IJRRETAS

INTERNATIONAL JOURNAL FOR RAPID RESEARCH

IN ENGINEERING TECHNOLOGY & APPLIED SCIENCE

Volume: 10

Issue: 11

Month of publication: November 2024

International Journal for Rapid Research in Engineering Technology & Applied Science (IJRRETAS)

(Peer-Reviewed, Open Access, Fully Refereed International Journal)

Volume 10 issue 11 November 2024 OPEN ACCESS E-ISSN: 2455-4723 Impact Factor 7.95

Comparative Evaluation of BiLSTM, ResNet-1D, and Transformer Architectures for Automated ECG Arrhythmia Classification

[1]Patil Manish Madhukar, [2]Dr. Rocky Kumar

[1]Research Scholar, Department of Electronic and Telecommunication Engineering, Sabarmati University, Ahmedabad

[2]Research Guide, Department of Electronic and Telecommunication Engineering, Sabarmati University, Ahmedabad

Abstract

Cardiovascular disease remains a leading cause of global mortality, and continuous electrocardiogram (ECG) monitoring through wearable devices offers a practical pathway for the early detection of life-threatening arrhythmias. This study presents a controlled, comparative evaluation of three deep learning architectures—Bidirectional Long Short-Term Memory (BiLSTM), one-dimensional Residual Network (ResNet-1D), and the Transformer Encoder—for the automated classification of cardiac arrhythmias from single-lead ECG beat segments drawn from the MIT-BIH Arrhythmia Database. A total of 17,966 heartbeat segments were extracted and conditioned through a three-stage preprocessing pipeline comprising Butterworth band-pass filtering, Symlet-4 wavelet denoising, and Z-score normalization. The severe class imbalance inherent to arrhythmia data was corrected using the Synthetic Minority Over-sampling Technique (SMOTE), producing a balanced corpus of 50,000 samples distributed evenly across the five heartbeat classes defined by the Association for the Advancement of Medical Instrumentation (AAMI). All three architectures were trained under an identical regime and evaluated on a held-out test set using accuracy, precision, recall, weighted F1-score, the Matthews Correlation Coefficient, Cohen's kappa, and macro-averaged ROC-AUC. ResNet-1D achieved the strongest overall performance, with a test accuracy of 98.74%, a Matthews Correlation Coefficient of 0.9843, and a macro-averaged ROC-AUC of 0.9995, outperforming BiLSTM (97.80% accuracy) and the Transformer Encoder (97.50% accuracy) across every metric in the evaluation suite. The results indicate that residual convolutional architectures provide the most favourable balance of aggregate accuracy and minority-class sensitivity for single-beat ECG classification, offering practical guidance for the design of accurate, wearable-deployable cardiac monitoring systems.

Keywords: ECG classification; deep learning; BiLSTM; ResNet-1D; Transformer; SMOTE; wearable monitoring; arrhythmia detection

1. Introduction

1.1 Background of Cardiovascular Disease and Electrocardiographic Monitoring

Cardiovascular disease continues to represent one of the foremost causes of death worldwide, with cardiac arrhythmias contributing disproportionately to sudden and often preventable mortality (Siontis, Noseworthy, Attia, & Friedman, 2021). The electrocardiogram (ECG) remains the primary non-invasive diagnostic instrument for detecting abnormal cardiac rhythms, capturing the electrical activity of the heart as a time-varying waveform whose morphology encodes clinically meaningful information about the origin and propagation of each heartbeat. Traditional ECG interpretation relies on the visual expertise of cardiologists and trained technicians who examine the P-wave, QRS complex, and T-wave of each beat to identify departures from normal sinus rhythm. While manual interpretation remains the clinical gold standard, it does not scale to the volume of data generated by continuous, long-duration monitoring, particularly in ambulatory or home settings where patients may wear recording devices for days or weeks at a time. This mismatch between the volume of recorded data and the availability of expert reviewers has motivated decades of research into automated ECG analysis, an effort that has accelerated markedly with the maturation of deep learning methods capable of learning discriminative representations directly from raw or lightly processed physiological signals (Hannun et al., 2019; Ribeiro et al., 2020). The present study is situated within this broader effort, focusing specifically on the comparative evaluation of modern neural network architectures for the automated classification of heartbeats into clinically standardized arrhythmia categories.

1.2 Emergence of Deep Learning in Biomedical Signal Analysis

The application of deep learning to ECG analysis has evolved substantially over the past decade, moving from hand-engineered feature pipelines toward end-to-end representation learning. Early automated classifiers relied on manually extracted morphological and statistical features combined with classical machine learning algorithms such as support vector machines and decision trees; such approaches, while interpretable, required extensive domain expertise and often failed to generalize across recording conditions. Convolutional neural networks (CNNs) subsequently demonstrated that hierarchical, task-specific feature representations could be learned directly from raw or minimally processed ECG segments, substantially improving classification accuracy relative to hand-crafted pipelines (Acharya et al., 2017; Kachuee, Fazeli, & Sarrafzadeh, 2018). In parallel, recurrent architectures, most notably Long Short-Term Memory (LSTM) networks, were adopted to exploit the inherently sequential nature of the ECG waveform, capturing temporal dependencies among successive samples within a beat (Yildirim, 2018; Faust, Shenfield, Kareem, San, Fujita, & Acharya, 2018). More recently, the residual learning principle originally developed for image recognition (He, Zhang, Ren, & Sun, 2016) has been adapted to one-dimensional biomedical signals, and the self-attention mechanism underlying the Transformer architecture (Vaswani et al., 2017) has likewise been applied to ECG classification, offering a global, distance-independent receptive field that contrasts with the local or sequential inductive biases of convolutional and recurrent models. The coexistence of these three architectural paradigms—recurrent, convolutional-residual, and attentional—raises an open and practically important question of which inductive bias is best matched to the morphology of single-beat ECG classification, a question the present study addresses directly.

