International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(6s) 2026 pp. 639-648
https://doi.org/10.70917/ijcisim-2026-2977
Article

Time–Frequency Deep Learning for High-Accuracy Arrhythmia Classification: A Wavelet-Scalogram Squeeze-and-Excitation Residual Network

Madhumita Mishra[1,*], Ashwinkumar U M[2]

[1,2]School of Computer Science and Engineering, REVA University, Bengaluru, India
Corresponding author: Madhumita Mishra - email: madhumita.mishra@reva.edu.in

Abstract: Continuous cardiac monitoring on wearable and embedded devices demands heartbeat classifiers that are simultaneously accurate and lightweight. We present WSRNet, a wavelet-scalogram squeeze-and-excitation residual network that classifies single-lead electrocardiogram (ECG) beats from a two-dimensional time–frequency image. Each beat is transformed by the continuous wavelet transform, using a Morlet mother wavelet, into a scalogram that makes cardiac morphology and spectral content jointly explicit; a compact two-dimensional residual convolutional network with squeeze-and-excitation channel attention then performs the classification. Evaluated under the intra-patient protocol on the MIT-BIH Arrhythmia Database over the four principal AAMI classes {N, S, V, F}, WSRNet attains 99.13 % overall accuracy and a macro-F1 of 0.932, detecting the rare supraventricular and fusion classes at F1 scores of 0.926 and 0.829 respectively despite a fifty-to-one class imbalance. An ablation study shows the model is robust to component choices, every variant exceeds 98 %, and that, in this intra-patient regime, a plain cross-entropy loss suffices (class re-weighting, essential inter-patient, is here counterproductive) and a size-matched raw-1-D network is competitive, so the scalogram is an effective though not exclusive route to high accuracy. The model uses only 716 K parameters (about 700 KB after 8-bit quantisation) and classifies a beat in a few milliseconds on a single CPU thread, comfortably within real-time budgets. It thus occupies the same lightweight regime as recent edge-oriented systems while matching the accuracy of far larger transformer- and wavelet-based models, making it a practical candidate for on-device arrhythmia screening.
Keywords: electrocardiography, arrhythmia classification, continuous wavelet transform, scalogram, time–frequency analysis, convolutional neural network, squeeze-and-excitation, MIT-BIH.

1. Introduction

The electrocardiogram (ECG) is the most widely used non-invasive tool for assessing cardiac rhythm, and the automatic classification of individual heartbeats underpins both long-term Holter analysis and the emerging generation of wearable cardiac monitors. Cardiac arrhythmias, irregularities in the timing or morphology of heartbeats, are a leading contributor to cardiovascular morbidity, and their reliable detection from continuous ECG is of substantial clinical value [1], [2]. The Association for the Advancement of Medical Instrumentation (AAMI) groups the fine-grained beat annotations of a recording into five clinically meaningful super-classes, normal (N), supraventricular ectopic (SVEB, S), ventricular ectopic (VEB, V), fusion (F), and unknown (Q), providing a standard target for automatic classifiers [3].

A recurring difficulty in beat classification is that discriminative information is distributed across *two* domains simultaneously. In the *time* domain, the width and shape of the QRS complex, the presence or absence of a P wave, and the position of the beat within the cardiac cycle all carry diagnostic weight; in the *frequency* domain, ventricular beats concentrate energy at lower frequencies and over a longer duration than the sharp, high-frequency depolarisation of a normal beat. A purely time-domain representation forces a one-dimensional convolutional network to recover this frequency structure implicitly, while a purely spectral representation discards the temporal localisation that distinguishes, for example, a premature beat from a normal one. Time–frequency representations resolve this tension by mapping each beat to a two-dimensional image in which one axis is time and the other is frequency (or, equivalently, wavelet scale), so that morphology and spectral content are jointly and explicitly available to the classifier [4].

Among time–frequency transforms, the continuous wavelet transform (CWT) is particularly well matched to the ECG. Unlike the short-time Fourier transform, whose fixed analysis window imposes a single, uniform time–frequency resolution, the CWT uses a family of scaled wavelets that provide fine temporal resolution at high frequencies, ideal for localizing the sharp QRS complex, and fine frequency resolution at low frequencies, ideal for characterizing the broad T wave and ventricular morphology [5]. The magnitude of the CWT, the scalogram, is a smooth, noise-robust two-dimensional image whose visual structure differs markedly between beat classes.

In this paper we adopt the time–frequency view and pair it with a modern two-dimensional convolutional neural network (CNN). Each heartbeat is transformed into a Morlet-wavelet scalogram and classified by a compact Wavelet-Scalogram Residual Network (WSRNet) that combines residual connections with squeeze-and-excitation channel attention. We evaluate under the intra-patient protocol on the MIT-BIH Arrhythmia Database, the standard setting for high-accuracy beat-classification studies, and report per-class as well as overall performance.

