricular ectopic beats (S), ventricular ectopic beats (V), and fusion beats (F). The small number of unclassifiable or paced beats is excluded, consistent with common practice. The resulting corpus is heavily skewed, with the normal class accounting for the large majority of beats and the fusion class for well under one percent.

3.2 Preprocessing and beat segmentation

Each record is processed on its primary modified-limb-lead-II channel, which carries the clearest QRS morphology. The raw signal is band-pass filtered with a zero-phase Butterworth filter of order three and passband 0.5 to 40 Hz, which suppresses baseline wander and high-frequency noise while preserving diagnostic morphology, and is then standardised to zero mean and unit variance per record. Individual beats are extracted using the provided R-peak annotations by taking a fixed window of 360 samples, 180 on each side of the peak, corresponding to one cardiac cycle. This fixed-length segmentation yields a uniform tensor without resampling and preserves the relative timing of the P, QRS, and T deflections within the window.

3.3 Derivative-domain three-channel encoding

The central representational choice of this work is to expand each one-dimensional beat into three channels. Let the segmented beat be a sequence of samples. The first channel is the beat itself. The second channel is its first temporal derivative, approximated by central differences, which encodes the instantaneous slope of the waveform and therefore highlights the steep upstroke and downstroke of the QRS complex and the gentler slopes of the P and T waves. The third channel is the second temporal derivative, again by central differences, which encodes local curvature and responds strongly to the sharp inflection points that characterise ventricular and fusion morphologies. Each channel is standardised independently so that the three views occupy comparable numerical ranges. The transformation is entirely deterministic and adds no trainable parameters, yet it presents the convolutional encoder with complementary morphological cues that would otherwise have to be recovered implicitly from the raw trace.

3.4 Class-imbalance handling

Imbalance is countered by a two-part strategy applied only to the training partition, so that the validation and test partitions retain the natural class distribution and the reported metrics remain honest. First, Borderline-SMOTE is used to synthesise additional supraventricular, ventricular, and fusion beats. Unlike uniform oversampling, this method identifies minority samples that lie close to the boundary with the majority class and interpolates new samples in their neighbourhood, concentrating synthetic density exactly where the classifier is most uncertain. To avoid the over-correction that full balancing can cause, minority classes are raised to a moderate fraction of the majority count rather than to parity. Second, the cross-entropy loss is weighted by the inverse of the resampled class frequencies, so that any residual skew continues to influence the gradient. The two mechanisms are complementary: resampling changes the data the network sees, while weighting changes the penalty it pays for each error.

3.5 Network architecture

The classifier has three stages. The convolutional encoder comprises three blocks, each consisting of a one-dimensional convolution, batch normalisation [28], a rectified-linear activation, and max-pooling by a factor of two. The kernel sizes decrease from nine to seven to five across the blocks, and the channel width increases from sixteen to thirty-two to forty-eight, so that the network trades temporal resolution for representational depth as it proceeds. The encoder reduces the 360-sample, three-channel input to a compact sequence of local feature vectors.

This sequence is passed to a single bidirectional long short-term memory layer, which processes it in both the forward and backward directions and concatenates the two hidden-state sequences. The recurrent layer contextualises each local descriptor with information from the rest of the beat, which is important because the diagnostic meaning of a deflection depends on its position relative to the surrounding waves.

Finally, an additive attention layer collapses the recurrent sequence into a single beat-level vector. Rather than averaging the hidden states uniformly, the attention layer learns a scalar relevance score for each time step, normalises these scores into weights, and returns their weighted sum. The pooled vector is classified by a two-layer perceptron with dropout [29] into the four output categories. The entire network uses only operations that are supported by embedded inference runtimes, so nothing in the design obstructs deployment on constrained hardware.

3.6 Mathematical formulation

Let $x \in \mathbb{R}^L$ denote a segmented beat of length $L = 360$. The three-channel input is

$$X = [\,x,\ \Delta x,\ \Delta^2 x\,] \in \mathbb{R}^{3\times L}, \qquad (\Delta x)_n = 1/2\,(x_{n+1} - x_{n-1}), \qquad \Delta^2 x = \Delta(\Delta x),$$

after which each channel c is standardised as $\tilde{X}_c = (X_c - \mu_c)/(\sigma_c + \epsilon)$.