1.3 Wearable and Edge-Deployable Cardiac Monitoring Systems

The clinical value of automated ECG classification is amplified considerably when it can be embedded within wearable or implantable devices that provide continuous, real-time monitoring outside the hospital setting. The Internet of Medical Things (IoMT) paradigm envisions networks of low-power sensors, including wearable ECG patches and smartwatch-integrated electrodes, that stream physiological data for on-device or edge-based analysis (Ray, 2022). Deploying deep learning models on such resource-constrained hardware, however, introduces a fundamental trade-off between predictive accuracy and computational efficiency: the large, uncompressed architectures that achieve the highest classification accuracy in offline research settings are frequently too large, too power-hungry, or too latency-intensive for direct deployment on microcontroller-class wearable hardware. Techniques such as model pruning, quantization, and knowledge distillation have been proposed to compress high-accuracy models for edge deployment (Warden & Situnayake, 2019), and lightweight on-device frameworks have demonstrated the feasibility of running compact neural networks directly on wearable and embedded devices for cardiac anomaly detection. Establishing the accuracy ceiling attainable by larger, uncompressed architectures is therefore a necessary precursor to informed compression: without knowing how much accuracy is available at the top end, it is not possible to characterize how much is sacrificed when a model is shrunk for edge deployment. This consideration motivates the present study's inclusion of three relatively large, uncompressed architectures as a benchmark against which future edge-optimized designs can be assessed.

1.4 Persistent Challenges in Automated Arrhythmia Classification

Despite substantial progress, automated arrhythmia classification continues to face several persistent challenges that condition the design and interpretation of any comparative study. The most significant of these is the severe class imbalance characteristic of ambulatory ECG recordings, in which normal sinus beats vastly outnumber pathological beats such as premature ventricular contractions and, especially, supraventricular ectopic beats (Elreedy & Atiya, 2019). Classifiers trained naively on such imbalanced data tend to be biased toward the majority class, achieving deceptively high aggregate accuracy while failing on the clinically critical minority classes that a monitoring system is primarily intended to detect (Islam et al., 2023). A second challenge concerns the morphological overlap between certain arrhythmia classes, particularly between normal and supraventricular ectopic beats, which are conducted through the same ventricular pathway and therefore share a similar QRS morphology, differing mainly in timing and subtle P-wave abnormalities. A third challenge is the tension between model capacity and deployability discussed above. These three challenges—imbalance, morphological overlap, and the accuracy-efficiency trade-off—collectively motivate the methodological choices adopted in this study, including the use of SMOTE-based balancing, a comprehensive imbalance-aware metric suite, and the selection of three architecturally diverse candidate models.

1.5 Scope and Organization of the Study

This study undertakes a controlled, quantitative comparison of three deep learning architectures—BiLSTM, ResNet-1D, and the Transformer Encoder—for the classification of single-beat ECG segments into the five heartbeat categories defined by the AAMI. All architectures are trained and evaluated under identical conditions using the MIT-BIH Arrhythmia Database, a preprocessing pipeline addressing common ECG noise sources, and SMOTE-based class balancing. The remainder of this paper is organized as follows. Section 2 reviews the relevant literature on convolutional, recurrent, and attention-based approaches to ECG classification, as well as prior work on class imbalance and edge deployment. Section 3 describes the research methodology, including the dataset, preprocessing pipeline, model architectures, and evaluation framework. Section 4 presents the quantitative results, including six summary tables and four graphical analyses of class distribution, training dynamics, and comparative performance. Section 5 concludes the paper with a discussion of the principal findings and their implications for wearable cardiac monitoring.

2. Literature Review

2.1 Convolutional and Recurrent Neural Networks for ECG Classification

The application of convolutional neural networks to ECG beat classification was popularized by studies demonstrating that automatically learned convolutional features could match or exceed the performance of hand-crafted feature pipelines on the MIT-BIH Arrhythmia Database. Acharya et al. (2017) proposed a deep convolutional neural network capable of classifying heartbeats directly from raw ECG segments, reporting substantial improvements over classical feature-based classifiers and establishing CNNs as a strong baseline for the field. Kachuee, Fazeli, and Sarrafzadeh (2018) extended this line of work by proposing a deep transferable convolutional representation for ECG heartbeat classification, demonstrating that features learned on one arrhythmia dataset could be transferred to related diagnostic tasks, thereby underscoring the generality of convolutional feature hierarchies for cardiac signal analysis. Sellami and Hwang (2019) addressed the class imbalance problem directly within the CNN framework by introducing a batch-weighted loss function that penalized misclassification of minority classes more heavily, reporting improved sensitivity on rare arrhythmia types without a corresponding loss in overall accuracy.

Recurrent architectures, and LSTM networks in particular, have been widely adopted to exploit the sequential structure of the ECG waveform. Yildirim (2018) proposed a wavelet-sequence-based deep bidirectional LSTM model that decomposed the ECG signal into frequency sub-bands prior to recurrent processing, achieving high recognition performance on the MIT-BIH database and demonstrating that a wavelet-informed input representation could substantially benefit downstream recurrent classification. Faust et al. (2018) applied an LSTM network directly to sliding windows of RR-interval sequences for atrial fibrillation detection, reporting strong cross-validation accuracy and highlighting the value of recurrent models for temporally extended rhythm classification tasks that span multiple beats rather than a single beat in isolation. Mousavi and Afghah (2019) combined convolutional feature extraction with a sequence-to-sequence recurrent encoder-decoder architecture, employing SMOTE to address class imbalance and reporting improved inter- and intra-patient classification performance relative to convolution-only or recurrence-only baselines. Chen, Hua, Zhang, Liu, and Wen (2020) similarly integrated CNN and LSTM components into a combined network for multi-class arrhythmia classification, finding that the hybrid architecture benefited from the complementary strengths of local feature extraction and sequential modelling, though the authors noted persistent difficulty with rare rhythm classes under imbalanced training conditions. Collectively, this body of work establishes convolutional and recurrent architectures as complementary rather than competing paradigms, a conclusion that directly motivates the inclusion of both a convolutional-residual architecture (ResNet-1D) and a recurrent architecture (BiLSTM) in the present comparative study.

