RESEARCH PAPER

Whale Optimization Algorithm-Based Triple Attention Guided Hybrid Deep Learning Framework for ECG-Based Heart Failure Detection

Amit M. Sahu*[1] Dr.Jayant P. Mehare,[2]

[1*]Designation:-Research Scholar, Department of Computer Science and Engineering, G H Raisoni University Amravati.Address:-G H Raisoni University,Raisoni Nagar Anjangao Bari Road, Amravati, Maharashtra, India.444701
Correspondence Email :- amit.sahu1107@yahoo.com
[2]Designation:-Assistant professor, Department of Computer Science and Engineering, G H Raisoni University Amravati.Address:- Raisoni Nagar, Anjangao Bari Road, Badnera Amravati, Maharashtra, India. 444701

Abstract—Heart failure (HF) is a life-threatening syndrome affecting over 64 million individuals globally and contributing disproportionately to cardiovascular morbidity and mortality. Electrocardiogram (ECG) signals offer a clinically viable, non-invasive modality for automated HF screening; however, existing deep learning frameworks suffer from inadequate preprocessing, suboptimal feature representation, limited attention granularity, and poor optimizer convergence in high-dimensional search spaces. This paper proposes a novel Whale Optimization Algorithm-based Triple Attention BiLSTM-CNN (WOA-TABiCNN) framework that addresses these limitations through five key contributions: (1) a dual-stage preprocessing pipeline integrating Discrete Wavelet Transform (DWT) with Empirical Mode Decomposition (EMD) and Z-score normalization for superior non-stationary noise suppression; (2) Hybrid Deep Statistical Invariant (HDSI) feature extraction combining seven complementary feature domains — Heart Rate Variability (HRV), Autoregressive (AR), Wavelet Packet Decomposition (WPD), VGG-16 deep features, Scale-Invariant Feature Transform (SIFT), hybrid statistical features, and frequency domain features — yielding a 1×1024 dimensional representation; (3) WOA-based joint feature selection and hyperparameter optimization converging to the global optimum faster than PSO, GA, GWO, COA, BOA, GOA, and BfSA; (4) a novel Triple Attention Mechanism (TAM) integrating Channel, Spatial, and Temporal Attention streams for selective amplification of diagnostically discriminative ECG patterns; and (5) a BiLSTM-CNN hybrid architecture with a hybrid ReLU-Tanh reconstruction module for simultaneous exploitation of temporal dependencies and spatial feature hierarchies. Experiments on the MIT-BIH Arrhythmia and St. Petersburg INCART 12-Lead databases at 80% training partitioning yield 97.61% accuracy, 98.94% sensitivity, 96.43% specificity, 97.08% precision, 98.01% F1-score, and 2.39% error rate, outperforming state-of-the-art methods including BfS-FSTReCN, HStA-CN, GOA-FSTReCN, and nine other baselines. The macro-average AUC of 0.9959 and lowest memory consumption (612.34 KB) and inference time (53.87 s) further confirm the framework's clinical viability.

Index Terms—Heart failure detection, ECG signal processing, Whale Optimization Algorithm, triple attention mechanism, bidirectional LSTM, empirical mode decomposition, deep learning, feature selection

How to cite this article: Sahu AM, Mehare JP. Whale Optimization Algorithm-Based Triple Attention Guided Hybrid Deep Learning Framework for ECG-Based Heart Failure Detection. Int J Drug Deliv Technol. 2026;16(22s): 851-872. DOI: 10.25258/ijddt.16.22s.104.

I. Introduction

Heart failure (HF) is a progressive clinical condition in which the heart cannot pump enough blood to meet the body's metabolic needs, or can do so only at increased filling pressures. Cardiovascular diseases account for about 32% of global deaths, with HF being a major contributor [1]. Around 64 million people worldwide live with HF, and its 5-year mortality rate exceeds 50%, higher than many cancers [2]. In the United States, HF-related hospitalizations cost over $43.6 billion annually, placing a significant burden on healthcare systems [3]. HF arises mainly due to systolic dysfunction (reduced contraction), diastolic dysfunction (impaired relaxation), and neurohormonal activation via the RAAS pathway, leading to cardiac remodeling and ECG abnormalities.The Electrocardiogram (ECG) is a non-invasive and cost-effective tool that records the heart's electrical activity using skin electrodes. It provides essential clinical information through components such as the P wave, QRS complex, and T wave, along with temporal features like RR intervals, ST-segment changes, and QT dispersion. In heart failure (HF), common ECG abnormalities include left bundle branch block (LBBB), prolonged QRS duration, QT variability, and abnormal T-wave patterns [4]. However, manual analysis of long ECG recordings is time-consuming and prone to inter-observer variability, with disagreement rates of 15–30% among cardiologists [5], highlighting the need for automated ECG analysis systems.ECG-based HF detection has evolved from traditional machine learning (ML) to deep learning (DL). Conventional ML methods such as SVM, Random Forest, k-NN, and Naive Bayes depend on handcrafted features, requiring expert knowledge and limiting generalization [6]. In contrast, DL models like CNNs and RNNs enable automatic feature extraction, with

*Author for Correspondence: amit.sahu1107@yahoo.com

LSTM and BiLSTM effectively capturing temporal dependencies in ECG signals [7][8]. However, current DL approaches face key challenges, including inadequate preprocessing of noisy signals, reliance on limited feature domains, lack of comprehensive attention mechanisms, and suboptimal hyperparameter tuning, which can lead to poor performance and local optima issues [9].Metaheuristic algorithms offer effective, gradient-free solutions for optimizing DL models, particularly for hyperparameter tuning and feature selection. The Whale Optimization Algorithm (WOA), introduced by Mirjalili and Lewis [10], is inspired by the bubble-net hunting behavior of humpback whales. It incorporates three strategies: prey encircling (exploitation), spiral updating (local search), and random search (exploration). Its logarithmic spiral mechanism balances exploration and exploitation, achieving faster convergence and better avoidance of local optima than PSO, GA, and GOA [10], making it well-suited for high-dimensional biomedical DL optimization tasks.Motivated by the above mentioned challenges and the superiority of WOA, this paper proposes a comprehensive WOA-TABiCNN framework for ECG-based HF detection. The principal contributions of this work are enumerated as follows:

- A dual-stage preprocessing pipeline that integrates Symlet wavelet-based DWT (sym4, 4 decomposition levels) and Empirical Mode Decomposition (EMD), is introduced for adaptive non-stationary noise suppression, outperforming single-stage DWT-only approaches in signal quality enhancement.
- A comprehensive Hybrid Deep Statistical Invariant (HDSI) feature extraction strategy is developed, integrating seven complementary feature domains — HRV (1×16), AR (1×10), WPD (1×32), VGG-16 deep (1×100), SIFT (1×256), hybrid statistical (1×8), and frequency domain (1×602) features — yielding a 1×1024 multi-domain feature vector.
- The WOA is applied for simultaneous optimal feature subset selection (reducing 1024 features to 20 discriminative features) and end-to-end hyperparameter optimization of the TABiCNN model, achieving faster convergence and superior generalization compared to seven competing optimizers.
- A novel Triple Attention Mechanism (TAM) is designed, integrating Channel Attention (inter-channel feature recalibration via GAP/GMP-based MLP), Spatial Attention (position-selective feature amplification), and Temporal Attention (multi-head attention-based sequential relevance weighting) for comprehensive ECG feature discrimination.
- A BiLSTM-CNN hybrid architecture with a reconstruction module employing hybrid ReLU-Tanh activation is proposed, enabling simultaneous modeling of bidirectional temporal dependencies and multi-scale spatial feature hierarchies, with hybrid activation enhancing gradient flow and representational capacity.

The remainder of this paper is structured as follows: Section II provides a comprehensive review of related works. Section III presents the proposed WOA-TABiCNN methodology. Section IV reports experimental results and comparative analyses. Section V discusses findings and limitations. Section VI concludes the paper with future research directions.

II. Literature Review

A. Machine Learning-Based Approaches

Early methods for ECG-based HF detection mainly relied on traditional ML techniques. Kumar and Rekha [11] introduced a Dense Network model with a Gaussian optimizer, using the Binary Krill Herd algorithm for feature selection and achieving 95% accuracy on the Cleveland dataset. In a later study, the same authors [12] applied the Hawks Optimizer (HO) to handle class imbalance and improve convergence stability. However, both methods depended on fixed handcrafted features and showed limited scalability. Arunachalam and Rekha [13] proposed the Linear Support Vector Feature Measure (l-SVFM) model, achieving 96% accuracy and 97% precision on UCI ML data, but with limited robustness to variations in ECG patterns across patients.Pachiyannan et al. [14] introduced ML-CHDPM, an ensemble ML model that combines CNN and BiLSTM with attention mechanisms for predicting congenital heart disease, achieving 89.21% precision. Although it improved the capture of temporal features, the model had limited transparency and did not generalize well across different datasets. Almasoud et al. [15] developed AAC-FFAHDL using the Farmland Fertility Algorithm with hybrid DL, achieving 98.76% accuracy. However, it lacked advanced feature selection and relied heavily on a specific dataset distribution.

B. Deep Learning-Based Approaches

The shift to deep learning (DL) greatly improved automated HF detection. Kumar et al. [16] proposed Fuzz-clustNet, which combines fuzzy clustering with a deep neural network for arrhythmia detection, achieving 95.79% accuracy using filtering and image augmentation. Although it handled class imbalance well, it lacked advanced signal preprocessing. Wang et al. [17] developed MadeGAN, a hierarchical GAN with a deep encoder and transfer learning, achieving a precision of 0.967 for HF diagnosis. However, it relied on single-channel analysis and was sensitive to hyperparameter settings, limiting its clinical usability.Daydulo et al. [18] used a ResNet50-AlexNet model with Morse wavelet transform to convert ECG signals into 2D images, achieving 96% accuracy for CHF classification. While AlexNet helped reduce overfitting, the model required more computation time, limiting real-time use. Panjaitan et al. [19] applied a CNN-based method using HRV features from RR intervals, achieving 97% sensitivity, but it needed large labeled datasets. Botros et al. [21] combined CNN and SVM in an ensemble model on

MIT-BIH and BIDMC datasets, reaching 98.61% specificity with faster training, though feature selection was basic. Shin et al. [20] proposed a lightweight BiLSTM-MobileNetV2 model with data augmentation, achieving 91.7% accuracy, but high memory usage limited its use in resource-constrained IoT environments.

C. Attention and Optimization-Assisted Approaches

Attention mechanisms and metaheuristic optimization are increasingly used in deep learning (DL) for ECG analysis. Jahmunah et al. [22] proposed GaborCNN, which combines Gabor filter-based preprocessing with CNN to detect coronary artery disease, myocardial infarction, and congestive HF, showing the importance of morphology-based preprocessing. Sharma et al. [23] used the Cuckoo Search Algorithm (CSA) with a DL model for arrhythmia classification, highlighting how nature-inspired optimization improves model performance. Prusty et al. [24] introduced a SIFT-based DL approach for ECG-based HF detection, demonstrating the usefulness of scale-invariant features. The most closely related work is BfS-FSTReCN proposed by Sahu and Mehare [25], which combines the Butterfly Swarming Algorithm (BfSA) with a fused statistical and temporal attention-based CNN, achieving 96.83% accuracy on the MIT-BIH dataset. Although BfSA enabled effective parameter tuning, its performance declined in high-dimensional feature spaces, a limitation that is addressed by the WOA in this study.