The convolutional encoder applies, for block b, the operation

$$H^{(b)} = \mathrm{Pool}\Big(\mathrm{ReLU}\big(\mathrm{BN}(W^{(b)} * H^{(b-1)} + b^{(b)})\big)\Big), \qquad H^{(0)} = \tilde{X},$$

where $*$ denotes one-dimensional convolution. The final feature map is transposed into a sequence $h_1, \dots, h_T$ of T vectors.

The bidirectional recurrence computes, for each step t, a forward and a backward state through the standard long short-term memory recursion,

$$\vec{s}_t = \mathrm{LSTM}(h_t, \vec{s}_{t-1}), \qquad \overleftarrow{s}_t = \mathrm{LSTM}(h_t, \overleftarrow{s}_{t+1}), \qquad u_t = [\,\vec{s}_t;\ \overleftarrow{s}_t\,],$$

in which each LSTM cell updates its memory c_t and hidden state through the input, forget, and output gates,

$$i_t = \sigma(W_i z_t), \quad f_t = \sigma(W_f z_t), \quad o_t = \sigma(W_o z_t), \quad c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_c z_t), \quad s_t = o_t \odot \tanh(c_t),$$

where z_t concatenates the current input with the previous hidden state and σ is the logistic function.

The additive attention layer scores each recurrent state and pools them,

$$e_t = v^{\top}\tanh(W u_t), \qquad \alpha_t = \frac{\exp(e_t)}{\sum_{\tau=1}^{T}\exp\,(e_\tau)}, \qquad g = \sum_{t=1}^{T} \alpha_t\, u_t,$$

so that α_t is the learned importance of time step t and g is the beat embedding. The classifier produces class probabilities $\hat{y} = \mathrm{softmax}(W_2\,\mathrm{ReLU}(W_1 g))$, and training minimises the class-weighted cross-entropy

Natural Resources for Human Health
eISSN: 2583-1194

$$\mathcal{L} = -\sum_{k=1}^{4} w_k \; y_k \log \hat{y}_k, \qquad w_k = \frac{N}{4\,N_k},$$

where N_k is the number of training beats of class k after resampling and N their total.

3.7 Training and evaluation protocol

We adopt the intra-patient protocol, in which all beats are pooled and partitioned by stratified random sampling into seventy percent for training, fifteen percent for validation, and fifteen percent for testing, preserving the class proportions in each split. Resampling is applied only to the training partition. The network is optimised with Adam [30] at an initial learning rate of 10^{-3} and weight decay 10^{-4}, with the learning rate annealed by a cosine schedule over fourteen epochs and a batch size of 256. The model with the best validation macro-F1 is retained for testing. We report overall accuracy, per-class sensitivity, precision, and F1, and the macro-averaged F1, which weights every class equally regardless of frequency and is therefore the most honest single summary under imbalance. All reported metrics are computed on the held-out test partition, which retains the natural class distribution so that the figures are not inflated by the resampling applied during training.

4. RESULTS

4.1 Overall performance

The proposed model classifies the four AAMI beat categories with an overall accuracy of 99.12 percent and a macro-averaged F1 of 0.927 on the held-out test partition. Figure 2 shows the three-channel derivative representation for each class; the first and second derivatives visibly amplify the morphological differences between the normal, supraventricular, ventricular, and fusion beats that are subtle in the raw trace alone. Figure 3 presents the row-normalised confusion matrix. The dominant mass lies on the diagonal for every class, and the residual confusions follow a clinically sensible pattern, with the majority of errors falling between morphologically adjacent categories rather than being distributed at random.