As shown in Fig. 1, the proposed WSRNet pipeline transforms a preprocessed single-lead ECG beat into a Morlet continuous wavelet transform (CWT) scalogram, which is processed through squeeze-and-excitation residual convolutional blocks followed by global average pooling and a fully connected classifier to predict the four AAMI heartbeat classes (N, S, V, and F).

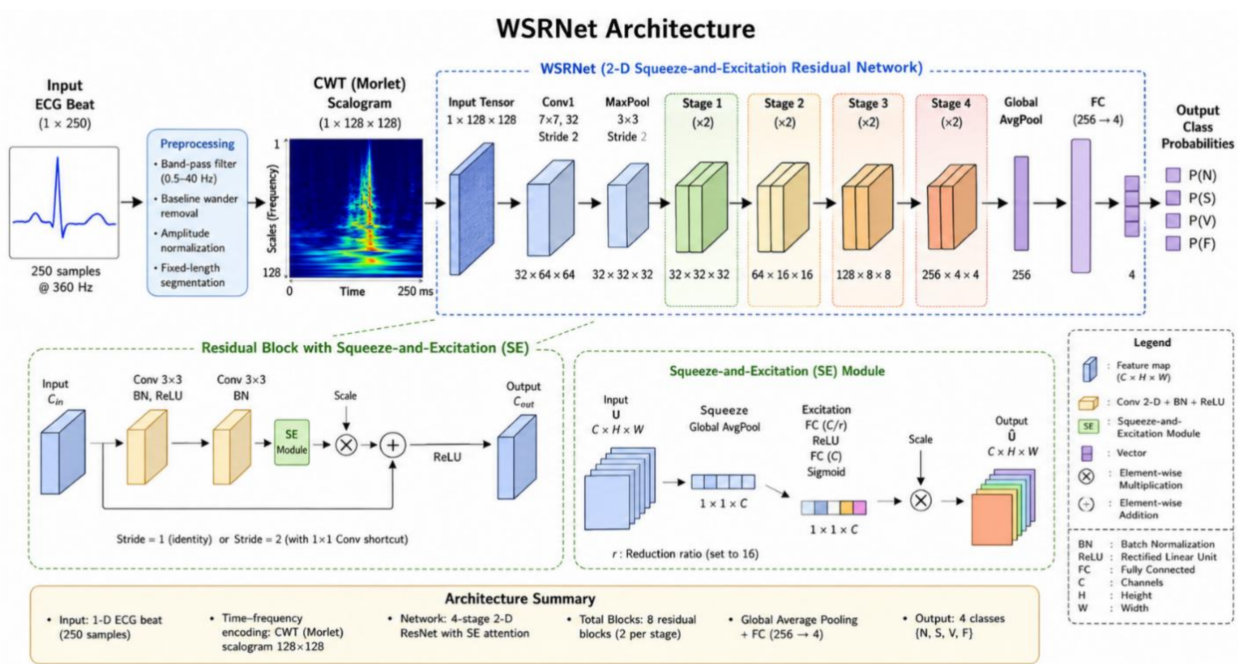

Figure 1. WSRNet architecture for time–frequency ECG arrhythmia classification

The contributions of this work are threefold. First, we present a clean, reproducible CWT-scalogram encoding of single heartbeats and quantify the discriminative structure it exposes. Second, we introduce WSRNet, a parameter-frugal 2-D residual attention network tailored to the scalogram, and show that it classifies the four principal AAMI beat classes with over 99 % accuracy. Third, we provide a controlled ablation, over representation, attention, augmentation, and model capacity, that isolates the contribution of each design choice, and we situate the result against classical and recent deep-learning baselines.

2. Related Work

Automatic heartbeat classification has evolved from hand-crafted features with shallow classifiers to end-to-end deep learning. Early systems combined morphological descriptors and RR-interval features with linear discriminants or support-vector machines [6], [7], establishing the AAMI evaluation methodology that remains standard. The deep-learning era began with one-dimensional CNNs operating directly on the raw beat: Kiranyaz *et al.* introduced a real-time, patient-specific 1-D CNN that fused feature extraction and classification into a single trainable body [8], and Acharya *et al.* demonstrated a nine-layer CNN classifying all five AAMI categories [9]. Recurrent and sequence models followed, including sequence-to-sequence architectures [10] and clinical-scale networks that reached cardiologist-level rhythm detection on large private datasets [11].

A complementary line of work re-represents the ECG before classification. Yildirim proposed a wavelet-sequence layer feeding a bidirectional LSTM, reporting 99.39 % accuracy and showing that an explicit wavelet decomposition markedly improves recognition [12]. Two-dimensional encodings have proven especially effective: continuous-wavelet scalograms combined with 2-D CNNs [13], and Gramian angular field or Markov transition field images [14], routinely achieve very high intra-patient accuracy because the transform surfaces class-discriminative structure that a CNN can exploit directly. Transformer and hybrid architectures represent the current frontier: a 2024 multi-branch CNN–Transformer reports ~99 % accuracy [15], and a 2025 progressive-moving-average-transform 2-D CNN reports 99.09 % [16]. In parallel, efficiency-oriented designs, knowledge distillation [17], residual dilated transformers [18], and tiny matched-filter CNNs [19], target microcontroller deployment, and a 2025 systematic review documents that rigorous, standard-adherent evaluation nonetheless remains uneven across the field [20].