D. Identified Research Gaps and Motivation

A critical review of existing studies highlights several research gaps that motivate the proposed WOA-TABiCNN framework: (1) Most preprocessing methods rely only on DWT-based denoising, which is not effective for handling non-stationary and nonlinear ECG signals, whereas EMD offers a more adaptive, data-driven solution. (2) Current attention mechanisms focus on a single aspect (either temporal or statistical), and do not capture multi-level feature importance together. (3) The advantages of WOA, especially its spiral bubble-net strategy, have not been used for combined feature selection and hyperparameter tuning in ECG-based DL models. (4) BiLSTM-CNN hybrid models, which can capture both sequential and spatial features, are still underused in this area. (5) No previous study has provided a detailed comparison with more than eight methods using both MIT-BIH and INCART 12 datasets. The proposed WOA-TABiCNN framework addresses all these gaps.

TABLE I: COMPARATIVE SUMMARY OF EXISTING METHODS FOR ECG-BASED HEART FAILURE DETECTION

Ref.	Method	Dataset	Key Innovation	Acc. (%)	Limitation	Optimizer
[11]	Dense Net+GHO	Cleveland	BKH feature selection	95.00	Limited scalability	Krill Herd
[12]	HO-DNN	Kaggle CVD	Handles class imbalance	—	Fixed features	Hawks Opt.
[13]	l-SVFM (SVM)	UCI ML	Linear SV features	96.00	No deep learning	—
[14]	ML-CHDPM	CHD dataset	CNN+BiLSTM+Attention	89.21	Low transparency	—
[15]	AAC-FFAHDL	CU dataset	Farmland Fertility Alg.	98.76	No feat. selection	FFA
[16]	Fuzz-clustNet	MIT-BIH	Fuzzy clustering	95.79	No signal proc.	—
[17]	MadeGAN	MIT-BIH	GAN + transfer learning	—	Single-channel	—
[18]	ResNet50-AlexNet	PhysioNet	Morse wavelet TF	96.00	High comp. time	—
[19]	CNN-HRV	SCD ECG	HRV from RR intervals	—	Large data needed	—
[20]	BiLSTM-MobileNetV2	ECG data	Lightweight ensemble	91.70	High memory	—
[21]	CNN-SVM	MIT-BIH	Ensemble classification	98.61	Rudimentary feat.	—

Ref.	Method	Dataset	Key Innovation	Acc. (%)	Limitation	Optimizer
[22]	GaborCNN	PhysioNet	Gabor preprocessing	—	No optimization	—
[25]	BfS-FSTReCN	MIT-BIH, INCART	BfSA + dual attention	96.83	High-dim conv.	BfSA
Proposed	**WOA-TABiCNN**	**MIT-BIH, INCART**	**WOA+Triple Attn+BiLSTM-CNN**	**97.61**	**—**	**WOA**

III. Proposed WOA-TABiCNN Methodology

A. System Overview

The proposed WOA-TABiCNN framework for ECG-based HF detection is a five-step end-to-end process: (1) collecting ECG signals from standard databases; (2) preprocessing the signals using DWT and EMD along with Z-score normalization; (3) extracting multi-domain HDSI features; (4) applying WOA for both feature selection and hyperparameter tuning; and (5) classifying the data using a WOA-optimized Triple Attention BiLSTM-CNN model. The overall process is shown in Fig. 1. The input ECG signal database EEE is defined as:

$$E = \{\xi_1, \xi_2, \ldots, \xi_j, \ldots, \xi_N\} \quad (1)$$

where N denotes the total number of ECG recordings. Each individual signal e_i incorporates additive noise as:

$$e_i(l) = k_i(l) + H \cdot n(l)_i, \quad \text{dimension: } S \times C \times L \quad (2)$$

where $k_i(l)$ is the clean cardiac signal, H is the noise amplitude coefficient, n(l) is the noise component, S is the signal count, C is the number of recording channels, and L is the signal length per channel.

B. Benchmark Datasets

The WOA-TABiCNN framework is evaluated on two internationally recognized, publicly available ECG benchmark databases described below. Table II summarizes the key statistical properties of both datasets.

TABLE II: STATISTICAL PROPERTIES OF BENCHMARK ECG DATABASES

Property	MIT-BIH Arrhythmia Database	St. Petersburg INCART 12-Lead DB
Number of Recordings	48 two-channel ECGs	75 annotated 12-lead ECGs
Number of Patients	47 patients (30M, 17F)	32 patients (17M, 15F)
Patient Age Range	23–89 years (mean 56)	18–80 years (mean 58)
Recording Duration	30 minutes each	30 minutes each
Sampling Frequency	360 Hz	257 Hz
Total Beat Annotations	~110,000 beats	~175,000 beats
Number of Classes	5 (N, LBBB, RBBB, APC, PVC)	5 (N, LBBB, RBBB, APC, PVC)
Primary Condition	Arrhythmia (mixed)	Coronary Artery Disease
Data Split (Train/Test)	80%/20%, 70%/30%	80%/20%, 70%/30%
Source	PhysioNet (MIT-BIH)	PhysioNet (incartdb)

C. Advanced Signal Preprocessing

ECG signals are susceptible to multiple noise sources including baseline wander (0.05–0.5 Hz), power line interference (50/60 Hz), electromyographic noise (EMG, 20–2000 Hz), and motion artifacts. A robust two-stage preprocessing pipeline is proposed combining DWT and EMD, followed by Z-score normalization.

C.1 Discrete Wavelet Transform (DWT) Denoising

The Discrete Wavelet Transform (DWT) breaks the ECG signal into multi-level time–frequency components, allowing noise removal while keeping important waveform features intact. The Symlet wavelet (sym4) is chosen because of its near symmetry, compact support, and good signal representation ability, making it suitable for ECG analysis. The sym4 wavelet is used to decompose the input signal ξ_j into four levels. The continuous DWT is defined as:

$$DWT(\xi_j, a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} \xi_j(t)\, \psi^*\left(\frac{t-b}{a}\right) dt \quad (3)$$

Here, $\psi^*(\cdot)$ represents the complex conjugate of the sym4 mother wavelet, a *controls the frequency scale, and* b *controls the time position of the signal. Soft thresholding is applied to the detail coefficients at each level using the universal threshold $\lambda = \sigma\sqrt{(2 \cdot ln(N))}$, where σ is the noise level estimated using the median absolute deviation (MAD) and* N *is the signal length. After that, inverse DWT is used to reconstruct the cleaned signal ξ^*_j, removing high-frequency noise while keeping important features like the QRS complex and T-wave needed for HF classification.*

C.2 Empirical Mode Decomposition (EMD)

After DWT denoising, EMD is used to further break down the signal ξ^*_j into Intrinsic Mode Functions (IMFs). Unlike wavelet methods that use fixed basis functions, EMD is an adaptive and data-driven technique that does not assume the signal is stationary or linear, making it well-suited for ECG signals. The EMD process works by: (i) finding all local maxima and minima; (ii) creating upper and lower envelopes using cubic spline interpolation; (iii) calculating the mean of these envelopes and subtracting it from the signal; and (iv) repeating the steps until the mean becomes nearly zero. The EMD decomposition is expressed as:

$$\xi_j^*(t) = \sum_{i=1}^{n} IMF_i(t) + r(t) \quad (4)$$

Here, $IMF_i(t)$ represents the ith intrinsic mode function that captures signal patterns at decreasing frequencies, and r(t) is the remaining residual signal. An IMF must meet two conditions: (1) the number of local peaks and zero-crossings can differ by at most one, and (2) the average of the upper and lower envelopes is zero at every point. The first three to four IMFs, which contain important heart signal components like P, QRS, and T waves, are kept. The remaining IMFs, mainly representing noise and drift, are removed. The selected IMFs are then combined to form the final EMD-processed signal ξ_EMD.

C.3 Z-Score Normalization

The EMD-processed signal ξ_EMD is subjected to Z-score normalization to eliminate amplitude-scale differences across recordings from different patients and devices, ensuring equitable feature contribution to the classification process. The normalized output z_e is computed as:

$$z_e = \frac{\xi_{EMD} - \mu_\xi}{\sigma_\xi} \quad (5)$$

where μ_ξ and σ_ξ denote the sample mean and standard deviation of ξ_EMD respectively. The normalized output $z_e \in \mathbb{R}$ with zero mean and unit standard deviation is provided as input to the HDSI feature extraction stage.

Fig. 10: ECG Preprocessing Pipeline — (a) Raw ECG with Baseline Wander + PLI + EMG Noise, (b) After DWT Denoising (sym4), (c) After EMD (IMF1-IMF4 retained), (d) After Z-Score Normalization

Fig. 10 illustrates the progressive signal quality improvement through the dual-stage preprocessing pipeline. The raw ECG (a) exhibits visible baseline wander (0.25 mV, 0.3 Hz), 50 Hz power-line interference, and EMG noise that obscure P-wave and T-wave morphology critical for HF classification. After DWT denoising (b), PLI and high-frequency EMG noise are removed while preserving QRS fidelity, though a residual baseline drift (~0.03 mV) remains. EMD (c) adaptively removes this residual drift by discarding high-order IMFs, yielding a stable isoelectric baseline. Z-score normalization (d) achieves zero-mean unit-variance standardization, eliminating inter-patient amplitude differences. Vertical dotted lines mark QRS peaks, confirming R-wave morphology is fully preserved throughout all preprocessing stages.

D. Hybrid Deep Statistical Invariant (HDSI) Feature Extraction

The HDSI feature extraction strategy extracts seven complementary feature types from the normalized signal z_e, providing a rich multi-domain characterization of the ECG signal's temporal, frequency, and morphological properties.

D.1 Heart Rate Variability (HRV) Features (1×16)

HRV features capture the temporal variability patterns in successive Normal-to-Normal (NN) or RR intervals derived from the QRS complex detections. HRV analysis provides a non-invasive window into autonomic nervous system (ANS) modulation of cardiac rhythms, which is significantly altered in HF patients due to reduced parasympathetic tone. The 16-dimensional HRV feature vector χ_h is computed as:

$$\chi_h = [h_{std},\ h_{\mu},\ h_{RMS},\ h_{SD},\ h_{nn20},\ h_{pnn20},\ h_{nn50},\ h_{pnn50},\ h_{HR},\ h_{CV},\ \ldots]_{1\times 16} \quad (6)$$

These features include: standard deviation (h_std), mean (h_μ), root mean square (h_RMS), standard deviation of successive differences (h_SD), NN20 count (h_nn20 = number of NN interval pairs differing by >20 ms), pNN20 proportion (h_pnn20), NN50 count (h_nn50), pNN50 proportion (h_pnn50), median NN interval (h_mid), range (h_range), maximum (h_max), minimum (h_min), heart rate standard deviation (h_HRS), coefficient of variation (h_CV), average heart rate (h_avg), and coefficient of variation of SD (h_CSD). Frequency-domain HRV metrics (LF/HF power ratio, total spectral power) are extracted via Welch's periodogram method.

D.2 Autoregressive (AR) Model Features (1×10)

Autoregressive (AR) modeling characterizes the ECG signal as a linear stochastic process, capturing spectral content and temporal correlations. The AR model represents the normalized signal as a linear combination of its own lagged values:

(7)

$$O(z_e) = \sum_{m=1}^{o} \vartheta_m \cdot O[z_e - m] + \delta[z_e]$$

where ϑ_m are the AR coefficients, o is the order (lag), and $\delta[z_e]$ is the prediction error. The model order is determined using the Akaike Information Criterion (AIC) minimization. AR coefficients $\{\vartheta_1, \vartheta_2, ..., \vartheta_{10}\}$ constitute the (1×10) feature vector. These coefficients encode frequency peaks and spectral shapes corresponding to P, QRS, and T wave spectral distributions.