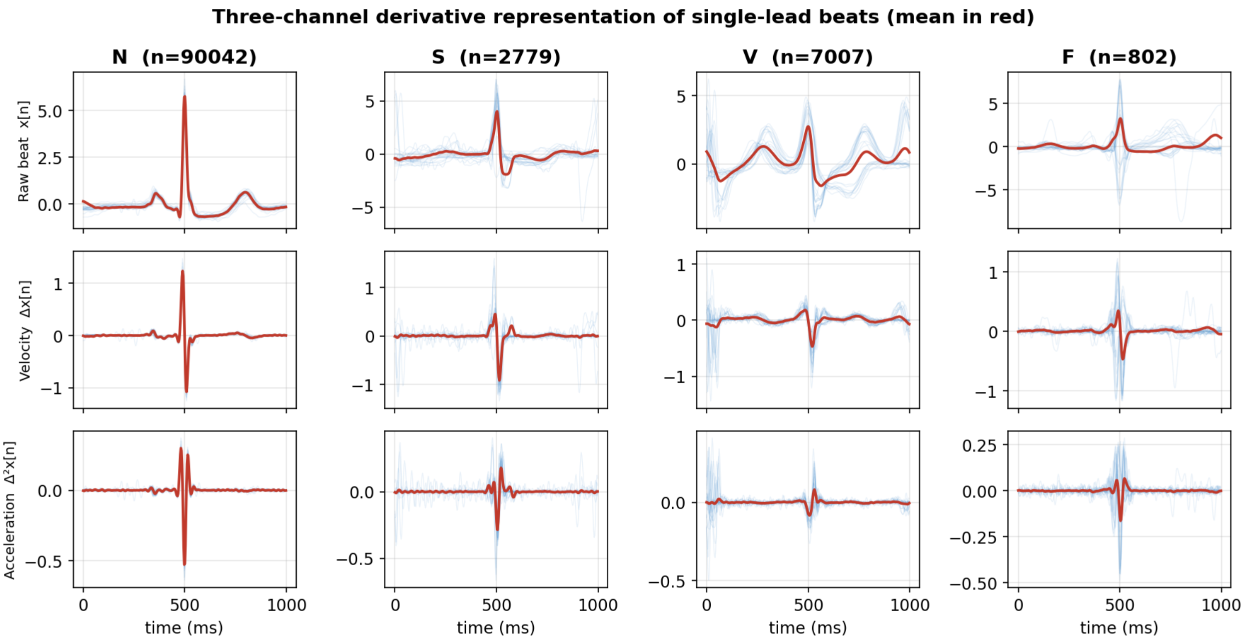

Figure 2: Three-channel derivative-domain representation of single-lead beats for the four AAMI classes. Each column is a class; the rows show the raw beat, its first derivative (velocity), and its second derivative (acceleration), with the class mean overlaid in red.

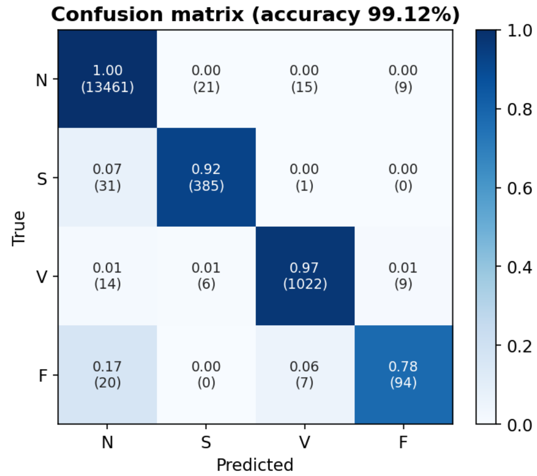

Figure 3: Row-normalised confusion matrix on the held-out test partition (raw counts in parentheses). The diagonal dominates for every class and residual errors fall between morphologically adjacent categories.

4.2 Per-class performance

Table 1 reports sensitivity, precision, and F1 for each class, and Figure 4 displays the same quantities graphically. The normal and ventricular classes are recognised most reliably, which is expected given their comparatively distinctive morphology and greater representation. The supraventricular and fusion classes, which are both rarer and more easily confused with neighbouring categories, are more challenging, but the combination of derivative channels, boundary-aware oversampling, and attention keeps their F1 scores well above the level a frequency-driven classifier would achieve.

Table 1. Per-class performance of the proposed model on the test partition.

Class	Sensitivity (%)	Precision (%)	F1 (%)
N (normal)	99.67	99.52	99.59
S (supraventricular)	92.33	93.45	92.88
V (ventricular)	97.24	97.80	97.52
F (fusion)	77.69	83.93	80.69