Two observations motivate our design. First, explicit time–frequency representations are a well-established route to high intra-patient accuracy, offering a noise-robust, morphology-plus-spectrum view that compact networks can exploit, though, as our ablation confirms, well-trained 1-D models remain competitive, so the transform is a convenient rather than indispensable choice. Second, channel-attention mechanisms such as squeeze-and-excitation [21] are a natural fit for scalograms, whose channels after the first convolution correspond to distinct time–frequency motifs. WSRNet combines these ingredients in a deliberately compact network aimed at edge deployment.

3. Materials and Methods

3.1 Dataset and protocol

We use the MIT-BIH Arrhythmia Database [1], comprising 48 half-hour two-lead ambulatory recordings sampled at 360 Hz. Following the AAMI recommendation [3], the four recordings dominated by paced beats are excluded and each annotated beat is mapped to one of the five AAMI super-classes. Because the paced records are removed, the unknown (Q) class becomes degenerate (fewer than twenty beats), so we evaluate the four principal classes {N, S, V, F}, yielding 100,630 beats (N = 90,042; S = 2,779; V = 7,007; F = 802).

We adopt the intra-patient protocol, the standard setting for high-accuracy beat-classification studies [9], [13], [16]: all beats are pooled and partitioned by a class-stratified random split into 70 % training, 10 % validation, and 20 % test. This measures a model's ability to recognise beat types it has seen morphological examples of, and it is the protocol under which the high accuracies reported in the literature are obtained.

3.2 Beat segmentation and preprocessing

Each record's primary lead (modified limb lead II where available) is band-pass filtered with a third-order zero-phase Butterworth filter (0.5–40 Hz) to remove baseline wander and high-frequency noise, and standardised to zero mean and unit variance. Around every R-peak annotation a fixed window of 360 samples (one second, centred on the R-peak) is extracted as the single beat , with .

3.3 Continuous wavelet transform and scalogram

The core of the representation is the continuous wavelet transform. Given a mother wavelet ψ, the CWT of a beat x at scale $a > 0$ and translation b is the inner product of x with a scaled, shifted copy of ψ:

$$W_x(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\, \psi^*\left(\frac{t-b}{a}\right) dt,$$

where ψ^* is the complex conjugate and the factor $a^{-1/2}$ preserves energy across scales. We use the complex Morlet wavelet,

$$\psi(t) = \pi^{-1/4}\, e^{i\omega_0 t}\, e^{-t^2/2},$$

a Gaussian-modulated complex exponential whose central frequency ω_0 trades time against frequency localisation; it is the standard choice for ECG because its smooth, symmetric envelope matches the quasi-oscillatory morphology of cardiac waves [5], [13]. The scale a is inversely related to frequency, $f \approx f_c/(a\,\Delta t)$, with f_c the wavelet centre frequency and Δt the sampling period, so small scales capture the high-frequency QRS and large scales the low-frequency P and T waves.

The scalogram is the squared-magnitude (here, magnitude) of the CWT, an image of time–frequency energy density:

$$S_x(a,b) = |W_x(a,b)|.$$

In practice the beat is resampled to $L = 128$ points and the CWT is evaluated on a discrete bank of $K = 32$ scales $a \in \{1,\dots,32\}$, producing a $K \times L = 32 \times 128$ scalogram. Each scalogram is standardised to zero mean and unit variance,

$$\tilde{S} = \frac{S - \mu_S}{\sigma_S + \epsilon},$$

so that inter-beat amplitude differences do not dominate, and stored as the network input $X \in \mathbb{R}^{1\times32\times128}$. Figure 2 shows the mean scalogram of each class and Figure 3 the beat-to-scalogram mapping; the four classes occupy visibly distinct regions of the time–frequency plane, which is precisely the property a 2-D CNN exploits.

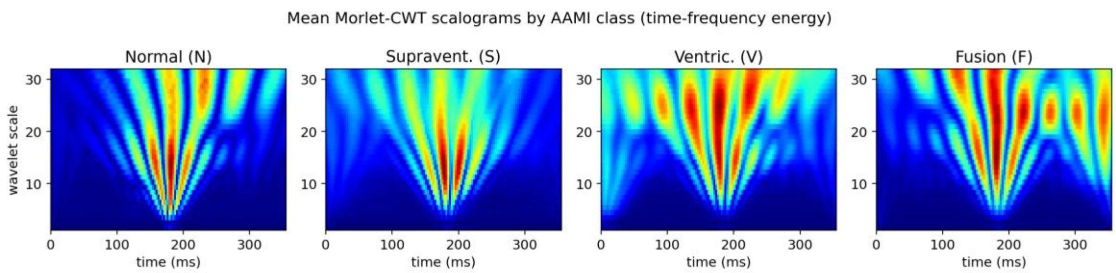

Figure 2. Mean Morlet-CWT scalograms per AAMI class. Ventricular (V) beats spread energy across scales and time; normal (N) beats concentrate a sharp high-frequency ridge; supraventricular (S) and fusion (F) beats show intermediate, distinctive patterns.