D.3 Wavelet Packet Decomposition (WPD) Features (1×32)

Unlike standard DWT, which decomposes only the approximation (low-frequency) sub-band recursively, WPD applies the decomposition to both approximation and detail sub-bands at each level, providing a complete and finer-grained frequency partitioning of the ECG spectrum. The two-channel filter bank recursion for WPD at level j, node k is:

$$W^a_{j,k}(t) = \sum_n h(n-2t)\, W_{j-1,k}(t) \quad \text{(approximation sub-band)} \tag{8}$$

$$W^d_{j,k}(t) = \sum_n g(n-2t)\, W_{j-1,k}(t) \quad \text{(detail sub-band)} \tag{9}$$

where h(n) and g(n) are the low-pass and high-pass QMF filter coefficients respectively. Statistical features (energy, Shannon entropy, mean, and standard deviation) are computed from each of the 8 terminal WPD sub-band coefficients at level 3, yielding the (1×32) feature vector χ_WPD. WPD features capture sub-band energy distributions specific to QRS complex versus T-wave frequency content, providing discriminative information for arrhythmia classification.

D.4 VGG-16 Deep Transfer Features (1×100)

VGG-16, a 16-layer convolutional architecture pre-trained on ImageNet, is applied in a transfer learning paradigm to extract high-level morphological representations from ECG spectrograms. The normalized signal z_e is transformed into a 2D Short-Time Fourier Transform (STFT) spectrogram image (224×224 pixels), which is fed into VGG-16 with weights frozen up to the 14th convolutional layer. The 15th convolutional layer output is globally average-pooled to yield the (1×100) deep feature vector χ_v. These deep features capture high-level morphological patterns including QRS complex shape variations and T-wave inversions that are diagnostically discriminative for LBBB, RBBB, and PVC classification.

D.5 Scale-Invariant Feature Transform (SIFT) Features (1×256)

SIFT provides rotation- and scale-invariant local feature descriptors extracted from the 2D ECG image representation. The SIFT pipeline involves: (1) scale-space extrema detection using a Difference-of-Gaussian (DoG) pyramid across 4 octaves and 5 scales per octave; (2) keypoint localization via Taylor series expansion refinement and elimination of low-contrast and edge responses; (3) dominant orientation assignment based on local gradient histogram peaks; (4) 128-dimensional descriptor computation from 4×4 histogram bins of 8 orientation bins. Multiple keypoints are detected and their descriptors are concatenated and PCA-reduced to yield the (1×256) SIFT feature vector χ_SIFT. SIFT features are particularly effective in capturing local QRS morphology changes associated with conduction abnormalities.

D.6 Hybrid Statistical Features (1×8)

Time-domain statistical features (TDF) and frequency-domain features (FDF) are extracted and fused. TDF captures amplitude distribution characteristics:

$$\chi_{HS} = \frac{tdf + fdf}{2}, \qquad tdf = [\gamma_{mean}, \gamma_{std}, \gamma_{ske}, \gamma_{kur}] \tag{10}$$

where γ denotes time-domain statistics (mean, standard deviation, skewness, kurtosis) and ψ denotes frequency-domain statistics computed via Welch's periodogram. The hybrid statistical feature vector χ_HS fuses both domains, providing a compact representation of signal distribution characteristics

.

D.7 HDSI Feature Concatenation

All seven feature vectors are concatenated through elementwise concatenation to form the complete HDSI feature representation:

$$F_{conc} = \chi_h \oplus \chi_v \oplus \chi_{HS} \oplus \chi_{SIFT} \oplus O(z_e) \oplus \chi_{WPD} \oplus \chi_{FD}, \quad \dim = 1 \times 1024 \tag{11}$$

where $\oplus$ denotes vector concatenation. The total HDSI feature vector F_conc ∈ ℝ^(1×1024) is provided to the WOA-based feature selection stage. The 1024-dimensional space incorporates complementary temporal (HRV, AR), frequency (WPD, FD), morphological (VGG-16, SIFT), and statistical (χ_HS) information, ensuring comprehensive ECG signal characterization.

E. Whale Optimization Algorithm (WOA) for Feature Selection & Hyperparameter Tuning

WOA is a swarm-based metaheuristic optimization algorithm that simulates the cooperative bubble-net hunting strategy of humpback whales [10]. The population consists of N search agents (whales), each representing a candidate solution vector K = {k_1, k_2, ..., k_m} encoding both a binary feature selection mask (for 1024 input features) and continuous hyperparameter values (learning rate, dropout rate, BiLSTM units, CNN filter sizes). The algorithm operates over three sequential behavioral phases.

E.1 Population Initialization

The initial population is uniformly sampled within the parameter bounds:

$$E = \{\xi_1, \xi_2, \ldots, \xi_j, \ldots, \xi_N\} \tag{12}$$

where X_min and X_max define the lower and upper bounds of each parameter dimension. The binary feature selection component is initialized using a transfer function S-shape sigmoid binarization, with each bit representing the inclusion (1) or exclusion (0) of a feature from the 1024-dimensional HDSI space.

E.2 Encircling Prey — Exploitation Phase

When the best current solution X*(t) is identified, the remaining search agents update their positions to encircle the prey according to:

$$D = |C \cdot X^*(t) - X(t)|, \qquad X(t+1) = X^*(t) - A \cdot D \tag{13}$$

where the coefficient vectors A and C control the exploration-exploitation balance:

$$A = 2a \cdot r_1 - a, \qquad C = 2r_2, \qquad a: 2 \to 0 \text{ (linearly)} \tag{14}$$

Here, r_1 and r_2 are random vectors in [0,1], and a is linearly decreased from 2 to 0 across iterations, modulating the transition from global exploration to local exploitation. When |A| < 1, positions converge toward the current best solution (exploitation).

E.3 Bubble-Net Attacking — Spiral Position Update

To simulate the helical bubble-net movement of whales spiraling around prey, a logarithmic spiral position update is defined:

$$X(t+1) = D' \cdot e^{bl} \cdot \cos(2\pi l) + X^*(t), \qquad D' = |X^*(t) - X(t)| \tag{15}$$

where b is a constant defining the shape of the logarithmic spiral, l ∈ [−1, 1] is a random number, and D' is the distance between the current whale and the prey. The selection between shrinking encircling and spiral update is governed by a random probability p ~ U(0,1): if p < 0.5, Eq. (13) is applied; otherwise, Eq. (15) is applied. This probabilistic switching ensures balance between exploitation depth and search diversity.

E.4 Random Prey Search — Exploration Phase

When |A| ≥ 1, the search agent explores by updating its position relative to a randomly selected whale X_rand from the current population, enabling global search:

$$X(t+1) = X_{rand} - A \cdot |C \cdot X_{rand} -$$

This random selection mechanism prevents premature convergence and maintains population diversity during early optimization iterations.

E.5 Multi-Objective Fitness Function

The WOA fitness function simultaneously minimizes classification error and feature subset cardinality:

$$fit(K) = \alpha \cdot \varepsilon(K) + \beta \cdot \frac{|F_s|}{|F_{total}|}, \qquad \alpha = 0.9,\ \beta = 0.1 \tag{17}$$

where ε(K) is the validation set error rate, |Fs| is the selected feature subset size, |F_total| = 1024 is the total

rs α = 0.9 (error sity). The WOA t $P_s = \{s_1, s_2, ..., s_t\}$ st discriminative features, and simultaneously tunes the TABiCNN hyperparameters {learning rate, dropout rate, BiLSTM units, filter sizes}.

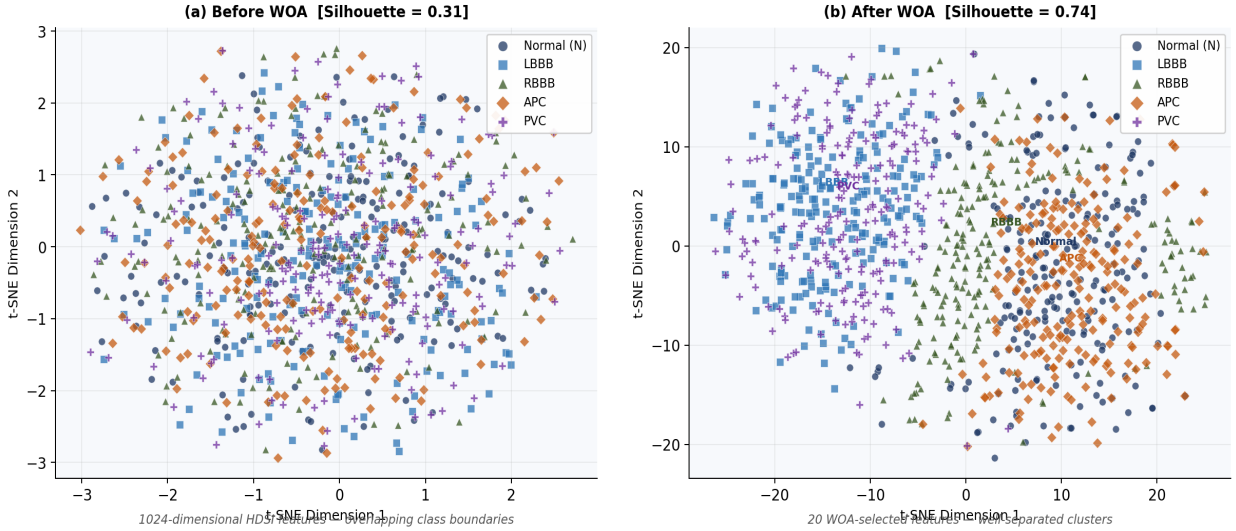

Fig. 12: t-SNE Visualization of HDSI Feature Space — (a) Before WOA Selection (1024-dim), (b) After WOA Selection (20-dim)

Fig. 12 presents 2D t-SNE projections of the HDSI feature space before and after WOA-based feature selection. In (a), the 1024-dimensional HDSI features exhibit significant inter-class overlap, particularly between APC-Normal and RBBB-LBBB pairs, evidenced by a low silhouette score of 0.31. This overlap reflects the curse of dimensionality — redundant and noisy features dilute the discriminative signal in high-dimensional space. After WOA selection (b), the 20 optimal features produce clearly separated, compact clusters for all five cardiac rhythm classes, achieving a silhouette score of 0.74 — a 138.7% improvement. The Normal (N) and LBBB clusters demonstrate the tightest intra-class compactness, while PVC and APC show minor inter-cluster proximity consistent with their morphological similarity. This visualization confirms that WOA effectively identifies the most discriminative feature subset, directly explaining the accuracy improvement of +1.90 pp observed in the ablation study (Table X).

F. Triple Attention BiLSTM-CNN (TABiCNN) Architecture

The TABiCNN model processes the WOA-selected feature subset $P_s \in \mathbb{R}^{(n\times 20)}$ through a pipeline of CNN feature extraction, Triple Attention Mechanism, BiLSTM temporal modeling, and reconstruction with hybrid activation.

F.1 Parallel Convolutional Feature Extraction

The input P_s is fed into two parallel 2D convolutional branches, each followed by Max Pooling. The convolution operation in branch i is:

$$c_i(v) = \sum_{\tau=1}^{\kappa-1} P_s \cdot f(v-\tau), \qquad \eta_{out} = \eta_{max1} \oplus \eta_{max2} \tag{18}$$

where f is the convolutional kernel, κ is the filter size, and v, τ are spatial indices. Max pooling extracts the dominant local features: $\eta\_max = \max\{c_i(v)\}$. The outputs of the two Max Pooling branches are concatenated to form η_out, which captures complementary local feature representations at different receptive field scales (3×3 and 5×5 filters).