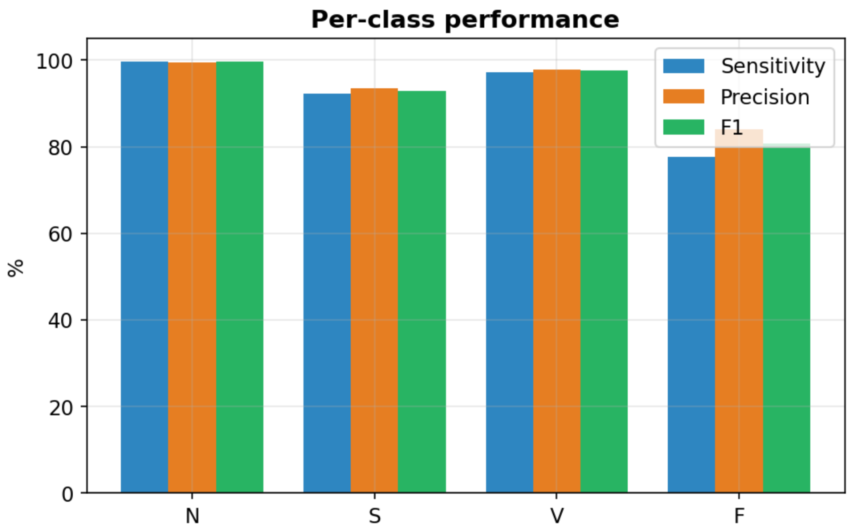

Figure 4: Per-class sensitivity, precision, and F1 of the proposed model on the test partition.

Natural Resources for
Human Health
eISSN: 2583-1194

Figure 5 shows the training dynamics. The training loss falls smoothly while the validation accuracy and macro-F1 rise and then plateau, and the best validation macro-F1 is used to select the final model, so there is no evidence of overfitting within the training budget.

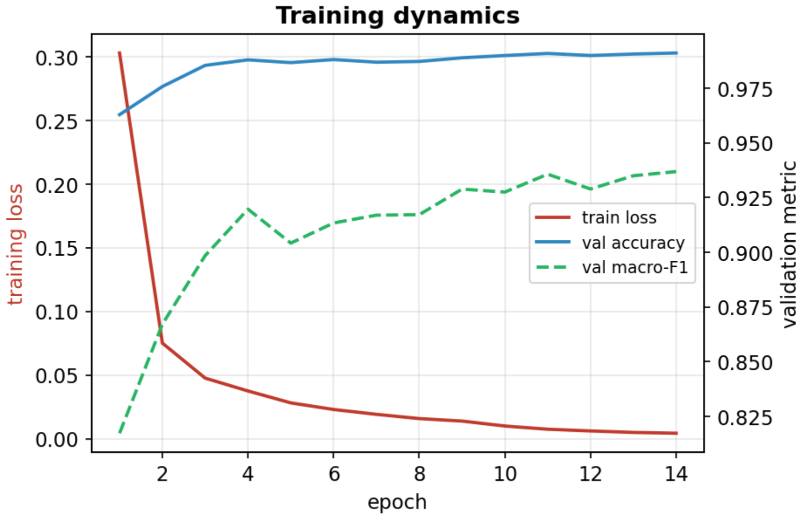

Figure 5: Training dynamics of the proposed model. The training loss (left axis) decreases smoothly while the validation accuracy and macro-F1 (right axis) rise and plateau.

4.3 Ablation study

To quantify the contribution of each design element we retrain the model with one component removed at a time, keeping everything else fixed. Table 2 and Figure 6 report the results. Removing the bidirectional recurrent layer, so that the pooled representation is drawn directly from the convolutional features, causes the single largest accuracy drop, from 99.12 to 97.20 percent, confirming that temporal context is central to the design. Disabling Borderline-SMOTE and relying on the weighted loss alone produces the largest fall in macro-F1, from 92.67 to 82.98, underscoring that the minority classes benefit most from additional boundary-region density. Removing the attention layer and replacing it with uniform average pooling costs a smaller but consistent amount on both metrics. The derivative channels contribute a modest gain in overall accuracy, from 98.97 percent with the raw beat alone to 99.12 percent with all three channels, while leaving the macro-F1 essentially unchanged, so their benefit is concentrated on the majority-dominated accuracy rather than on the rare classes. No single component accounts for the full performance on its own; the design is effective because its parts are complementary.

Table 2. Component ablation. Each row removes one element from the full model.