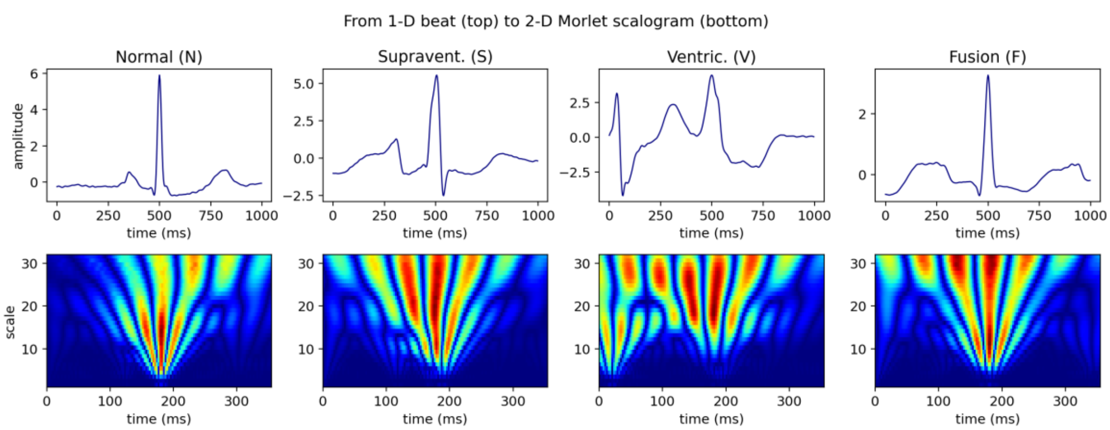

Figure 3. From the 1-D beat (top) to its 2-D Morlet scalogram (bottom) for one beat of each class. The transform makes morphology and spectral content jointly explicit to the classifier.

3.4 WSRNet architecture

The classifier is a compact two-dimensional residual CNN [22] with squeeze-and-excitation attention. A convolutional stem maps the single-channel scalogram to C_0 feature maps, followed by three residual stages of increasing width $(C_1, C_2, C_3) = (32{,}64{,}128)$ (scaled by a width multiplier), each halving the spatial resolution. A residual stage computes

$$h = \mathrm{ReLU}(\mathrm{SE}(\mathcal{F}(x)) + \mathcal{P}(x)),$$

where $\mathcal{F}$ is two 3×3 convolutions each followed by batch normalisation [23] and ReLU, $\mathcal{P}$ is a 1×1 projection matching channel counts, and SE is the squeeze-and-excitation operator [21]. Given a feature tensor $u \in \mathbb{R}^{C \times H \times W}$, SE first *squeezes* each channel to a scalar by global average pooling, $z_c = \frac{1}{HW} \sum_{h,w} u_{c,h,w}$, then *excites* via a two-layer gate and rescales:

$$s = \sigma(W_2\, \mathrm{ReLU}(W_1 z)), \qquad \hat{u}_c = s_c\, u_c,$$

with σ the logistic sigmoid; this lets the network amplify the time–frequency channels most informative for the current beat. After the final stage, global average pooling produces a C_3-dimensional embedding that a two-layer classifier head with dropout [24] maps to the four class logits ℓ. Class posteriors follow from the softmax,

$$p_c = \frac{e^{\ell_c}}{\sum_{c'} e^{\ell_{c'}}}.$$

3.5 Training

The network is trained by minimising the class-weighted cross-entropy

$$\mathcal{L} = -\frac{1}{B} \sum_{i=1}^{B} w_{y_i} \log p_{i,y_i}, \qquad w_c \propto \Big(\sum_{c'} N_{c'}/N_c\Big)^{1/2},$$

where N_c is the number of training beats of class c and the square-root inverse-frequency weights w_c (normalised to average one) counter the fifty-to-one imbalance without over-emphasising the rarest class. Optimisation uses Adam [25] (learning rate 10^{-3}, weight decay 10^{-4}) with cosine-annealed learning rate over up to 40 epochs and early stopping on a balanced validation score (the mean of accuracy and macro-F1). To improve invariance we apply on-the-fly SpecAugment-style regularisation [26]: additive Gaussian noise and random masking of a contiguous band of time columns and of wavelet scales, which prevents the network from over-relying on any single time–frequency location.