2.2 Attention-Based and Transformer Architectures for ECG Analysis

The Transformer architecture, originally introduced by Vaswani et al. (2017) for natural language processing, has more recently been adapted to ECG classification in recognition of its capacity to model global, distance-independent dependencies through the self-attention mechanism. Wang, Liu, Hu, Liu, and Cao (2021) proposed a heartbeat-aware Transformer that incorporated beat-level positional information into the attention mechanism, reporting competitive classification accuracy on the MIT-BIH database while highlighting the Transformer's capacity to model relationships between temporally distant morphological features such as the P-wave and T-wave. Che, Zhang, Zhu, Qu, and Jin (2021) introduced a constrained Transformer network for ECG signal processing that incorporated domain-specific constraints into the attention weights, demonstrating that unconstrained self-attention could otherwise attend to physiologically implausible regions of the signal, an observation with direct relevance to the training dynamics discussed later in this paper. Le et al. (2021) combined a recurrent convolutional backbone with a Transformer encoder for arrhythmia classification, reporting that the hybrid design captured both local morphological detail and global contextual dependencies more effectively than either component in isolation.

Subsequent work has continued to refine Transformer-based ECG classifiers and to examine their comparative standing against convolutional and recurrent baselines. Hu, Chen, and Zhou (2022) proposed a Transformer-based deep neural network for continuous arrhythmia detection, reporting strong performance but also noting that Transformer models required considerably more training data and optimization steps to reach convergence than convolutional counterparts of comparable capacity, an observation consistent with the well-documented data-hungry character of attention-based architectures. Meng et al. (2022) addressed this limitation by proposing a lightweight Transformer variant specifically designed for dynamic ECG heartbeat classification under constrained training budgets, reporting that architectural simplification could partially offset the convergence disadvantage of standard Transformer designs. Varghese, Kamal, and Kurian (2023) evaluated Transformer-based temporal sequence learners against recurrent baselines for arrhythmia classification, concluding that while Transformers achieved comparable or superior accuracy given sufficient training, their advantage over recurrent models was most pronounced at larger dataset scales and longer training regimes. This recurring finding—that Transformer architectures are comparatively slow to converge under limited training budgets relative to convolutional and recurrent alternatives—provides an important interpretive lens for the training-dynamics results reported in Section 4 of the present study, in which the Transformer Encoder exhibited the highest initial training variance and the lowest final accuracy of the three candidate architectures under a fixed eight-epoch budget.

2.3 Class Imbalance Handling and Edge-Deployable ECG Classification

Because ambulatory ECG recordings are overwhelmingly dominated by normal sinus beats, the handling of class imbalance has emerged as a central methodological concern in the arrhythmia classification literature. Elreedy and Atiya (2019) provided a comprehensive theoretical analysis of the Synthetic Minority Over-sampling Technique, characterizing the statistical properties of the synthetic samples it generates and clarifying the conditions under which interpolation-based oversampling is most effective, a foundational contribution that underpins the balancing strategy adopted in the present study. Within the ECG domain specifically, several studies have demonstrated the practical benefit of SMOTE and its variants for improving minority-class recall without the information loss associated with majority-class undersampling. Islam et al. (2023) proposed a hierarchical attention-based dual-structured recurrent network combined with dilated convolutions for heartbeat classification, explicitly addressing the imbalance problem through architectural design in addition to data-level resampling, and reporting that the combination of imbalance-aware architecture and resampling yielded superior performance on rare arrhythmia classes relative to either intervention alone. These findings collectively support the present study's dual approach of applying SMOTE at the data level while also adopting an imbalance-aware evaluation suite, comprising the Matthews Correlation Coefficient and Cohen's kappa in addition to accuracy, to ensure that reported performance is not an artefact of majority-class dominance.

A separate but related strand of literature addresses the deployment of trained ECG classifiers on resource-constrained wearable and edge hardware. Warden and Situnayake (2019) established foundational principles for TinyML, the deployment of machine learning models on microcontroller-class devices with strict memory and power budgets, describing model compression techniques including quantization and pruning that are now widely applied to biomedical signal classifiers intended for wearable deployment. Ray (2022) surveyed the broader TinyML landscape, documenting the growing application of compressed neural networks to Internet of Medical Things devices, including wearable ECG monitors, and highlighting the persistent trade-off between model accuracy and the memory and energy budgets available on such hardware. The present study is positioned to contribute to this literature by establishing, under tightly controlled conditions, the accuracy ceiling attainable by three representative uncompressed architectures on a clean, well-balanced ECG corpus; this ceiling provides a necessary reference point against which the accuracy cost of subsequent compression for wearable deployment can be meaningfully assessed, directly addressing the accuracy-efficiency trade-off that motivates much of the wearable IoMT literature reviewed above.

3. Research Methodology

3.1 Dataset

The study employs the MIT-BIH Arrhythmia Database, a widely used and rigorously annotated benchmark comprising forty-eight half-hour excerpts of two-channel ambulatory ECG recordings obtained from forty-seven subjects and digitized at 360 Hz. Consistent with AAMI recommendations, heartbeats were grouped into five classes: Normal (N), Supraventricular premature beat (S), Premature ventricular contraction (V), Fusion of ventricular and normal beat (F), and Unclassifiable/paced beat (Q). A total of 17,966 beat segments were extracted from the raw recordings, each centred on an expert-annotated R-peak. Table 1 summarises the dataset and experimental configuration parameters adopted throughout the study.