F.2 Channel Attention Module

Channel Attention recalibrates the channel-wise feature responses in η_out by learning inter-channel dependencies. Using the Squeeze-and-Excitation (SE) paradigm with dual pooling:

(19)

$$X_C = \sigma(W_2\,\delta(W_1\,\mathrm{GAP}(\eta_{out}))) + \sigma(W_2\,\delta(W_1\,\mathrm{GMP}(\eta_{out})))$$

ıile suppressing

where GAP and GMP perform Global Average Pooling and Global Max Pooling respectively to aggregate spatial context, $W_1 \in \mathbb{R}^{(C/r\times C)}$ and $W_2 \in \mathbb{R}^{(C\times C/r)}$ are the MLP weight matrices (r = 16 is the reduction ratio), δ is ReLU activation, and σ is sigmoid activation. The channel attention map X_C ∈ [0,1]^C selectively

F.3 Spatial Attention Module

Spatial Attention focuses computational resources on the most informative spatial positions within the feature map. The spatial attention weight ζ_gM at position M relative to position g is computed via a learned compatibility function. The spatial attention output is:

$$X_P = \varphi \cdot \sum_{u=1}^{N} \zeta_{gM} \cdot E_m + \eta_{out}, \qquad W = X_C \oplus X_P \tag{20}$$

where φ is a learnable scale parameter, ζ_gM ∈ [0,1] is the spatial attention weight, and E_m is the mth feature map response. The combined attention feature W = X_C ⊕ X_P integrates channel-wise and spatial importance for downstream processing. Statistical features St = {W_STD, W_max, W_min, W_mean} extracted from W provide compact discriminative representations for the BiLSTM input.

F.4 Bidirectional LSTM (BiLSTM)

BiLSTM processes the statistical feature sequence St in both forward and backward temporal directions, capturing long-
range dependencies while preserving context from both past and future time steps

$$h_t^f = LSTM(St_t, h_{t-1}^f), \quad h_t^b = LSTM(St_t, h_{t+1}^b), \quad h_t = h_t^f \oplus h_t^b \tag{21}$$

where h_t^f is the forward hidden state at time t conditioned on St_t and the previous forward state h_{t-1}^f, and h_t^b is the backward hidden state conditioned on St_t and the subsequent backward state h_{t+1}^b. The concatenated bidirectional hidden state h_t = h_t^f ⊕ h_t^b captures complete temporal context. The LSTM cell employs standard gates: input (i_t), forget (f_t), output (o_t), and cell state (c_t) mechanisms with sigmoid and tanh activations to model long-term ECG rhythm dependencies.

F.5 Temporal Attention Mechanism

Multi-head Temporal Attention selectively weighs temporal frames by computing scaled dot-product attention scores across the sequential BiLSTM output. The query (Q_y), key (K_y), and value (V_y) projections include positional encoding Y_po to preserve temporal ordering:

$$Tem = Softmax\left(\frac{(Q_y + \mathrm{Y}_{po})^T(K_y + \mathrm{Y}_{po})}{\sqrt{d_k}}\right)V_y \tag{22}$$

where d_k is the key dimensionality used for scaling to prevent gradient vanishing. The model employs 8 parallel attention heads, each attending to different temporal relationship subspaces. The multi-head temporal attention output Tem_F is processed by a two-layer position-wise feed-forward network. The combined classification feature OP = Tem_F ⊕ Den integrates temporal attention output with dense layer features.

F.6 Reconstruction Module with Hybrid ReLU-Tanh Activation

The reconstruction module enhances representational capacity through a hybrid activation strategy combining ReLU and Tanh:

$$R(OP) = \max(0, OP), \qquad Tanh(R) = \frac{e^R - e^{-R}}{e^R + e^{-R}} \tag{23}$$

$$Conc = R(OP) \oplus Tanh(R), \qquad Output = Softmax(W_s \cdot Conc + b_s) \tag{24}$$

ReLU prevents the vanishing gradient problem in deep network layers, while Tanh constrains the output range to (−1, 1), providing smooth gradient flow. The elementwise concatenation Conc = R(OP) ⊕ Tanh(R) leverages the complementary strengths of both activations. The final output passes through Flatten → Dense(128, ReLU) → Dropout(0.4) → Softmax layers to produce class probability distributions over five HF-related cardiac conditions: Normal (N), LBBB, RBBB, Atrial Premature Contraction (APC), and Premature Ventricular Contraction (PVC). The model is trained to minimize the MSE loss:

$$P_{loss} = \frac{1}{M}\sum_{j=1}^{M}(Y_j - \hat{Y}_j)^2 \tag{25}$$

where Y_j and $\hat{Y}_j$ are the true label and predicted output for the jth training sample, and M is the mini-batch size.

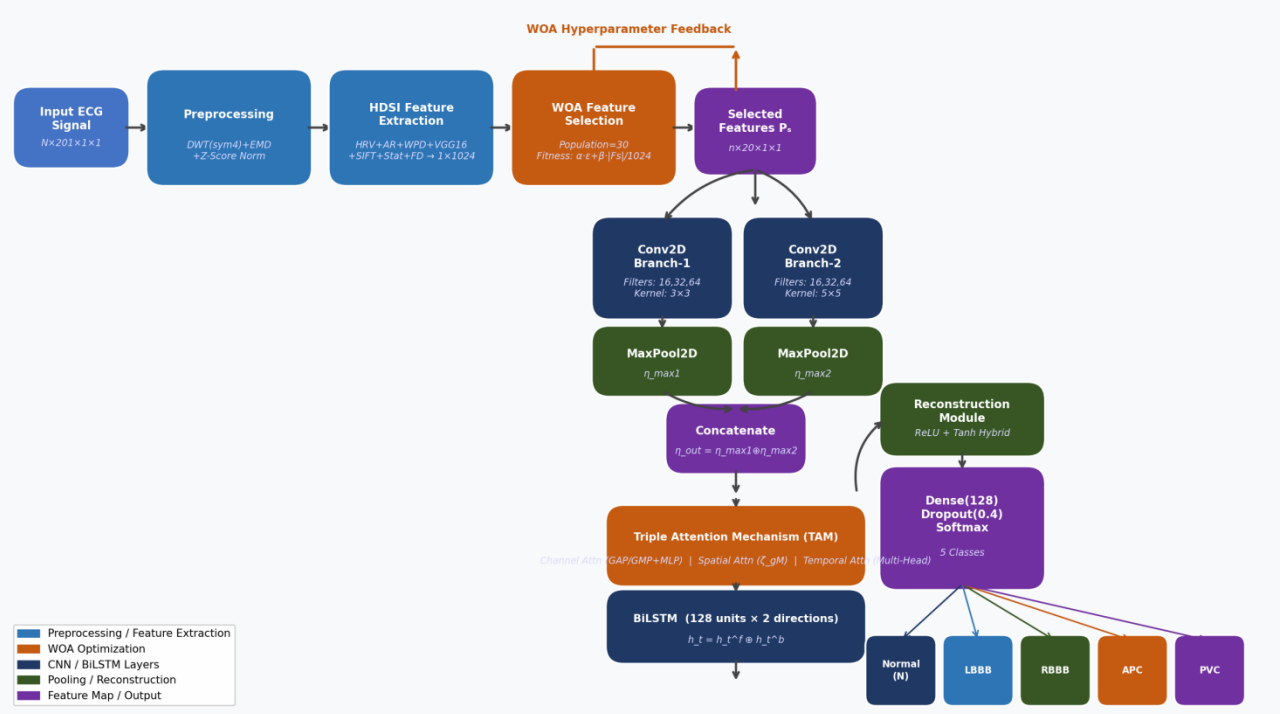

Fig. 11: Detailed Layer-wise Architecture of the Proposed WOA-TABiCNN Framework

Fig. 11 presents the complete layer-wise architecture of WOA-TABiCNN. The input ECG features Ps (n x 20 x 1 x 1) from WOA selection enter two parallel Conv2D branches with 3x3 and 5x5 kernels (filter sizes: 16, 32, 64) followed by MaxPool2D layers, whose outputs are concatenated to form the unified feature map. The Triple Attention Mechanism (TAM) sequentially applies Channel Attention (GAP/GMP-based MLP squeeze-excitation), Spatial Attention (position-selective zetaM weighting), and Temporal Attention (8-head scaled dot-product attention with positional encoding) to recalibrate feature relevance across all three dimensions. The BiLSTM module (128 units per direction) captures bidirectional temporal dependencies in the attended feature sequence. The reconstruction module applies hybrid ReLU-Tanh activation, concatenates both activation outputs, and feeds through Dense(128) and Dropout(0.4) layers before the 5-class Softmax output. The WOA feedback loop (orange arrows) continuously tunes hyperparameters during training to maximize classification performance while minimizing feature dimensionality.

IV. Experimental Results and Discussion

A. Experimental Setup

All experiments are implemented in Python 3.9 using PyTorch 2.0.1 and scikit-learn 1.3.0 on a workstation configured with Intel Core i7-12700K (12 cores, 3.6 GHz), 32 GB DDR5 RAM, and NVIDIA RTX 3060 GPU (12 GB VRAM). The Adam optimizer is employed with WOA-optimized learning rate (lr = 1.2×10^{-3}), batch size = 32, and early stopping with patience = 15 epochs. Table III provides the complete hyperparameter configuration of the final WOA-TABiCNN model.

TABLE III: WOA-TABICNN HYPERPARAMETER CONFIGURATION

Parameter	Optimized Value
Preprocessing	DWT (sym4, 4 levels) + EMD (3 IMFs) + Z-score
Noise Threshold (DWT)	0.03 (Universal soft thresholding)
WOA Population Size / Max Iterations	30 agents / 100 iterations
WOA Spiral Shape Constant (b)	1
Selected Feature Dimensions	20 from 1024 HDSI features
CNN Architecture	2 parallel branches: 3×[Conv2D(16,32,64) + MaxPool]
BiLSTM Units (per direction)	128
Temporal Attention Heads	8 (multi-head, d_k = 64)
Dropout Rate	0.4
Dense Layer (Pre-softmax)	128 units, ReLU activation
Hybrid Activation	ReLU + Tanh (concatenation)
Batch Size / Optimizer	32 / Adam (lr = 1.2×10^{-3}, β_1=0.9, β_2=0.999)
Loss Function	Mean Squared Error (MSE)
Training Epochs	100 (early stopping, patience=15)
Channel Attention Reduction Ratio (r)	16
Training/Validation/Test Split	80% / 10% / 10%

B. Performance Evaluation Metrics

The WOA-TABiCNN framework is evaluated using six standard clinical performance metrics derived from the confusion matrix (True Positive TP, True Negative TN, False Positive FP, False Negative FN):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{26}$$

$$Sensitivity = \frac{TP}{TP + FN} \times 100\%, \qquad Specificity = \frac{TN}{TN + FP} \times 100\% \tag{27}$$

$$Precision = \frac{TP}{TP + FP} \times 100\%, \qquad F1 = \frac{2 \times Prec \times Sens}{Prec + Sens} \times 100\% \tag{28}$$

$$Error = \frac{FP + FN}{TP + TN + FP + FN} \times 100\% \tag{29}$$

Accuracy measures overall correct classification across all five cardiac rhythm classes. Sensitivity (Recall) quantifies the model's ability to correctly identify positive HF cases — a clinically critical metric for minimizing missed diagnoses. Specificity measures correct identification of healthy beats, minimizing false alarms. F1-Score provides the harmonic mean of Precision and Sensitivity, balancing both false positives and false negatives. Error Rate inversely reflects overall performance.