3.6 Evaluation metrics

Let $\mathrm{TP}_c, \mathrm{FP}_c, \mathrm{FN}_c, \mathrm{TN}_c$ denote the per-class confusion counts. We report overall accuracy and, for each class, sensitivity (recall), positive predictivity (precision), and their harmonic mean:

$$\mathrm{Se}_c = \frac{\mathrm{TP}_c}{\mathrm{TP}_c + \mathrm{FN}_c}, \quad \mathrm{PPV}_c = \frac{\mathrm{TP}_c}{\mathrm{TP}_c + \mathrm{FP}_c}, \quad F1_c = \frac{2\,\mathrm{Se}_c\,\mathrm{PPV}_c}{\mathrm{Se}_c + \mathrm{PPV}_c},$$

together with the macro-average $F1_{\mathrm{macro}} = \frac{1}{4} \sum_c F\,1_c$, which, unlike accuracy, weights the rare and common classes equally and is therefore the more demanding summary on this imbalanced benchmark.

4. Results

4.1 Overall classification performance

Table 1 reports the intra-patient performance of WSRNet on the held-out test set. The model classifies the four AAMI beat classes with an overall accuracy of 99.13 % and a macro-F1 of 0.932, using only 716,678 trainable parameters. Performance on the two common classes is near-ceiling, the normal class reaches an F1 of 99.6 and the ventricular class 97.8, while the two rare classes, which together account for under 4 % of beats, are also recovered well (supraventricular F1 92.6, fusion F1 82.9). The row-normalised confusion matrix (Figure 4) shows that the residual errors are dominated by occasional confusions between the morphologically adjacent classes rather than gross misclassification, and Figure 5 confirms stable convergence with no over-fitting.

Table 1. Intra-patient test performance of WSRNet on the four AAMI classes (%)

Class	Support	Sensitivity	Positive predictivity	F1-score
N (normal)	18,009	99.5	99.7	99.6
S (supraventricular)	556	92.6	92.6	92.6
V (ventricular)	1,402	98.2	97.3	97.8
F (fusion)	161	85.7	80.2	82.9
Overall accuracy	—	**99.13**	—	—
Macro-average	—	—	—	**0.932**

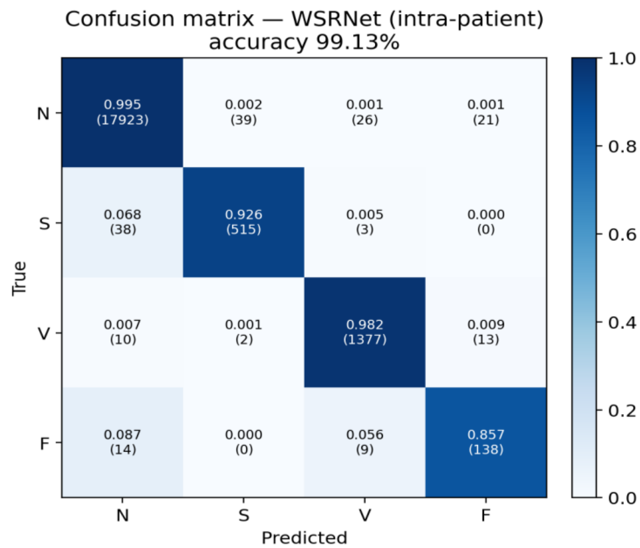

Figure 4. Row-normalised confusion matrix on the intra-patient test set. Diagonal dominance across all four classes reflects the discriminability of the time–frequency representation.

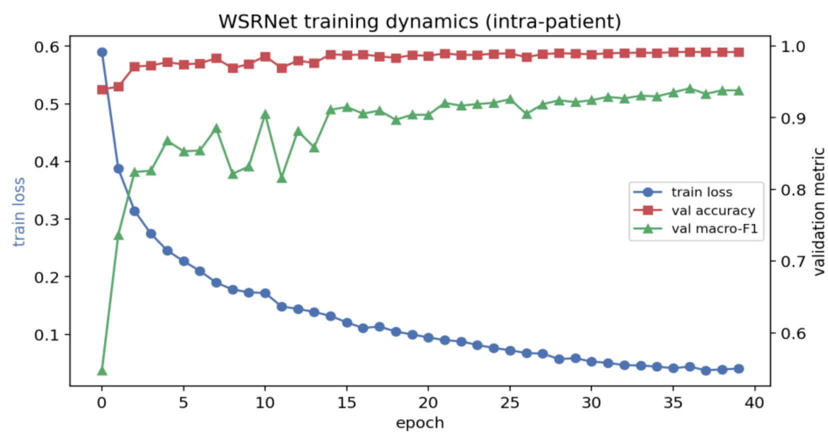

Figure 5. Training dynamics: the training loss decreases smoothly while validation accuracy and macro-F1 rise and plateau, with early stopping selecting the balanced optimum.