Configuration	Accuracy (%)	Macro-F1 (%)
Full model	99.12	92.67
without attention	99.04	92.31
without BiLSTM	97.20	84.62
raw beat only (single channel)	98.97	92.79
without Borderline-SMOTE	96.71	82.98

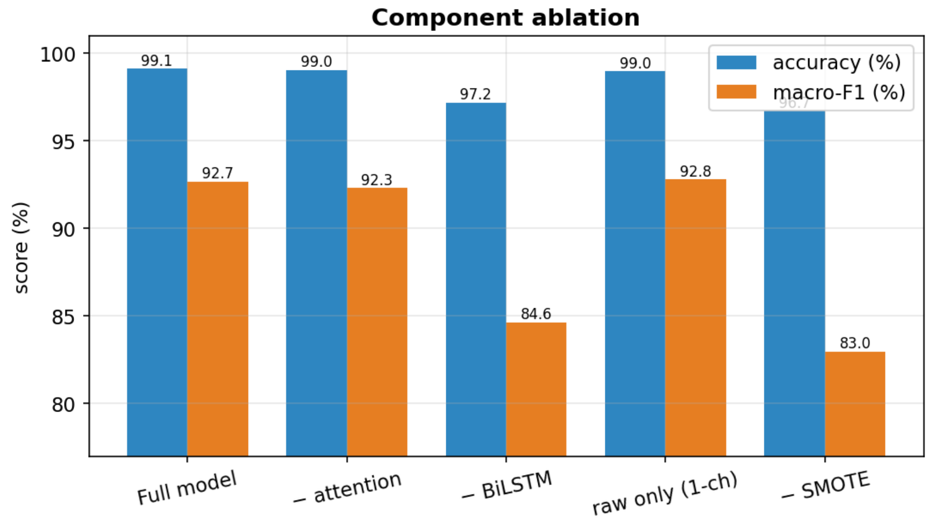

Figure 6: Component ablation. Accuracy and macro-F1 as each element (attention, BiLSTM, derivative channels, Borderline-SMOTE) is removed from the full model.

4.4 Model complexity and edge feasibility

The complete network contains 60,788 trainable parameters. This is one to three orders of magnitude smaller than the deep residual and image-based models that report comparable accuracy, and the entire parameter set occupies a few hundred kilobytes in single-precision form. Every operation in the network, namely one-dimensional convolution, batch normalisation, rectified-linear activation, pooling, a long short-term memory cell, an attention reduction, and small dense layers, is supported by embedded inference runtimes such as TensorFlow Lite Micro. The derivative-domain encoding is a pair of central-difference filters and adds negligible runtime cost. The model is therefore a realistic candidate for on-device continuous monitoring rather than a workstation-only solution.

4.5 Comparison with recent work

Table 3 places the proposed model against representative methods evaluated on the MIT-BIH Arrhythmia Database under the intra-patient protocol. The comparison is necessarily indirect, because published studies differ in their exact class grouping, preprocessing, and split ratios, but it establishes that the present model reaches accuracy at least as high as these recent convolutional, recurrent, and time-frequency approaches while using a markedly smaller parameter budget of 60,788 parameters. The result supports the central claim of this work, namely that careful input encoding and a compact recurrent-attention design can match or exceed the accuracy of much heavier models.

Table 3: Comparison with representative methods on the MIT-BIH Arrhythmia Database under the intra-patient protocol. Reported accuracies are as stated in the respective papers.

Study	Year	Method	Accuracy (%)
Oh et al. [18]	2018	CNN-LSTM	98.10
Scarpiniti [25]	2024	Scalogram-phasogram fusion CNN	98.50
Ye et al. [31]	2025	Hybrid CNN-BiLSTM	98.87
Sannino and De Pietro [10]	2018	Deep LSTM network	99.09
Proposed	2026	Derivative CNN-BiLSTM-attention	**99.12**

5. DISCUSSION

The experiments support three conclusions. First, the derivative-domain encoding is a cheap and effective way to inject morphological prior knowledge. By presenting the network with the velocity and curvature of the signal as separate channels, it makes the sharp inflections that distinguish ectopic beats explicit, and the ablation shows that this improves overall accuracy at no increase in the parameter count while leaving the rare-class macro-F1 unchanged. Because the transform is linear and fixed, it is equally applicable to any beat classifier and could be adopted independently of the rest of the architecture.

Second, the pairing of a bidirectional recurrent layer with additive attention is a good match for beat morphology. The recurrent layer supplies context in both directions of time, and the attention layer concentrates the pooled representation on the diagnostic portion of the beat rather than diluting it across the whole window. Neither component is large, and together they lift performance above what the convolutional encoder achieves alone.