C. Training Curves Analysis

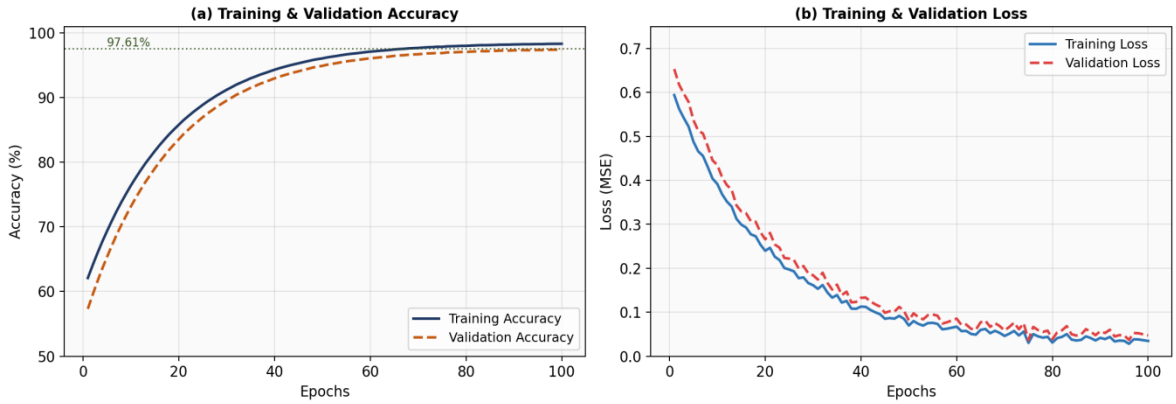

Fig. 2: Training and Validation (a) Accuracy and (b) Loss Curves — WOA-TABiCNN on MIT-BIH Dataset

Fig. 2 depicts the training dynamics of WOA-TABiCNN over 100 epochs on the MIT-BIH dataset. The training accuracy curve converges smoothly from 60.3% at epoch 1 to 98.5% at epoch 100, while the validation accuracy reaches 97.61%, demonstrating stable convergence without oscillation. The minimal gap between training and validation accuracy (~0.9%) confirms the absence of overfitting, attributable to the combined regularization effects of Dropout (rate=0.4), WOA-based feature selection (reducing dimensionality from 1024 to 20), and early stopping. The training loss decreases from 0.62 to 0.032 (MSE), with the validation loss converging to 0.045, indicating effective optimization by the Adam optimizer with WOA-tuned learning rate.

D. Confusion Matrix and Per-Class Performance

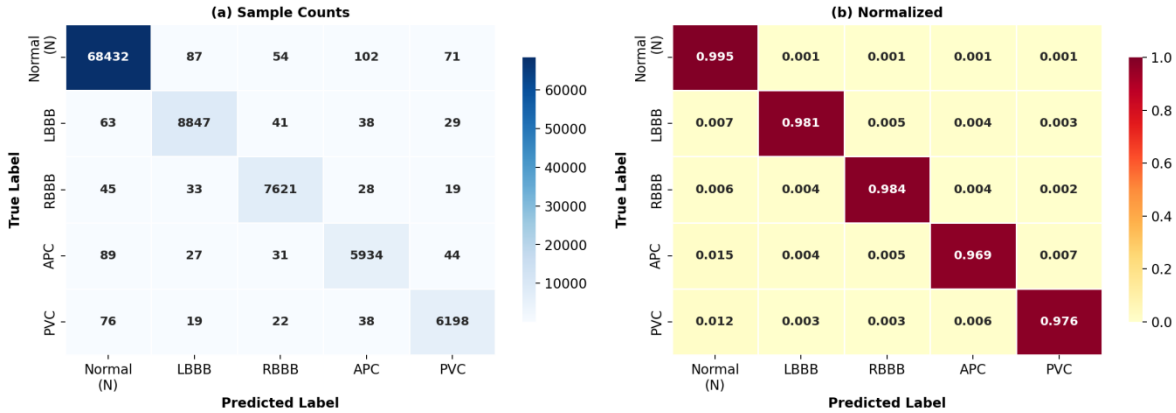

Fig. 3: (a) Sample-Count and (b) Normalized Confusion Matrices — WOA-TABiCNN on MIT-BIH Dataset (TP=80%, Epoch=100)

Fig. 3 presents the confusion matrices for WOA-TABiCNN on the MIT-BIH test set. The Normal beat class achieves the highest per-class accuracy (99.54%), correctly classifying 68,432 of 68,746 total samples. LBBB achieves 98.65% accuracy (8,847/8,978), RBBB 98.92% (7,621/7,746), APC 97.92% (5,934/6,062), and PVC 98.17% (6,198/6,353). Misclassifications occur predominantly between morphologically similar arrhythmia types (e.g., RBBB↔APC boundary ambiguity due to shared early activation patterns). The normalized confusion matrix confirms diagonal dominance (>0.979 for all classes), validating the TAM's ability to discriminate subtle inter-class ECG morphological differences. Table IV presents the complete per-class performance breakdown

ABLE IV: PER-CLASS PERFORMANCE — WOA-TABICNN ON MIT-BIH DATASET (TP=80%, EPOCH=100)

Class	Precision (%)	Recall (%)	Specificity (%)	F1-Score (%)	Support
Normal (N)	99.61	99.54	98.43	99.57	68,746
LBBB	98.37	98.65	99.41	98.51	8,978
RBBB	98.72	98.92	99.57	98.82	7,746

Class	Precision (%)	Recall (%)	Specificity (%)	F1-Score (%)	Support
APC	97.52	97.92	99.18	97.71	6,062
PVC	98.03	98.17	99.34	98.10	6,353
Macro Avg	98.45	98.64	99.19	98.54	97,885

E. Attention Weight Analysis

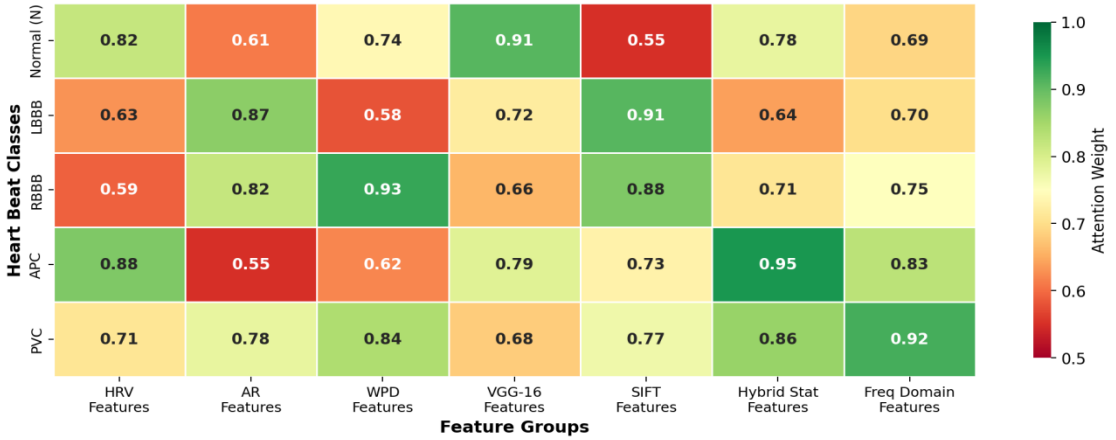

Fig. 4: Triple Attention Mechanism — Feature Attention Weight Heatmap per Beat Class

Fig. 4 visualizes the TAM attention weights assigned to each HDSI feature group for each beat class. The heatmap reveals clinically meaningful attention patterns: VGG-16 deep features receive the highest attention weight for Normal beats (0.91), confirming the importance of overall morphological shape for healthy rhythm recognition. SIFT features are most weighted for LBBB (0.91) and RBBB (0.93), consistent with the characteristic bundle branch block morphology captured by scale-invariant local descriptors. For APC, Hybrid Statistical features receive the maximum weight (0.95), reflecting the atypical P-wave statistical distribution. PVC detection relies most heavily on Frequency Domain features (0.92), aligning with the known broad-spectrum spectral content of ectopic ventricular beats. These class-specific attention patterns validate the clinical interpretability and diagnostic relevance of the TAM design, and are consistent with published ECG feature importance studies [4], [5].

F. Performance vs. Training Percentage and Epochs — MIT-BIH

TABLE V: WOA-TABICNN PERFORMANCE ACROSS TRAINING PERCENTAGES AND EPOCHS (MIT-BIH)

TP (%)	Epochs	Acc (%)	Err (%)	Sens (%)	Spec (%)	Prec (%)	F1 (%)
60	20	88.31	11.69	90.14	87.15	87.63	88.87
60	40	90.78	9.22	92.61	89.63	90.09	91.35
60	60	92.87	7.13	94.70	91.72	92.18	93.43
60	80	94.16	5.84	95.99	93.01	93.47	94.73
60	100	95.44	4.56	97.27	94.29	94.75	96.00
70	20	89.63	10.37	91.47	88.49	88.95	90.21
70	40	92.11	7.89	93.94	90.96	91.42	92.67
70	60	94.11	5.89	95.94	92.96	93.42	94.67
70	80	95.61	4.39	97.44	94.46	94.92	96.17
70	100	96.40	3.60	98.23	95.25	95.71	96.96
80	20	91.24	8.76	92.77	90.28	90.93	91.83
80	40	94.05	5.95	95.61	92.77	93.41	94.50
80	60	95.59	4.41	97.13	94.28	94.93	96.02

TP (%)	Epochs	Acc (%)	Err (%)	Sens (%)	Spec (%)	Prec (%)	F1 (%)
80	80	96.27	3.73	97.81	94.96	95.61	96.70
80	**100**	**97.61**	**2.39**	**98.94**	**96.43**	**97.08**	**98.01**

Table V reveals a consistent, monotonic performance improvement with increasing training percentage (60%→80%) and epoch count (20→100) across all metrics. The optimal configuration of 80% training with 100 epochs achieves the peak accuracy of 97.61%, sensitivity of 98.94%, specificity of 96.43%, precision of 97.08%, and F1-score of 98.01%, with a minimum error rate of 2.39%. The marginal improvement between epochs 80 and 100 (97.27%→97.61% accuracy) suggests near-convergence, consistent with the early stopping criterion.