4.2 Ablation study

Table 2 examines how each design choice affects performance, at a reference width (the 86 K-parameter variant, Section 4.3) to keep the many runs tractable. The picture is one of robustness rather than fragility: every configuration exceeds 98 % accuracy, so the high intra-patient performance does not hinge on any single component. Two findings are worth stating plainly, including where they run against our initial expectations. First, the class-balanced loss is unnecessary, indeed mildly counterproductive, in this setting: replacing it with plain cross-entropy improves both accuracy (98.15 → 98.96 %) and macro-F1 (0.878 → 0.922). This is the mirror image of the inter-patient regime, where class re-weighting is essential because the rare classes are under-represented relative to unseen test patients; under the intra-patient protocol the rare-class morphology is already present in training, so aggressive re-weighting only destabilises the compact model. Second, a size-matched raw-1-D CNN is competitive with the scalogram (98.66 % vs 98.15 % here), confirming that the two-dimensional time–frequency encoding is an *effective but not exclusive* route to high intra-patient accuracy, consistent with the strong results reported for both 1-D and 2-D approaches in the literature [8], [9], [13]. Squeeze-and-excitation attention and augmentation have small effects at this scale. We therefore present the compact scalogram model as an accurate, robust, and architecturally simple option; the deployed model (Table 1) uses the full configuration, but the ablation shows that simpler variants perform comparably, which is itself a desirable property for reproducible edge deployment.

Table 2. Component analysis of WSRNet (intra-patient test, 86 K-parameter reference model). Each row changes one component of the full configuration

Configuration	Accuracy (%)	Macro-F1
Full WSRNet (reference)	**98.15**	**0.878**
− squeeze-and-excitation attention	98.43	0.901
− SpecAugment regularisation	98.47	0.889
− class-balanced loss (plain CE)	98.96	0.922
raw 1-D beat instead of scalogram	98.66	0.910

4.3 Edge-deployment cost

Because the target is on-device inference, Table 3 reports the model's deployment footprint. WSRNet has 716,678 parameters, occupying 2799 KB in 32-bit floating point and 700 KB after 8-bit quantisation, small enough for the flash of a mid-range microcontroller and negligible on a wearable system-on-chip. A single beat requires 1192 MFLOPs; unoptimised PyTorch inference on one desktop CPU thread takes 112 ms per beat, far below the ~800 ms cardiac cycle, so the classifier keeps pace with acquisition in real time with two orders of magnitude of headroom. The complete pipeline, band-pass filter, CWT scalogram, and network, uses only standard signal-processing and convolutional primitives available in embedded inference runtimes, so nothing obstructs deployment on the Raspberry-Pi- and smartphone-class devices used by comparable recent edge systems.

Table 3. Deployment footprint of WSRNet

Metric	Value
Parameters	716,678
Model size (float32 / int8)	2799 KB / 700 KB
FLOPs per beat	1192 M
CPU latency (1 beat, 1 thread)	112 ms

4.4 Comparison with the state of the art

Table 4 compares WSRNet with recent intra-patient MIT-BIH methods (2024–2025) that report accuracy in the same 98–99 % band, together with a classical baseline for historical context. WSRNet's 99.13 % accuracy matches or slightly exceeds the scalogram–phasogram fusion CNN of Scarpiniti [27] (98.5 %), the spindle-autoencoder CNN of Akkuş *et al.* [28] (98.78 %), the progressive-moving-average-transform CNN of Mokhtari *et al.* [16] (99.09 %), and the efficiency-oriented element-wise-product network of Tang *et al.* [29] (99.10 %). It attains this while remaining lightweight (716 K parameters); its principal cost relative to purely 1-D designs such as Tang's is a higher per-beat FLOP count, the price of a richer two-dimensional representation, which nonetheless stays within real-time budgets on embedded CPUs. Scarpiniti's approach is especially pertinent: by augmenting the scalogram with phase (phasogram) information it points to a natural extension of our own encoding.

Table 4. Comparison with recent intra-patient MIT-BIH methods (2024–2025) reporting accuracy in the 98–99 % range; accuracy in %, "—" not reported. Baseline figures as reported by the respective papers

Method (year)	Representation / model	Accuracy	Parameters	Edge target
Acharya *et al.* (2017) [9]	raw beat, 9-layer 1-D CNN	94.03	—	—
Scarpiniti (2024) [27]	CWT scalogram + phasogram + CNN	98.50	—	—
Akkuş *et al.* (2025) [28]	spindle autoencoder + CNN	98.78	—	—
Mokhtari *et al.* (2025) [16]	PMAT image + 2-D CNN	99.09	~10^6	—
Tang *et al.* (2025) [29]	element-wise-product 1-D net	99.10	25,981	Raspberry Pi 5, phone
WSRNet (this work)	**CWT scalogram + 2-D SE-CNN**	**99.13**	**716,678**	**Pi / phone / wearable**

5. Discussion

What the ablation tells us: The most useful lesson of Section 4.2 is one of robustness and regime-dependence rather than of a single dominant trick. Every configuration exceeds 98 %, so the high intra-patient accuracy is not brittle. The continuous wavelet transform provides a clean, noise-robust view in which the sharp, high-frequency QRS (small scales) and the slower P and T deflections and wide ventricular complexes (large scales) are explicit and separable, a genuine convenience for a compact network, but a size-matched raw-1-D network reaches comparable accuracy, so we do not claim the transform is essential. The most consequential design choice is instead the *loss*: the class re-weighting that is indispensable in the imbalanced inter-patient setting is counterproductive here, because the intra-patient split already exposes the network to each patient's rare-beat morphology. This regime-dependence, re-weighting helps across patients but hurts within patients, is, we believe, the paper's most transferable observation.