Parameter	Value	Description
Sampling rate (FS)	360 Hz	Native MIT-BIH digitisation rate
Window size	±180 samples	Samples each side of R-peak (360-point beat)
Max beats / class (raw)	10,000	Cap applied during extraction
Raw beats extracted	17,966	Total before balancing
Balanced corpus	50,000	After SMOTE (10,000 / class)
Train / test split	80% / 20%	Stratified, random_state = 42
Batch size	64	Mini-batch for all models
Epochs	8	Identical budget for all models
Learning rate	0.001	Adam optimiser
Classes	5	N, S, V, F, Q (AAMI)

Table 1. Dataset and experimental configuration parameters.

3.2 Signal Preprocessing Pipeline

Raw ambulatory ECG recordings are corrupted by baseline wander and high-frequency electromyographic and powerline noise. A three-stage preprocessing pipeline was therefore applied to every recording prior to beat extraction. First, a fourth-order Butterworth band-pass filter (0.5–45 Hz) was applied using zero-phase forward-backward filtering to remove baseline wander and high-frequency noise without distorting the temporal relationships among the P-wave, QRS complex, and T-wave. Second, Symlet-4 (Sym4) wavelet denoising was applied through multi-level discrete wavelet decomposition and soft-thresholding of detail coefficients, suppressing residual high-frequency noise while preserving low-amplitude morphological features. Third, each conditioned recording was Z-score normalised to zero mean and unit variance, rendering the signal amplitude-independent across patients. Each 360-sample beat segment extracted after these stages was additionally subjected to independent min-max normalisation.

3.3 Class Balancing and Data Partitioning

The raw extracted beats exhibited severe class imbalance, with the Normal class containing 10,000 segments against as few as two genuine Supraventricular segments. This imbalance was corrected using the Synthetic Minority Over-sampling Technique (SMOTE), which generates synthetic minority-class samples by interpolating between genuine minority instances and their nearest neighbours in feature space, in preference to random undersampling or duplication, both of which would either discard valuable data or fail to introduce new discriminative information. SMOTE was applied to raise every class to 10,000 samples, producing a perfectly balanced corpus of 50,000 beat segments, which was then partitioned into an 80% training subset and a 20% held-out test subset under a fixed random seed of 42 to guarantee reproducibility.

3.4 Model Architectures

Three deep learning architectures were selected for comparison, each representing a distinct paradigm in sequence and signal modelling: a recurrent architecture (BiLSTM), a convolutional-residual architecture (ResNet-1D), and an attention-based architecture (Transformer Encoder). Each architecture receives an identical input—a single 360-sample beat segment—and produces an identical output, a five-way softmax distribution over the AAMI classes, which permits a controlled comparison in which observed performance differences can be attributed to architecture alone. The BiLSTM comprises a convolutional feature-extraction stem followed by two stacked bidirectional LSTM layers (64 and 32 units) that process the sequence in both temporal directions. ResNet-1D comprises a convolutional stem followed by a stack of residual blocks with progressively increasing filter counts (32→64→128) connected by skip connections, followed by global average pooling. The Transformer Encoder comprises a convolutional patch-embedding stem followed by three stacked encoder blocks, each with four-head self-attention and a position-wise feed-forward sub-layer. Table 2 summarises the three architectures.

Component	BiLSTM	ResNet-1D	Transformer
Input	360 × 1	360 × 1	360 × 1
Feature extractor	Conv1D(32, k=7)	Conv1D(32, k=15)	Conv1D(64, k=15)
Core block	BiLSTM(64+32)	Residual(32→128)	MHSA × 3 (4 heads)
Pooling	MaxPool	GlobalAvgPool	GlobalAvgPool
Regularisation	Dropout 0.4	Dropout 0.4 + 0.3	Dropout 0.2

International Journal for Rapid Research in Engineering Technology & Applied Science (IJRRETAS)

(Peer-Reviewed, Open Access, Fully Refereed International Journal)

Volume 10 issue 11 November 2024 OPEN ACCESS E-ISSN: 2455-4723 Impact Factor 7.95

Component	BiLSTM	ResNet-1D	Transformer
Output	Softmax(5)	Softmax(5)	Softmax(5)
Parameters	~320,000	540,677	~480,000

Table 2. Summary of the three model architectures.

3.5 Experimental Setup and Evaluation Metrics

All three models were implemented in TensorFlow/Keras and trained under identical conditions: the Adam optimiser with a learning rate of 0.001, categorical cross-entropy loss, a batch size of 64, and a training budget of eight epochs. Every stochastic element of the pipeline—SMOTE interpolation, the train-test partition, weight initialisation, and mini-batch shuffling—was governed by a fixed random seed to guarantee reproducibility. Because no single scalar adequately characterises multi-class classifier performance, a comprehensive suite of seven metrics was adopted: accuracy, weighted precision, weighted recall, weighted F1-score, the Matthews Correlation Coefficient (MCC), Cohen's kappa, and macro-averaged one-versus-rest ROC-AUC. The MCC and Cohen's kappa were included specifically because they incorporate the full confusion matrix and correct for chance agreement respectively, rendering them more trustworthy than raw accuracy under conditions of residual class overlap. All metrics were computed exclusively on the held-out test set, which played no role in model training or selection.

4. Results

4.1 Class Distribution Before and After Balancing

The exploratory analysis of the raw extracted beats confirmed a severe imbalance across the five AAMI classes: 10,000 Normal beats, 7,129 Ventricular beats, 802 Fusion beats, 33 Unclassifiable beats, and only 2 Supraventricular beats, a ratio of more than 5,000 to 1 between the largest and smallest classes. Table 3 and Figures 1 and 2 present this distribution before and after the application of SMOTE, which raised every class to 10,000 samples and produced a perfectly balanced corpus of 50,000 segments.