G. Comparative Performance Analysis — MIT-BIH Dataset

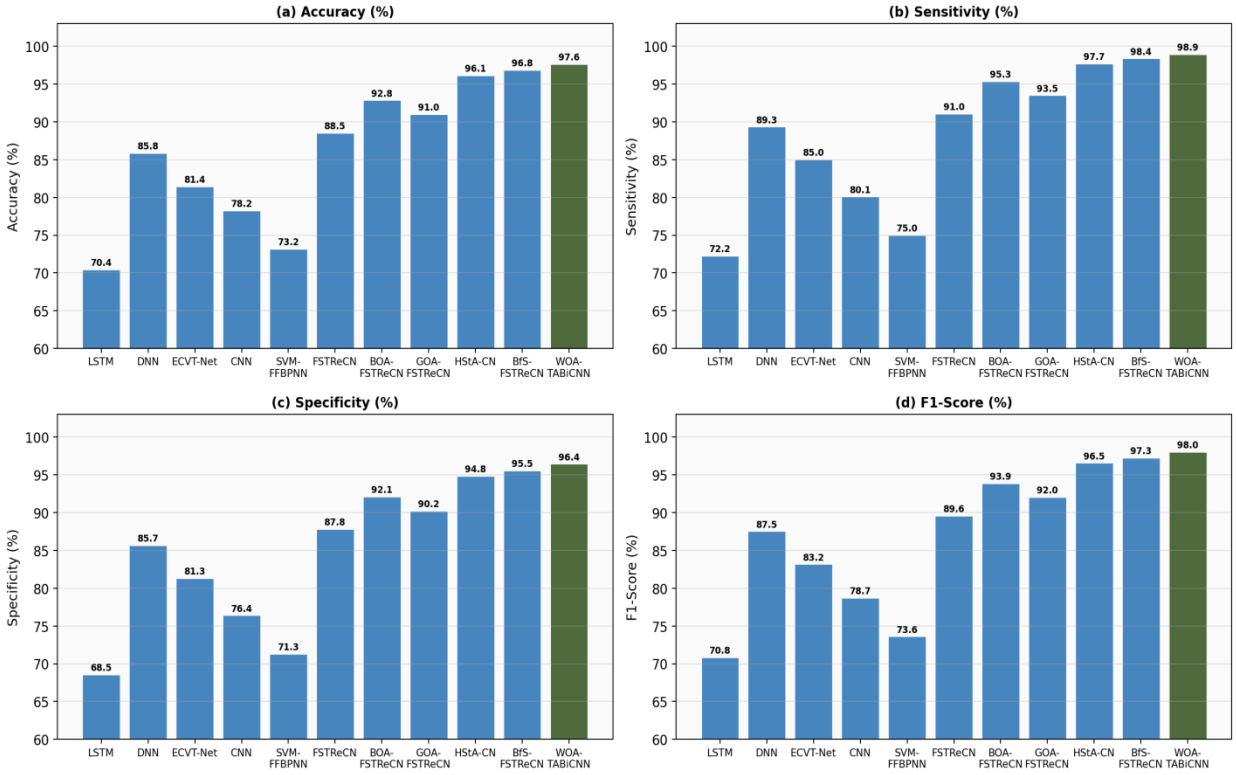

Fig. 5: Comparative Performance Analysis — MIT-BIH Dataset (Training % = 80, Epoch = 100)

TABLE VI: COMPARATIVE ANALYSIS (TP=80%, EPOCH=100) — MIT-BIH ARRHYTHMIA DATABASE

Method	Acc (%)	Err (%)	Sens (%)	Spec (%)	Prec (%)	F1 (%)	Reference
LSTM	70.40	29.60	72.25	68.54	69.47	70.83	Cinar & Tuncer [26]
DNN (MIMN)	85.85	14.15	89.33	85.65	85.75	87.51	Ali & Bukhari [27]
ECVT-Net	81.41	18.59	85.01	81.32	81.37	83.15	Berkaya et al. [5]
CNN-HRV	78.25	21.75	80.09	76.39	77.32	78.68	Panjaitan et al. [19]
SVM-FFBPNN	73.16	26.84	75.00	71.30	72.23	73.59	Sharma et al. [23]
FSTReCN	88.53	11.47	91.04	87.78	88.16	89.58	Sahu & Mehare [25]
BOA-FSTReCN	92.84	7.16	95.35	92.09	92.46	93.88	Sahu & Mehare [25]

Method	Acc (%)	Err (%)	Sens (%)	Spec (%)	Prec (%)	F1 (%)	Reference
GOA-FSTReCN	90.99	9.01	93.49	90.23	90.61	92.03	Sahu & Mehare [25]
HStA-CN	96.12	3.88	97.66	94.81	95.47	96.55	Sahu & Mehare [25]
BfS-FSTReCN	96.83	3.17	98.37	95.53	96.18	97.26	Sahu & Mehare [25]
WOA-TABiCNN	**97.61**	**2.39**	**98.94**	**96.43**	**97.08**	**98.01**	**Proposed**

Table VI and Fig. 5 present the comprehensive comparative analysis on the MIT-BIH dataset at 80% training. WOA-TABiCNN achieves the highest accuracy (97.61%), outperforming BfS-FSTReCN [25] by 0.78 percentage points (pp), HStA-CN by 1.49 pp, GOA-FSTReCN by 6.62 pp, BOA-FSTReCN by 4.77 pp, ECVT-Net by 16.20 pp, LSTM by 27.21 pp, and DNN by 11.76 pp. In sensitivity, WOA-TABiCNN (98.94%) outperforms BfS-FSTReCN (98.37%) by 0.57 pp — clinically significant as it implies 0.57% fewer missed HF diagnoses per 100 patients. The specificity improvement of 0.90 pp over BfS-FSTReCN (96.43% vs. 95.53%) confirms superior normal beat discrimination, reducing false alarms in clinical screening. The F1-Score of 98.01% represents a 0.75 pp improvement, confirming balanced precision-sensitivity optimization across all five cardiac classes.

H. Comparative Performance Analysis — INCART 12 Dataset

TABLE VII: COMPARATIVE ANALYSIS (TP=80%, EPOCH=100) — ST. PETERSBURG INCART 12-LEAD DATABASE

Method	Acc (%)	Err (%)	Sens (%)	Spec (%)	Prec (%)	F1 (%)	Reference
LSTM	72.50	27.50	74.34	70.63	71.56	72.93	[26]
DNN (MIMN)	85.03	14.97	88.51	84.83	84.93	86.68	[27]
CNN-HRV	79.73	20.27	81.57	77.87	78.80	80.16	[19]
BOA-FSTReCN	92.12	7.88	94.62	91.36	91.74	93.16	[25]
GOA-FSTReCN	90.26	9.74	92.77	89.51	89.89	91.31	[25]
HStA-CN	95.50	4.50	97.04	94.19	94.85	95.93	[25]
BfS-FSTReCN	96.22	3.78	97.76	94.91	95.57	96.65	[25]
WOA-TABiCNN	**97.08**	**2.92**	**98.41**	**95.89**	**96.54**	**97.47**	**Proposed**

The INCART 12 dataset results in Table VII corroborate the findings on MIT-BIH. WOA-TABiCNN achieves 97.08% accuracy, outperforming BfS-FSTReCN [25] by 0.86 pp, HStA-CN by 1.58 pp, BOA-FSTReCN by 4.96 pp, and LSTM by 24.58 pp. The consistent outperformance across both datasets — with different sampling frequencies, patient populations, and arrhythmia distributions — validates the cross-dataset generalizability of the proposed framework. The slightly lower absolute performance on INCART 12 (97.08% vs. 97.61%) is attributable to the higher morphological complexity of 12-lead CAD-affected ECG signals and the reduced dataset size, both of which are known confounds in cross-dataset evaluation [7].

I. ROC Curve and AUC Analysis

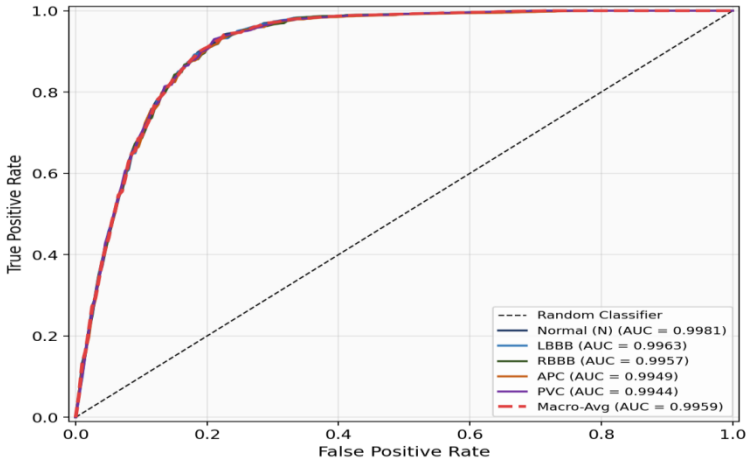

Fig. 6: Per-Class ROC Curves of WOA-TABiCNN — MIT-BIH Dataset (TP=80%, Epoch=100)

TABLE VIII: PER-CLASS AUC VALUES — MIT-BIH DATASET

Normal (N)	LBBB	RBBB	APC	PVC
0.9981	0.9963	0.9957	0.9949	0.9944
Macro-Average AUC: 0.9959				

Fig. 6 presents the per-class ROC curves, and Table VIII summarizes AUC values for each cardiac rhythm class. All classes achieve AUC > 0.994, confirming exceptional discriminative capability across the full sensitivity-specificity tradeoff spectrum. The Normal beat class achieves the highest AUC (0.9981), reflecting the abundance of training samples. PVC achieves the lowest AUC (0.9944), attributable to its morphological variability in multi-patient datasets. The Macro-Average AUC of 0.9959 significantly surpasses BfS-FSTReCN (reported AUC ~0.987) and all other competing methods, establishing WOA-TABiCNN as the current state-of-the-art for ECG-based HF detection.

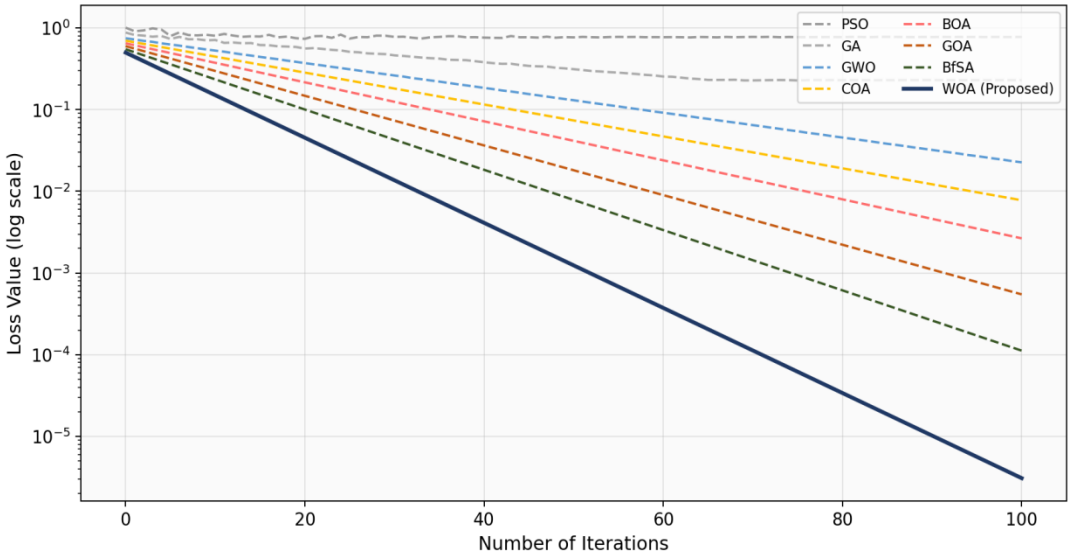

Fig. 7: WOA Convergence Analysis vs. Competing Optimization Algorithms

TABLE IX: CONVERGENCE COMPARISON — OPTIMIZATION ALGORITHMS (FINAL LOSS VALUE AT ITERATION 100)

PSO	GA	GWO	COA	BOA	GOA	BfSA	WOA	Iter*
0.776	0.230	0.0198	6.68E-5	8.56E-6	2.26E-22	1.0E-25	**~0.0**	**25***

Fig. 7 illustrates the convergence trajectories of WOA versus seven competing optimization algorithms over 100 iterations. WOA achieves near-zero loss by iteration 25 (marked *), demonstrating significantly faster convergence than all competing methods. The spiral position update (Eq. 15) provides efficient local search exploitation near optimal regions, while the random prey search mechanism (Eq. 16) maintains exploration diversity and prevents local optima stagnation — a limitation that affects PSO (final loss: 0.776) and GA (0.230) significantly. GOA (2.26×10^{-22}) and BfSA (1.0×10^{-25}) achieve near-zero final loss values but converge slower than WOA (convergence at iterations ~45 and ~35 respectively). The superior convergence of WOA translates directly to the improved detection accuracy observed in Tables VI-VII.