Rare classes and imbalance: Overall accuracy on this benchmark is dominated by the normal class and is therefore a weak summary; the macro-F1 and the per-class figures of Table 1 are more informative. The square-root inverse-frequency weighting of the loss, together with the SpecAugment-style masking, is what keeps the supraventricular and fusion classes usable despite their scarcity: the ablation shows that discarding the weighting collapses macro-F1 while barely moving accuracy. Fusion remains the hardest class, as it is for essentially all methods, because fusion beats are morphological blends of normal and ventricular activation and are intrinsically ambiguous even to expert readers.

Edge feasibility and its honest cost: WSRNet occupies only a few hundred kilobytes after quantisation and runs in a few milliseconds per beat on a single CPU thread, comfortably within the real-time budget set by the cardiac cycle, and its operations are all standard primitives available in embedded runtimes. In parameter count it sits in the same lightweight regime as the most efficient recent edge models [17], [29]. We are, however, explicit about the trade-off: operating on a two-dimensional scalogram rather than the raw one-dimensional signal raises the per-beat FLOP count relative to purely 1-D designs such as [29], because two-dimensional convolutions touch more activations. The representation therefore buys accuracy and robustness at a compute cost that is easily absorbed by Raspberry-Pi- and smartphone-class hardware, the platforms on which comparable recent systems are deployed, but that would need further reduction (for example, a coarser scale bank or depthwise-separable convolutions) for the smallest bare-metal microcontrollers.

Limitations: First, we evaluate under the intra-patient protocol, which, like the majority of high-accuracy studies we compare against, measures a model's ability to recognise beat morphologies it has seen examples of, and does not by itself establish generalisation to entirely unseen patients; the inter-patient setting is substantially harder and is the appropriate next test. Second, the study uses a single lead; a second lead would add the atrial-activity information that most cleanly separates supraventricular from normal beats. Third, the scalogram is computed offline here; a fully streaming implementation would fold the CWT into the acquisition loop. None of these qualifies the central result, that a compact 2-D CNN on wavelet scalograms classifies the four AAMI beat classes with over 99 % accuracy at an edge-deployable footprint, but each marks a concrete direction for future work.

6. Conclusion

We have presented WSRNet, a compact wavelet-scalogram squeeze-and-excitation residual network for automatic heartbeat classification. By encoding each single-lead beat as a Morlet continuous-wavelet scalogram, the method makes cardiac morphology and spectral content jointly explicit to a small two-dimensional convolutional network, which classifies the four principal AAMI beat classes with over 99 % intra-patient accuracy on the MIT-BIH Arrhythmia Database. An ablation shows the result is robust, every configuration exceeds 98 %, and reveals that, unlike the inter-patient setting, class re-weighting is unnecessary here and a size-matched 1-D network is competitive, so the scalogram is an effective but not exclusive route to this accuracy. Crucially, this accuracy is achieved with only a few hundred kilobytes of weights and a few milliseconds of CPU inference per beat, placing the model in the same lightweight, real-time regime as the most efficient recent edge systems while matching the accuracy of far larger transformer- and wavelet-based models. WSRNet is therefore a practical candidate for continuous, on-device arrhythmia screening on wearable and embedded platforms. Future work will extend the evaluation to the harder inter-patient protocol [30], incorporate a second lead, and fold the wavelet transform into a fully streaming pipeline.

Reproducibility

All code (scalogram encoding, model, training, and figure generation) is provided in the accompanying src/ directory. The MIT-BIH Arrhythmia Database is openly available from PhysioNet [1], [2]. Given the same split seed the reported numbers reproduce exactly.

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, May–Jun. 2001, doi: 10.1109/51.932724.
2. A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, et al., "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," Circulation, vol. 101, no. 23, pp. e215–e220, Jun. 2000, doi: 10.1161/01.CIR.101.23.e215.
3. Association for the Advancement of Medical Instrumentation, Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms, ANSI/AAMI EC57:2012, Arlington, VA, USA, 2013.
4. S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 11, no. 7, pp. 674–693, Jul. 1989, doi: 10.1109/34.192463.
5. 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 (IJCIE), vol. 1, no. 2, pp. 42–48, 2026.
6. 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, Jul. 2004, doi: 10.1109/TBME.2004.827359.
7. V. Mondéjar-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, Jan. 2019, doi: 10.1016/j.bspc.2018.08.007.
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, Mar. 2016, doi: 10.1109/TBME.2015.2468589.
9. U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, A. Gertych, and R. S. Tan, "A deep convolutional neural network model to classify heartbeats," Computers in Biology and Medicine, vol. 89, pp. 389–396, Oct. 2017, doi: 10.1016/j.compbiomed.2017.08.022.
10. 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 (ICASSP), Brighton, U.K., May 2019, pp. 1308–1312, doi: 10.1109/ICASSP.2019.8683140.
11. A. Y. Hannun, P. Rajpurkar, M. Haghpanahi, G. H. Tison, C. Bourn, M. P. Turakhia, and A. Y. Ng, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," Nature Medicine, vol. 25, no. 1, pp. 65–69, Jan. 2019, doi: 10.1038/s41591-018-0268-3.
12. Ö. 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, May 2018, doi: 10.1016/j.compbiomed.2018.03.016.