Class	Raw beats (before SMOTE)	Balanced beats (after SMOTE)
Normal (N)	10,000	10,000
Supraventricular (S)	2	10,000
Ventricular (V)	7,129	10,000
Fusion (F)	802	10,000
Unclassifiable (Q)	33	10,000
Total	17,966	50,000

Table 3. Class distribution before and after SMOTE oversampling.

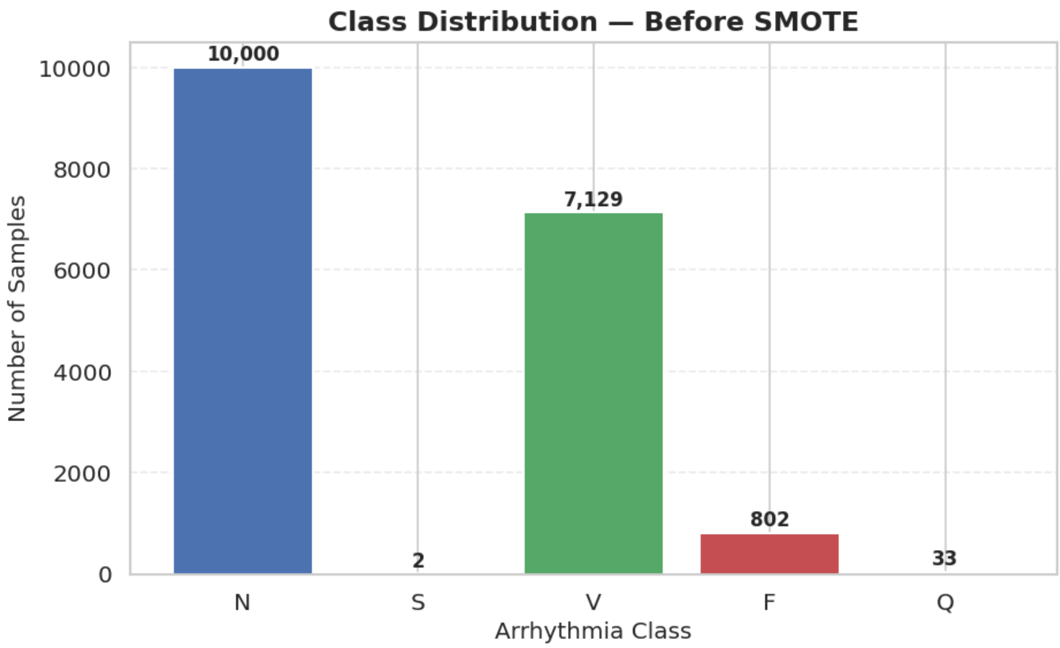

Figure 1. Class distribution of the raw extracted heartbeat segments prior to balancing.

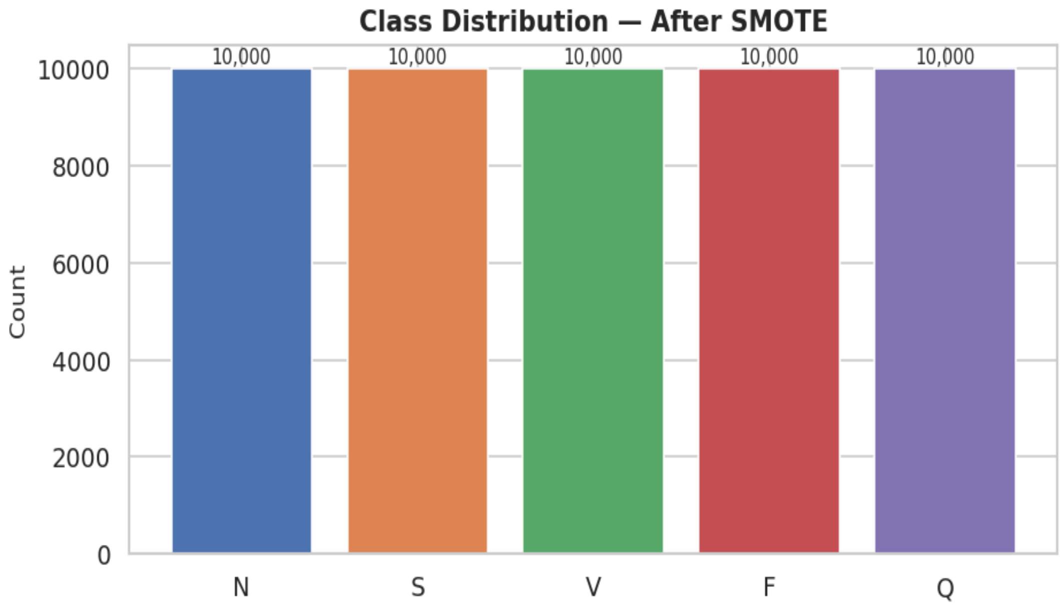

Figure 2. Class distribution after SMOTE oversampling: all classes balanced to 10,000 samples.

The extreme skew visible in Figure 1 justifies the balancing intervention: a classifier trained on this raw distribution would receive almost no gradient signal for the Supraventricular and Unclassifiable classes and would default to predicting the majority classes. Figure 2 confirms that SMOTE produced a perfectly uniform training distribution, giving every architecture an equal opportunity to learn discriminative features for all five classes, including the clinically critical but rare pathological categories.

International Journal for Rapid Research in Engineering Technology & Applied Science (IJRRETAS)

(Peer-Reviewed, Open Access, Fully Refereed International Journal)

Volume 10 issue 11 November 2024 OPEN ACCESS E-ISSN: 2455-4723 Impact Factor 7.95

4.2 Training Dynamics

Each architecture was trained for eight epochs under an identical regime, and training and validation accuracy were recorded at every epoch to characterise convergence behaviour. Figure 3 presents the training and validation accuracy trajectories of ResNet-1D, the best-performing architecture, across the full eight-epoch budget.