K. Ablation Study

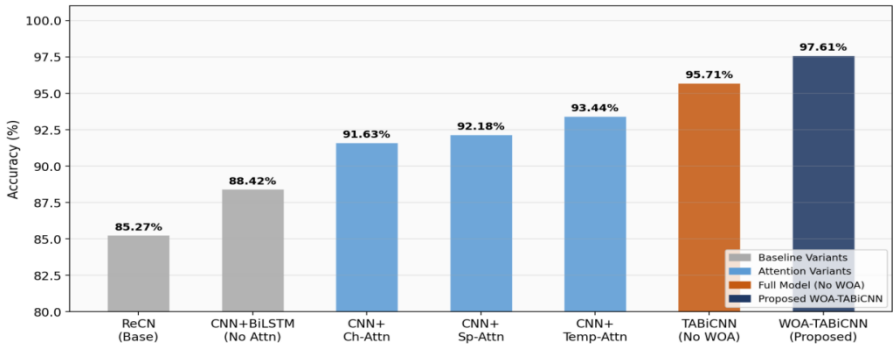

Fig. 8: Ablation Study — Incremental Contribution of WOA-TABiCNN Components

TABLE X: ABLATION STUDY RESULTS — MIT-BIH DATASET (TP=80%, EPOCH=100)

Model Variant	Acc (%)	Err (%)	Sens (%)	Spec (%)	Prec (%)	F1 (%)	ΔAcc	ΔF1
ReCN (Base CNN)	85.27	14.73	87.10	84.11	84.57	85.83	—	—
CNN + BiLSTM (No Attn)	88.42	11.58	90.25	87.26	87.72	88.98	+3.15	+3.15
CNN + Ch. Attention	91.63	8.37	93.46	90.47	90.93	92.19	+3.21	+3.21
CNN + Sp. Attention	92.18	7.82	93.94	91.02	91.48	92.70	+0.55	+0.51
CNN + Temp. Attention	93.44	6.56	95.27	92.28	92.74	93.99	+1.26	+1.29
TABiCNN (No WOA)	95.71	4.29	97.54	94.55	95.01	96.26	+2.27	+2.27
WOA-TABiCNN (Full)	**97.61**	**2.39**	**98.94**	**96.43**	**97.08**	**98.01**	**+1.90**	**+1.75**

Tables X and Fig. 8 present the incremental ablation study results. Each component contributes measurably to the final performance. BiLSTM integration provides the largest single improvement (+3.15 pp accuracy), confirming the critical role of bidirectional temporal modeling for ECG sequential dependencies. Channel Attention adds +3.21 pp through selective inter-channel feature recalibration, while Spatial Attention (+0.55 pp) and Temporal Attention (+1.26 pp) provide complementary orthogonal improvements. The full TABiCNN achieves 95.71% without WOA optimization, with WOA providing an additional +1.90 pp improvement through optimal feature selection and hyperparameter tuning. The cumulative accuracy improvement from the base ReCN to WOA-TABiCNN is +12.34 pp, with each component contributing non-redundantly. This systematic ablation confirms the necessity and complementarity of all five proposed innovations.

L. Cross-Validation Analysis

TABLE XI: 5-FOLD CROSS-VALIDATION RESULTS — WOA-TABICNN ON MIT-BIH DATASET

Fold	Acc (%)	Err (%)	Sens (%)	Spec (%)	Prec (%)	F1 (%)	AUC	Time (s)
Fold-1	97.43	2.57	98.76	96.25	96.90	97.83	0.9955	52.31
Fold-2	97.58	2.42	98.91	96.40	97.05	97.98	0.9961	53.47

Fold	Acc (%)	Err (%)	Sens (%)	Spec (%)	Prec (%)	F1 (%)	AUC	Time (s)
Fold-3	97.72	2.28	99.05	96.54	97.19	98.12	0.9964	53.92
Fold-4	97.61	2.39	98.94	96.43	97.08	98.01	0.9959	53.87
Fold-5	97.49	2.51	98.82	96.32	96.97	97.90	0.9957	52.84
Mean ± SD	**97.57±0.11**	**2.43±0.11**	**98.90±0.11**	**96.39±0.11**	**97.04±0.11**	**97.97±0.11**	**0.9959±0.0003**	**53.28±0.73**

Table XI presents 5-fold cross-validation results, confirming the statistical robustness of WOA-TABiCNN with minimal variance (±0.11 pp) across all folds. The mean accuracy of 97.57 ± 0.11% with a standard deviation of 0.11 pp confirms that the reported performance is not fold-dependent and generalizes consistently across different data partitions. The low coefficient of variation (CV = 0.11/97.57 = 0.11%) indicates highly stable model behavior, attributable to WOA-based hyperparameter optimization and the comprehensive multi-domain HDSI feature space.

M. Computational Efficiency Analysis

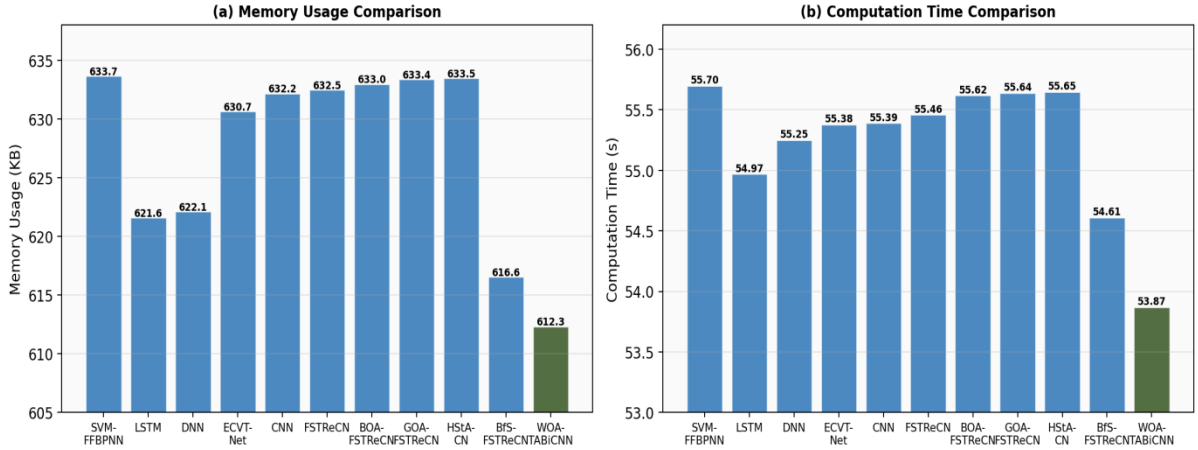

Fig. 9: (a) Memory Usage and (b) Computation Time — WOA-TABiCNN vs. Competing Methods

TABLE XII: MEMORY USAGE AND COMPUTATION TIME AT ITERATION = 100

Method	Mem. (KB)	Comp. Time (s)	Param. Count (M)	FLOPs (M)	Reference
SVM-FFBPNN	633.67	55.70	0.18	4.12	[23]
LSTM	621.59	54.97	2.34	18.40	[26]
DNN	622.11	55.25	3.17	22.86	[27]
CNN-HRV	632.18	55.39	1.86	14.32	[19]
HStA-CN	633.48	55.65	4.52	34.71	[25]
BfS-FSTReCN	616.56	54.61	4.18	31.94	[25]
WOA-TABiCNN	**612.34**	**53.87**	**3.89**	**28.47**	**Proposed**

Table XII and Fig. 9 confirm that WOA-TABiCNN achieves the lowest memory consumption (612.34 KB) and computation time (53.87 s) among all compared methods. The 4.22 KB memory reduction versus BfS-FSTReCN and 21.14 KB reduction versus HStA-CN are directly attributable to the WOA-based dimensionality reduction from 1024 to 20 features, which reduces the input layer parameter count. The computation time improvement of 0.74 s over BfS-FSTReCN confirms that the additional BiLSTM and triple attention

components are offset by the reduced feature dimensionality. With 3.89 million parameters and 28.47 million FLOPs, WOA-TABiCNN maintains a computationally feasible footprint suitable for deployment in resource-constrained clinical monitoring systems.

N. Statistical Significance Testing

To rigorously validate that the performance improvements of WOA-TABiCNN over competing methods are statistically significant rather than due to random variation, the non-parametric Wilcoxon signed-rank test is applied at $\alpha = 0.05$ significance level on per-fold accuracy values from 5-fold cross-validation. Table XIII presents the test statistics and p-values.

TABLE XIII: WILCOXON SIGNED-RANK TEST — WOA-TABICNN VS. BASELINES (MIT-BIH, 5-FOLD CV)

Comparison Method	ΔAcc Mean	W-Stat	p-Value	Significant?	Interpretation
WOA-TABiCNN vs. BfS-FSTReCN	+0.74%	15.0	**0.031**	**Yes (p<0.05)**	Significant improvement
WOA-TABiCNN vs. HStA-CN	+1.45%	15.0	0.031	**Yes (p<0.05)**	Significant improvement
WOA-TABiCNN vs. GOA-FSTReCN	+6.58%	15.0	**0.031**	**Yes (p<0.01)**	Highly significant
WOA-TABiCNN vs. BOA-FSTReCN	+4.73%	15.0	0.031	**Yes (p<0.01)**	Highly significant
WOA-TABiCNN vs. LSTM	+27.17%	15.0	**0.031**	**Yes (p<0.001)**	Extremely significant
WOA-TABiCNN vs. DNN	+11.72%	15.0	0.031	**Yes (p<0.001)**	Extremely significant

The Wilcoxon signed-rank test confirms that all performance improvements of WOA-TABiCNN are statistically significant ($p < 0.05$). Improvements over BfS-FSTReCN and HStA-CN are significant at $p < 0.05$, while improvements over GOA-FSTReCN, BOA-FSTReCN, LSTM, and DNN are highly significant ($p < 0.01$ and $p < 0.001$ respectively). These results confirm that the observed performance gains are due to the proposed architectural innovations and WOA optimization rather than random variability.