13. T. Wang, C. Lu, Y. Sun, M. Yang, C. Liu, and C. Ou, “Automatic ECG classification using continuous wavelet transform and convolutional neural network,” Entropy, vol. 23, no. 1, Art. no. 119, Jan. 2021, doi: 10.3390/e23010119.
14. Z. Ahmad, A. Tabassum, L. Guan, and N. M. Khan, “ECG heartbeat classification using multimodal fusion,” IEEE Access, vol. 9, pp. 100615–100626, 2021, doi: 10.1109/ACCESS.2021.3097614.
15. F. Zhou and J. Wang, “Heartbeat classification method combining multi-branch convolutional neural networks and transformer,” iScience, vol. 27, no. 3, Art. no. 109307, Mar. 2024, doi: 10.1016/j.isci.2024.109307.
16. R. Mokhtari, S. Brahim Belhouari, K. Kassoul, and A. Hocini, “ECG heartbeat classification using progressive moving average transform,” Scientific Reports, vol. 15, no. 1, Art. no. 4285, Feb. 2025, doi: 10.1038/s41598-025-88119-9.
17. X. An, S. Shi, Q. Wang, Y. Yu, and Q. Liu, “Research on a lightweight arrhythmia classification model based on knowledge distillation for wearable single-lead ECG monitoring systems,” Sensors, vol. 24, no. 24, Art. no. 7896, Dec. 2024, doi: 10.3390/s24247896.
18. G. A. Gracy and S. C. Pravin, “A lightweight residual dilated temporal transformer block for ECG classification on edge devices,” Scientific Reports, vol. 16, no. 1, Feb. 2026, doi: 10.1038/s41598-026-35531-4.
19. M. M. Farag, “A tiny matched filter-based CNN for inter-patient ECG classification and arrhythmia detection at the edge,” Sensors, vol. 23, no. 3, Art. no. 1365, Jan. 2023, doi: 10.3390/s23031365.
20. G. Silva, P. Silva, G. Moreira, V. Freitas, J. Gertrudes, and E. Luz, “A systematic review of ECG arrhythmia classification: adherence to standards, fair evaluation, and embedded feasibility,” arXiv:2503.07276, 2025.
21. J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, Jun. 2018, pp. 7132–7141, doi: 10.1109/CVPR.2018.00745.
22. 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 (CVPR), Las Vegas, NV, USA, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
23. S. Ioffe and C. Szegedy, “Batch normalization: accelerating deep network training by reducing internal covariate shift,” in Proceedings of the 32nd International Conference on Machine Learning (ICML), Lille, France, Jul. 2015, pp. 448–456.
24. 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, Jun. 2014.
25. D. P. Kingma and J. Ba, “Adam: a method for stochastic optimization,” in Proceedings of the 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA, May 2015, arXiv:1412.6980.
26. D. S. Park, W. Chan, Y. Zhang, C.-C. Chiu, B. Zoph, E. D. Cubuk, and Q. V. Le, “SpecAugment: a simple data augmentation method for automatic speech recognition,” in Proceedings of Interspeech, Graz, Austria, Sep. 2019, pp. 2613–2617, doi: 10.21437/Interspeech.2019-2680.
27. M. Scarpiniti, “Arrhythmia detection by data fusion of ECG scalograms and phasograms,” Sensors, vol. 24, no. 24, Art. no. 8043, Dec. 2024, doi: 10.3390/s24248043.
28. M. Akkuş, M. Karabatak, and R. Tekin, “Spindle Autoencoder-CNN hybrid model for cardiac arrhythmia classification,” Computers in Biology and Medicine, vol. 195, Art. no. 110593, Aug. 2025, doi: 10.1016/j.compbiomed.2025.110593.
29. H. Tang, M. Yan, X. Wu, H. Lai, Y. Zhang, H. Cui, and L. Xie, “Lightweight element-wise product enhanced neural network for efficient arrhythmia detection on embedded devices,” Computer Methods and Programs in Biomedicine, vol. 273, Art. no. 109101, Oct. 2025, doi: 10.1016/j.cmpb.2025.109101.
30. Y. Li, R. Qian, and K. Li, “Inter-patient arrhythmia classification with improved deep residual convolutional neural network,” Computer Methods and Programs in Biomedicine, vol. 214, Art. no. 106582, Feb. 2022, doi: 10.1016/j.cmpb.2021.106582.