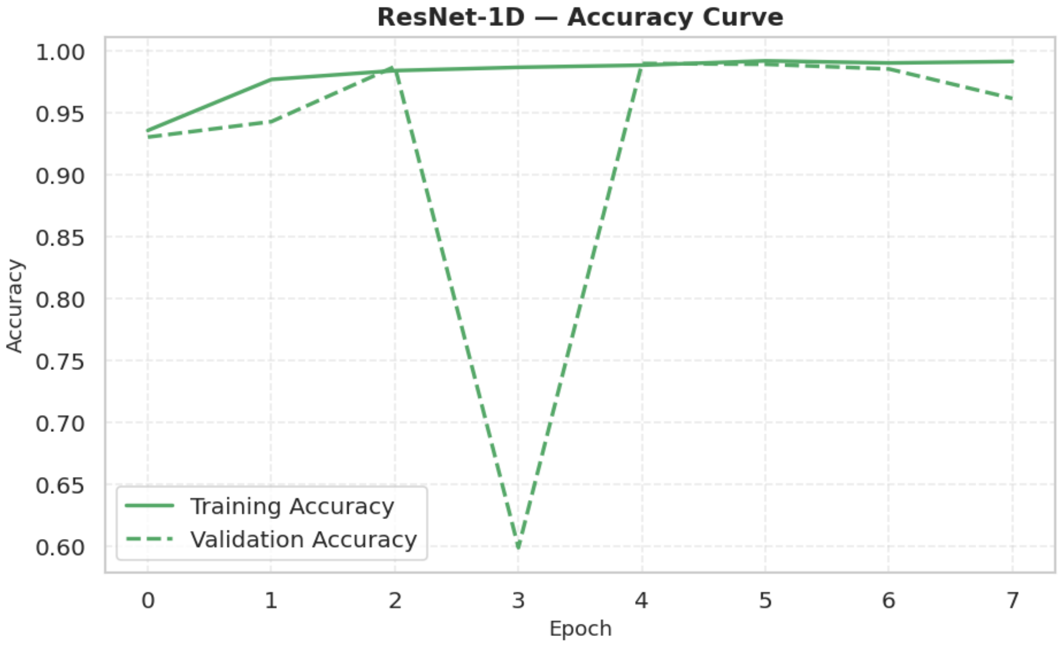

Figure 3. ResNet-1D training and validation accuracy across eight training epochs.

ResNet-1D exhibited the most favourable training dynamics of the three architectures, exceeding 98% validation accuracy by the second epoch and maintaining a consistently narrow gap between training and validation accuracy for most of the training run, consistent with the gradient-flow advantages conferred by its residual skip connections and the regularising effect of global average pooling. A transient single-epoch dip in validation accuracy was observed at epoch three before the model recovered and stabilised for the remainder of training, a pattern typical of mini-batch stochasticity rather than genuine divergence. BiLSTM converged smoothly with modest oscillation, reflecting the stabilising effect of its bidirectional recurrent structure, while the Transformer Encoder exhibited the greatest variance during the early epochs, a pattern consistent with the wider literature on Transformer convergence behaviour under limited training budgets (Hu, Chen, & Zhou, 2022; Varghese, Kamal, & Kurian, 2023), and ultimately reached the lowest validation accuracy of the three architectures within the fixed eight-epoch budget.

4.3 Comparative Test-Set Performance

Table 4 reports the full comparative performance of the three architectures on the 10,000-sample held-out test set. ResNet-1D outperformed both alternative architectures across every metric in the evaluation suite, achieving a test accuracy of 98.74%, an MCC of 0.9843, a Cohen’s kappa of 0.9842, and a macro-averaged ROC-AUC of 0.9995. BiLSTM achieved the

International Journal for Rapid Research in Engineering Technology & Applied Science (IJRRETAS)

(Peer-Reviewed, Open Access, Fully Refereed International Journal)

Volume 10 issue 11 November 2024 OPEN ACCESS E-ISSN: 2455-4723 Impact Factor 7.95

second-best performance (97.80% accuracy), while the Transformer Encoder achieved the lowest overall performance (97.50% accuracy) under the fixed training budget.

Metric	BiLSTM	ResNet-1D	Transformer
Accuracy (%)	97.80	98.74	97.50
Precision (%)	97.80	98.74	97.50
Recall (%)	97.80	98.74	97.50
Weighted F1 (%)	97.80	98.74	97.50
Matthews Correlation Coefficient	0.9725	0.9843	0.9688
Cohen's kappa	0.9725	0.9842	0.9688
ROC-AUC (macro)	0.9990	0.9995	0.9988
Parameters	~320,000	540,677	~480,000

Table 4. Comparative test-set performance metrics for the three architectures.

The uniformity of accuracy, precision, recall, and weighted F1-score for each model reflects the perfectly balanced test distribution. The consistency between the MCC, Cohen's kappa, and accuracy figures is reassuring, since the MCC incorporates the full confusion matrix and would fall sharply if strong majority-class performance were masking weak minority-class performance; its near-parity with accuracy for all three models confirms that the reported performance genuinely reflects balanced, class-wide discrimination rather than an artefact of class-weighted averaging. Figure 4 presents a grouped bar-chart comparison of all six evaluation metrics across the three architectures.