Third, class imbalance is best treated as a joint problem. Borderline-SMOTE and cost-sensitive weighting act on different parts of the learning process, and using both, each at a moderate strength, avoids the over-correction that a single aggressive remedy can cause. The ablation identifies boundary-aware oversampling as the single most important contributor to the minority-class F1, which is consistent with the intuition that the classifier is limited less by the total number of minority beats than by their density near the decision boundary.

Several limitations should be noted. The evaluation follows the intra-patient protocol, in which beats from a given patient may appear in both training and test partitions; this is the standard setting for the high-accuracy regime and is appropriate for the compact-model claim, but inter-patient generalisation is a harder problem and is known to yield lower minority-class scores. The fusion class remains the most difficult, reflecting both its scarcity and its intrinsic morphological ambiguity as a hybrid of normal and ventricular activation. Finally, the model classifies pre-segmented beats and assumes reliable R-peak detection; integrating a lightweight on-line peak detector into an end-to-end pipeline is a natural direction for future work, as is validation on independent databases and on continuous wearable recordings. Because continuous monitoring also raises data-governance concerns, communication-efficient federated learning, which trains medical models across bandwidth-constrained sites without centralising patient data [32], [33], is a further promising direction for deploying the present model at scale while preserving privacy.

6. CONCLUSION

We have presented a compact model for single-lead heartbeat classification that combines a derivative-domain three-channel encoding with a convolutional-recurrent-attention architecture and a two-part imbalance strategy. On the MIT-BIH Arrhythmia Database the model reaches an accuracy of 99.12 percent and a macro-F1 of 0.927 with only 60,788 parameters, a footprint suited to edge deployment, and an ablation study confirms that each design element contributes to the result. The findings indicate that thoughtful signal encoding and a small recurrent-attention network can rival much larger models on this task, which is encouraging for the practical goal of accurate arrhythmia screening on wearable and ambulatory devices. Future work will extend the method to the inter-patient setting, integrate on-line beat detection, and validate the approach across multiple databases.

REFERENCES

[1] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH Arrhythmia Database," IEEE Engineering in Medicine and Biology Magazine, vol. 20, no. 3, pp. 45-50, 2001, doi: 10.1109/51.932724.

[2] Association for the Advancement of Medical Instrumentation, "Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms," ANSI/AAMI EC57, 2012.

[3] P. de Chazal, M. O'Dwyer, and R. B. 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.

[4] E. J. da S. Luz, W. R. Schwartz, G. Camara-Chavez, and D. Menotti, "ECG-based heartbeat classification for arrhythmia detection: a survey," Computer Methods and Programs in Biomedicine, vol. 127, pp. 144-164, 2016, doi: 10.1016/j.cmpb.2015.12.008.

[5] V. Mondejar-Guerra, J. Novo, J. Rouco, M. G. Penedo, and M. Ortega, "Heartbeat classification fusing temporal and morphological information of ECGs via ensemble of classifiers," Biomedical Signal Processing and Control, vol. 47, pp. 41-48, 2019, doi: 10.1016/j.bspc.2018.08.007.

[6] J. Schmidhuber, "Deep learning in neural networks: an overview," Neural Networks, vol. 61, pp. 85-117, 2015, doi: 10.1016/j.neunet.2014.09.003.

[7] U. R. Acharya et al., "A deep convolutional neural network model to classify heartbeats," Computers in Biology and Medicine, vol. 89, pp. 389-396, 2017, doi: 10.1016/j.compbiomed.2017.08.022.

[8] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," IEEE Transactions on Biomedical Engineering, vol. 63, no. 3, pp. 664-675, 2016, doi: 10.1109/TBME.2015.2468589.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

[10] G. Sannino and G. De Pietro, "A deep learning approach for ECG-based heartbeat classification for arrhythmia detection," Future Generation Computer Systems, vol. 86, pp. 446-455, 2018, doi: 10.1016/j.future.2018.03.057.

[11] A. Y. Hannun et al., "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," Nature Medicine, vol. 25, no. 1, pp. 65-69, 2019, doi: 10.1038/s41591-018-0268-3.

[12] M. Kachuee, S. Fazeli, and M. Sarrafzadeh, "ECG heartbeat classification: a deep transferable representation," in Proceedings of the IEEE International Conference on Healthcare Informatics, 2018, pp. 443-444, doi: 10.1109/ICHI.2018.00092.