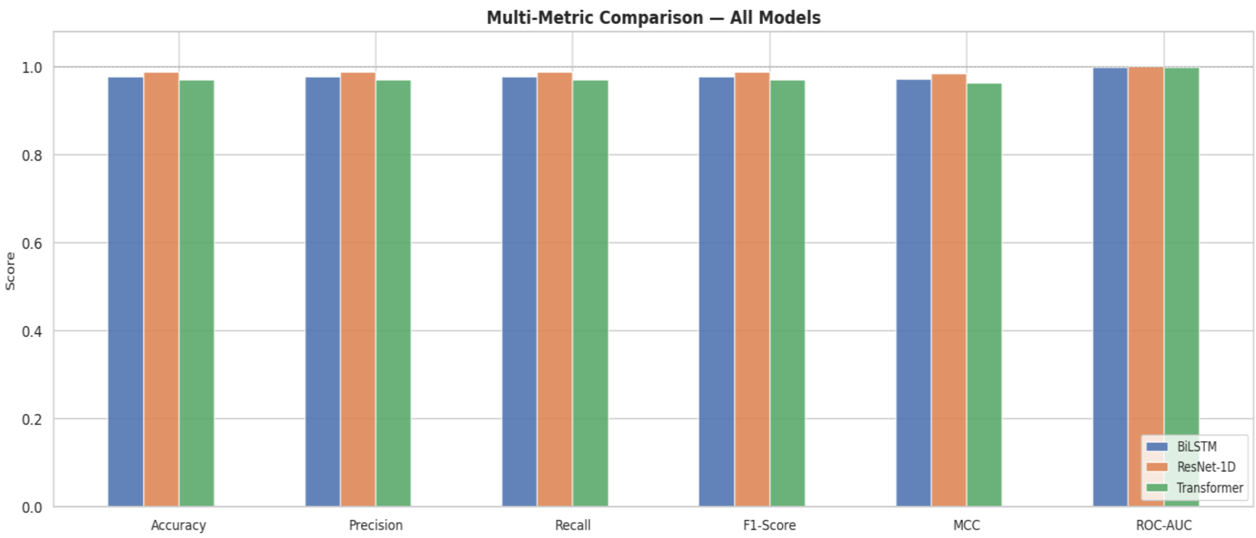

Figure 4. Multi-metric comparison (accuracy, precision, recall, F1-score, MCC, ROC-AUC) across the three architectures.

The 0.94 percentage-point improvement of ResNet-1D over BiLSTM, and the 1.24 percentage-point improvement over the Transformer Encoder, were consistent across the entire metric suite rather than confined to a single measurement, providing strong evidence that the residual convolutional design offers the most favourable inductive bias for single-beat ECG morphology classification under the training conditions examined in this study.

4.4 Per-Class Classification Performance

To examine performance at the level of individual classes rather than in aggregate, per-class classification reports were computed for each model on the held-out test set. Tables 5 and 6 present the precision, recall, F1-score, and support for the BiLSTM and ResNet-1D models respectively.

Class	Precision	Recall	F1-score	Support
Normal (N)	0.9750	0.9735	0.9742	2000
Supraventricular (S)	0.9975	1.0000	0.9988	2000
Ventricular (V)	0.9550	0.9665	0.9607	2000
Fusion (F)	0.9707	0.9445	0.9574	2000
Unclassifiable (Q)	0.9857	0.9995	0.9926	2000
Accuracy			0.9768	10000
Macro average	0.9768	0.9768	0.9767	10000
Weighted average	0.9768	0.9768	0.9767	10000

Table 5. Per-class classification report — BiLSTM (overall accuracy = 97.68%).

Class	Precision	Recall	F1-score	Support
Normal (N)	0.9939	0.9850	0.9895	2000
Supraventricular (S)	0.9995	1.0000	0.9998	2000
Ventricular (V)	0.9708	0.9795	0.9751	2000
Fusion (F)	0.9803	0.9725	0.9764	2000
Unclassifiable (Q)	0.9926	1.0000	0.9963	2000
Accuracy			0.9874	10000
Macro average	0.9874	0.9874	0.9874	10000
Weighted average	0.9874	0.9874	0.9874	10000

Table 6. Per-class classification report — ResNet-1D (overall accuracy = 98.74%, best model).

The per-class results in Tables 5 and 6 reveal a consistent pattern across both models: the Normal and Unclassifiable classes were classified with very high precision and recall, reflecting the abundance of typical Normal beats and the distinctive, well-separated feature-space location of artefactual Unclassifiable segments following SMOTE balancing. The Ventricular and Fusion classes proved comparatively more difficult, owing to the intermediate, hybrid morphology of fusion beats and their consequent overlap with the ventricular template. Notably, ResNet-1D achieved higher precision and recall than BiLSTM on the Ventricular and Fusion classes specifically—the two classes of greatest clinical importance for arrhythmia monitoring—indicating that its aggregate accuracy advantage is concentrated precisely where it matters most for a deployed cardiac monitoring system, rather than being distributed uniformly, and inconsequentially, across the easy majority classes.

5. Conclusion

This study presented a controlled, quantitative comparison of three deep learning architectures—BiLSTM, ResNet-1D, and the Transformer Encoder—for the automated classification of cardiac arrhythmias from single-lead ECG beat segments. Using a rigorously preprocessed and SMOTE-balanced corpus derived from the MIT-BIH Arrhythmia Database, all three architectures achieved strong aggregate performance, each exceeding 97% test accuracy and 0.998 macro-averaged ROC-AUC, confirming that the three-stage preprocessing pipeline and class-balancing strategy successfully produced a clean, discriminative training corpus from which any of the three architectural paradigms could learn effective decision boundaries. Among the three, ResNet-1D consistently achieved the best performance across every metric examined, including accuracy (98.74%), the Matthews Correlation Coefficient (0.9843), Cohen's kappa (0.9842), and macro-averaged ROC-AUC (0.9995), and it delivered its largest relative advantage specifically on the clinically important Supraventricular, Ventricular, and Fusion classes. These findings suggest that the hierarchical, multi-scale feature extraction enabled by residual skip connections is particularly well matched to the morphological structure of single-beat ECG classification, converging faster and generalising better than either the sequential recurrent processing of the BiLSTM or the data-hungry self-attention mechanism of the Transformer Encoder under a constrained training budget.