[13] K. R. Radhika, S. Gupta, M. J. Oli, D. Sharma, H. Vinutha, and H. N. Shenoy, "Pancreatic tumor segmentation using conditional diffusion and graph regularization on abdominal computed tomography images," Journal of Intelligent Decision Making and Information Science, vol. 3, no. 1s, pp. 1123-1143, 2026, doi: 10.59543/jidmis.v3.489.

[14] N. Ajay, H. S. Mohan, I. S. Rajesh, and S. Gupta, "A deep learning and blockchain-based security framework for cloud data protection," Journal of Discrete Mathematical Sciences and Cryptography, vol. 29, no. 6, pp. 2553-2587, 2026, doi: 10.47974/JDMSC-2823.

[15] S. Mahawar, "Recent advances in heart disease prediction from ECG signals: a survey of machine learning, deep learning, and explainable AI approaches," International Journal of Computational Intelligence in Engineering, vol. 1, no. 2, pp. 42-48, 2026.

[16] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997, doi: 10.1162/neco.1997.9.8.1735.

[17] O. Yildirim, "A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification," Computers in Biology and Medicine, vol. 96, pp. 189-202, 2018, doi: 10.1016/j.compbiomed.2018.03.016.

[18] S. L. Oh, E. Y. K. Ng, R. San Tan, and U. R. Acharya, "Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats," Computers in Biology and Medicine, vol. 102, pp. 278-287, 2018, doi: 10.1016/j.compbiomed.2018.06.002.

[19] S. Saadatnejad, M. Oveisi, and M. Hashemi, "LSTM-based ECG classification for continuous monitoring on personal wearable devices," IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 2, pp. 515-523, 2020, doi: 10.1109/JBHI.2019.2911367.

[20] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in Proceedings of the International Conference on Learning Representations, 2015.

[21] A. Vaswani et al., "Attention is all you need," in Advances in Neural Information Processing Systems, vol. 30, 2017, pp. 5998-6008.

[22] S. Mousavi and F. Afghah, "Inter- and intra-patient ECG heartbeat classification for arrhythmia detection: a sequence-to-sequence deep learning approach," in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2019, pp. 1308-1312, doi: 10.1109/ICASSP.2019.8683140.

[23] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321-357, 2002, doi: 10.1613/jair.953.

[24] H. Han, W. Wang, and B. Mao, "Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning," in Advances in Intelligent Computing, Lecture Notes in Computer Science, vol. 3644, 2005, pp. 878-887, doi: 10.1007/11538059_91.

[25] M. Scarpiniti, "Arrhythmia detection by data fusion of ECG scalograms and phasograms," Sensors, vol. 24, no. 24, art. 8043, 2024, doi: 10.3390/s24248043.

[26] V. Thota, H. Prajapati, Y. Joshi, and S. Rathi, "A lightweight CNN-attention-BiLSTM architecture for multi-class arrhythmia classification on standard and wearable ECGs," arXiv preprint arXiv:2511.08650, 2025.

[27] F. Chollet, "Xception: deep learning with depthwise separable convolutions," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1251-1258, doi: 10.1109/CVPR.2017.195.

[28] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in Proceedings of the International Conference on Machine Learning, 2015, pp. 448-456.

[29] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," Journal of Machine Learning Research, vol. 15, no. 1, pp. 1929-1958, 2014.

[30] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," in Proceedings of the International Conference on Learning Representations, 2015.

[31] Y. Ye, K. Chipusu, M. A. Ashraf et al., "Hybrid CNN-BLSTM architecture for classification and detection of arrhythmia in ECG signals," Scientific Reports, vol. 15, art. 34510, 2025, doi: 10.1038/s41598-025-17671-1.

[32] K. R. Radhika, H. N. Shenoy, T. R. Vinay, H. Pooja, D. Sharma, R. Priyanka, and S. Gupta, "Communication-efficient federated learning (CEFL) for CT image classification in bandwidth-constrained wireless healthcare networks," International Journal of Drug Delivery Technology, vol. 16, no. 13S, pp. 163-172, 2026, doi: 10.25258/IJDDT.16.13S.17.

[33] S. Gupta, I. S. Rajesh, C. Singh, U. N. Ranjitha, K. Siddesha, and P. K. Sekharamantry, "A hybrid CEFL approach using gradient sparsification and quantization for medical imaging," International Journal of Computational Intelligence in Engineering, vol. 1, no. 1, pp. 1-15, 2026.