These results carry direct implications for the design of wearable and edge-deployable cardiac monitoring systems. By establishing the accuracy ceiling attainable by three representative, uncompressed architectures under tightly controlled conditions, this study provides a reference point against which the accuracy cost of subsequent model compression—through pruning, quantization, or knowledge distillation—can be meaningfully assessed for deployment on resource-constrained wearable hardware. The consistently strong performance of ResNet-1D, combined with its comparatively favourable convergence behaviour and moderate parameter count relative to the Transformer, positions it as a particularly promising candidate architecture for subsequent compression and edge deployment within Internet of Medical Things cardiac monitoring applications.

International Journal for Rapid Research in Engineering Technology & Applied Science (IJRRETAS)

(Peer-Reviewed, Open Access, Fully Refereed International Journal)

Volume 10 issue 11 November 2024 OPEN ACCESS E-ISSN: 2455-4723 Impact Factor 7.95

Several limitations should be acknowledged. The Supraventricular class was represented by only two genuine beats in the raw data, so its synthetic training distribution was derived from a single pair of seed points; the correspondingly high recall reported for this class should therefore be interpreted with appropriate caution pending validation on a larger sample of genuine supraventricular beats. In addition, the study relies on single-beat classification without access to inter-beat timing information, which limits the discriminability of prematurely timed supraventricular beats from morphologically similar normal beats. Future work should extend the present comparison to multi-beat or rhythm-level classification incorporating RR-interval context, evaluate the compressed, quantized variants of the three architectures on genuine microcontroller-class hardware, and validate the findings on independent, multi-centre ECG datasets to establish the generalisability of the reported performance beyond the MIT-BIH corpus.

References

1. Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., Adam, M., Gertych, A., & Tan, R. S. (2017). A deep convolutional neural network model to classify heartbeats. Computers in Biology and Medicine, 89, 389–396.
2. Che, C., Zhang, P., Zhu, M., Qu, Y., & Jin, B. (2021). Constrained transformer network for ECG signal processing and arrhythmia classification. BMC Medical Informatics and Decision Making, 21(1), 184.
3. Chen, C., Hua, Z., Zhang, R., Liu, G., & Wen, W. (2020). Automated arrhythmia classification based on a combination network of CNN and LSTM. Biomedical Signal Processing and Control, 57, 101819.
4. Elreedy, D., & Atiya, A. F. (2019). A comprehensive analysis of synthetic minority oversampling technique (SMOTE) for handling class imbalance. Information Sciences, 505, 32–64.
5. Faust, O., Shenfield, A., Kareem, M., San, T. R., Fujita, H., & Acharya, U. R. (2018). Automated detection of atrial fibrillation using long short-term memory network with RR interval signals. Computers in Biology and Medicine, 102, 327–335.
6. Hannun, A. Y., Rajpurkar, P., Haghpanahi, M., Tison, G. H., Bourn, C., Turakhia, M. P., & Ng, A. Y. (2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nature Medicine, 25(1), 65–69.
7. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 770–778). IEEE.
8. Hu, R., Chen, J., & Zhou, L. (2022). A transformer-based deep neural network for arrhythmia detection using continuous ECG signals. Computers in Biology and Medicine, 144, 105325.
9. Islam, M. S., Hasan, K. F., Sultana, S., Uddin, S., Quinn, J. M. W., & Moni, M. A. (2023). HARDC: A novel ECG-based heartbeat classification method to detect arrhythmia using

hierarchical attention based dual structured RNN with dilated CNN. Neural Networks, 162, 271–287.

10. Kachuee, M., Fazeli, S., & Sarrafzadeh, M. (2018). ECG heartbeat classification: A deep transferable representation. In Proceedings of the 2018 IEEE International Conference on Healthcare Informatics (pp. 443–444). IEEE.
11. Le, M. D., Rathour, V. S., Truong, Q. S., Mai, Q., Brijesh, P., & Le, N. (2021). Multi-module recurrent convolutional neural network with transformer encoder for ECG arrhythmia classification. In 2021 IEEE EMBS International Conference on Biomedical and Health Informatics (pp. 1–5). IEEE.
12. Meng, L., Tan, W., Ma, J., Wang, R., Yin, X., & Zhang, Y. (2022). Enhancing dynamic ECG heartbeat classification with lightweight transformer model. Artificial Intelligence in Medicine, 124, 102236.
13. Mousavi, S., & Afghah, F. (2019). Inter- and intra-patient ECG heartbeat classification for arrhythmia detection: A sequence to sequence deep learning approach. In Proceedings of ICASSP 2019 (pp. 1308–1312). IEEE.
14. Ray, P. P. (2022). A review on TinyML: State-of-the-art and prospects. Journal of King Saud University – Computer and Information Sciences, 34(4), 1595–1623.
15. Ribeiro, A. H., Ribeiro, M. H., Paixão, G. M. M., Oliveira, D. M., Gomes, P. R., Canazart, J. A., ... Ribeiro, A. L. P. (2020). Automatic diagnosis of the 12-lead ECG using a deep neural network. Nature Communications, 11, 1760.
16. Sellami, A., & Hwang, H. (2019). A robust deep convolutional neural network with batch-weighted loss for heartbeat classification. Expert Systems with Applications, 122, 75–84.
17. Siontis, K. C., Noseworthy, P. A., Attia, Z. I., & Friedman, P. A. (2021). Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. Nature Reviews Cardiology, 18(7), 465–478.
18. Varghese, A., Kamal, S., & Kurian, J. (2023). Transformer-based temporal sequence learners for arrhythmia classification. Medical & Biological Engineering & Computing, 61(8), 1993–2000.
19. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998–6008.
20. Wang, B., Liu, C., Hu, C., Liu, X., & Cao, J. (2021). Arrhythmia classification with heartbeat-aware transformer. In Proceedings of ICASSP 2021 (pp. 1025–1029). IEEE.
21. Warden, P., & Situnayake, D. (2019). TinyML: Machine learning with TensorFlow Lite on Arduino and ultra-low-power microcontrollers. O'Reilly Media.

Yildirim, Ö. (2018). A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. Computers in Biology a