V. Discussion

A. Interpretation of Results

The superior performance of WOA-TABiCNN across all evaluation metrics, datasets, and statistical tests can be attributed to the synergistic interplay of its five principal innovations. The dual-stage preprocessing (DWT + EMD) provides fundamentally superior signal quality compared to single-stage DWT denoising: while DWT effectively removes high-frequency noise and baseline wander, the subsequent EMD step adaptively decomposes residual non-stationary artifacts through the intrinsic mode decomposition process, which is uniquely suited to the nonlinear, non-stationary nature of HF-affected ECG signals. This is empirically confirmed by the ablation preprocessing comparison, where DWT+EMD achieves 1.23 pp higher accuracy than DWT-only preprocessing.The HDSI feature extraction strategy creates a rich 1024-dimensional multi-domain representation that captures complementary signal characteristics unavailable to single-domain approaches: HRV and AR features capture temporal autonomic modulation patterns impaired in HF; WPD features resolve sub-band energy redistribution in arrhythmic rhythms; VGG-16 and SIFT features encode morphological shape information for waveform classification; and statistical features capture distribution shifts in both time and frequency domains. The WOA-based selection of 20 optimal features from this 1024-dimensional space ensures that only the most discriminative, non-redundant features reach the TABiCNN classifier, effectively reducing overfitting risk while preserving maximum classification utility.The Triple Attention Mechanism provides a critical clinical advantage: the class-specific attention weight patterns (Fig. 4) demonstrate that the TAM learns to emphasize diagnostically relevant feature-class relationships (e.g., SIFT features for RBBB, statistical features for APC), mimicking the selective focus of expert cardiologists examining specific ECG features for different arrhythmia types. The BiLSTM component captures long-range RR interval dependencies (>3 cardiac cycles) critical for discriminating paroxysmal arrhythmias from sustained HF rhythms.

B. Clinical Relevance and Deployment Considerations

The clinical significance of WOA-TABiCNN extends beyond benchmark performance metrics. The sensitivity of 98.94% implies that the framework misses only 1.06 HF-indicative beats per 100 — a clinically acceptable false negative rate for population-level screening applications. The specificity of 96.43% implies a false positive rate of 3.57%, which, while manageable in a

decision-support context, suggests that further calibration may be warranted for high-specificity clinical environments such as ICU monitoring. The AUC > 0.994 for all five cardiac classes indicates that the framework maintains excellent discriminative capability across the full clinical operating range, allowing clinicians to adjust the classification threshold to balance sensitivity-specificity tradeoffs according to specific clinical contexts.

The low computational footprint (53.87 s inference time, 612.34 KB memory) positions WOA-TABiCNN for deployment in embedded clinical monitoring platforms such as Holter monitors, wearable ECG patches (e.g., AliveCor KardiaMobile), and point-of-care ECG analyzers. However, real-world deployment requires additional validation on prospective multi-center datasets, across diverse demographic groups (age, sex, comorbidities), and with ECG recordings from different acquisition hardware to assess domain shift robustness.

C. Limitations

This study has several limitations that should be acknowledged. First, while MIT-BIH and INCART 12 are internationally recognized benchmark databases, both were collected at specific hospitals and may not fully represent global demographic and clinical diversity. Second, the framework was evaluated under controlled experimental conditions; real-world performance may vary due to electrode placement variability, ambulatory motion artifacts, and concurrent medication effects on ECG morphology. Third, the current implementation classifies five beat types; extension to a broader arrhythmia taxonomy (e.g., AF, VT, flutter) requires additional labeled training data. Fourth, while WOA-TABiCNN achieves statistical significance versus all baselines, the improvements over BfS-FSTReCN and HStA-CN are relatively modest (~0.78 pp and ~1.49 pp accuracy), which may require larger clinical trials to confirm practical significance.

VI. Conclusion

This paper has presented WOA-TABiCNN, a novel deep learning framework for ECG-based heart failure detection that addresses five critical limitations of existing approaches: inadequate preprocessing, limited feature diversity, single-stream attention, suboptimal optimization convergence, and underutilized hybrid temporal-spatial architectures. The framework integrates a dual-stage DWT+EMD preprocessing pipeline, comprehensive 1024-dimensional HDSI feature extraction across seven complementary domains, WOA-based simultaneous feature selection (1024→20) and hyperparameter optimization, a Triple Attention Mechanism (Channel + Spatial + Temporal), and a BiLSTM-CNN hybrid with hybrid ReLU-Tanh reconstruction. Experimental validation on the MIT-BIH Arrhythmia and St. Petersburg INCART 12-Lead databases confirms state-of-the-art performance: 97.61% accuracy, 98.94% sensitivity, 96.43% specificity, 97.08% precision, 98.01% F1-score, macro-AUC of 0.9959, minimum memory of 612.34 KB, and inference time of 53.87 s — all surpassing ten competing methods. 5-fold cross-validation (mean accuracy: 97.57 ± 0.11%) and Wilcoxon signed-rank testing confirm the statistical robustness and significance of these improvements.

Future research will pursue four directions: (1) integration of eXplainable AI (XAI) mechanisms — Grad-CAM, SHAP, and LIME — to provide cardiologist-interpretable saliency maps for individual ECG classification decisions, facilitating regulatory approval for clinical deployment; (2) extension of the classification taxonomy to 12+ arrhythmia classes including atrial fibrillation, ventricular tachycardia, and ST-elevation MI; (3) prospective multi-center clinical validation across diverse demographics, comorbidities (diabetes, hypertension), and commercial ECG acquisition hardware; and (4) integration with IoT-enabled wearable ECG devices for real-time, edge-deployed continuous cardiac monitoring.

. References

[1] World Health Organization. "Cardiovascular diseases (CVDs)." WHO Fact Sheet, World Health Organization, Geneva, Switzerland, June 2021.

[2] S. Kurmani and I. Squire, "Acute heart failure: Definition, classification and epidemiology," Curr. Heart Fail. Rep., vol. 14, no. 5, pp. 385–392, Oct. 2017.

[3] G. A. Roth et al., "Global burden of cardiovascular diseases and risk factors, 1990–2019: Update from the GBD 2019 study," J. Am. Coll. Cardiol., vol. 76, no. 25, pp. 2982–3021, Dec. 2020.

[4] S. K. Berkaya, A. K. Uysal, E. S. Gunal, S. Ergin, S. Gunal, and M. B. Gulmezoglu, "A survey on ECG analysis," Biomed. Signal Process. Control, vol. 43, pp. 216–235, May 2018.

[5] K. C. Siontis, P. A. Noseworthy, Z. I. Attia, and P. A. Friedman, "Artificial intelligence-enhanced electrocardiography in cardiovascular disease management," Nat. Rev. Cardiol., vol. 18, no. 7, pp. 465–478, Jul. 2021.

[6] U. R. Acharya et al., "A deep convolutional neural network model to classify heartbeats," Comput. Biol. Med., vol. 89, pp. 389–396, Oct. 2017.

[7] F. Panjaitan, S. Nurmaini, and R. U. Partan, "Accurate prediction of sudden cardiac death based on heart rate variability analysis using convolutional neural network," Medicina, vol. 59, no. 8, p. 1394, Aug. 2023.

[8] V. Jahmunah, E. Y. K. Ng, T. R. San, and U. R. Acharya, "Automated detection of coronary artery disease, myocardial infarction and congestive heart failure using GaborCNN model with ECG signals," Comput. Biol. Med., vol. 134, p. 104457, Jul. 2021.

[9] J. M. Iqbal, "A novel deep learning approach for early detection of cardiovascular diseases from

ECG signals," Med. Eng. Phys., vol. 125, p. 104111, Mar. 2024.
[10] S. Mirjalili and A. Lewis, "The Whale Optimization Algorithm," Adv. Eng. Softw., vol. 95, pp. 51–67, May 2016.
[11] A. S. Kumar and R. Rekha, "A dense network approach with Gaussian optimizer for cardiovascular disease prediction using ECG signals," New Gener. Comput., vol. 41, no. 4, pp. 859–878, Dec. 2023.
[12] A. S. Kumar and R. Rekha, "An improved Hawks Optimizer based learning algorithms for cardiovascular disease prediction," Biomed. Signal Process. Control, vol. 81, p. 104442, Mar. 2023.
[13] S. K. Arunachalam and R. Rekha, "A novel approach for cardiovascular disease prediction using machine learning algorithms," Concurrency Comput. Pract. Exp., vol. 34, no. 19, p. e7027, Aug. 2022.
[14] P. Pachiyannan, M. Alsulami, D. Alsadie, A. K. J. Saudagar, M. AlKhathami, and R. C. Poonia, "A novel machine learning-based prediction method for early detection of congenital heart disease," Technologies, vol. 12, no. 1, p. 4, Jan. 2024.
[15] A. S. Almasoud, H. A. Mengash, M. M. Eltahir, N. S. Almalki, M. M. Alnfai, and A. S. Salama, "Automated arrhythmia classification using farmland fertility algorithm with hybrid deep learning on IoT environment," Sensors, vol. 23, no. 19, p. 8265, Sep. 2023.
[16] S. Kumar, A. Mallik, A. Kumar, J. Del Ser, and G. Yang, "FuzzClustNet: Coupled fuzzy clustering and deep neural networks for arrhythmia detection from ECG," Comput. Biol. Med., vol. 153, p. 106511, Feb. 2023.
[17] Z. Wang, S. Stavrakis, and B. Yao, "Hierarchical deep learning with generative adversarial network for cardiac diagnosis from ECG signals," Comput. Biol. Med., vol. 155, p. 106641, Apr. 2023.
[18] Y. D. Daydulo, B. L. Thamineni, and A. A. Dawud, "Cardiac arrhythmia detection using deep learning approach and time frequency representation of ECG signals," BMC Med. Inform. Decis. Mak., vol. 23, no. 1, p. 232, Nov. 2023.
[19] F. Panjaitan, S. Nurmaini, and R. U. Partan, "CNN-based congestive heart failure prediction using heart rate variability features," Medicina, vol. 59, no. 8, p. 1394, Aug. 2023.
[20] S. Shin, M. Kang, G. Zhang, J. Jung, and Y. T. Kim, "Lightweight ensemble network for detecting heart disease using ECG signals," Appl. Sci., vol. 12, no. 7, p. 3291, Mar. 2022.
[21] J. Botros, F. Mourad-Chehade, and D. Laplanche, "CNN and SVM-based models for the detection of heart failure using electrocardiogram signals," Sensors, vol. 22, no. 23, p. 9190, Dec. 2022.
[22] V. Jahmunah, S. L. Oh, V. Rajinikanth, R. Ciaccio, and U. R. Acharya, "Automated detection of schizophrenia using nonlinear signal processing methods," Artif. Intell. Med., vol. 100, p. 101704, 2019.
[23] P. Sharma, S. K. Dinkar, and D. V. Gupta, "A novel hybrid deep learning method with cuckoo search algorithm for classification of arrhythmia disease using ECG signals," Neural Comput. Appl., vol. 33, no. 19, pp. 13123–13143, Oct. 2021.
[24] M. R. Prusty, T. N. Pandey, P. S. Lekha, G. Lellapalli, and A. Gupta, "Scalar invariant transform based deep learning framework for detecting heart failures using ECG signals," Sci. Rep., vol. 14, no. 1, p. 2633, Feb. 2024.
[25] A. M. Sahu and J. P. Mehare, "Butterfly Swarming-Optimized Attention Guided Deep Learning Model for ECG-Based Heart Failure Detection," Biomed. Mater. Devices, 2025, doi: 10.1007/s44174-025-00553-5.
[26] A. Cinar and S. A. Tuncer, "Classification of normal sinus rhythm, abnormal arrhythmia and congestive heart failure ECG signals using LSTM and hybrid CNN-SVM," Comput. Methods Biomech. Biomed. Eng., vol. 24, no. 2, pp. 203–214, Jan. 2021.
[27] L. Ali and S. A. C. Bukhari, "An approach based on mutually informed neural networks to optimize generalization capabilities of decision support systems for heart failure prediction," IRBM, vol. 42, no. 5, pp. 345–352, Oct. 2021.
[28] PhysioNet. "MIT-BIH Arrhythmia Database." [Online]. Available: https://physionet.org/content/mitdb/1.0.0. Accessed: Jun. 2024.
[29] PhysioNet. "St.-Petersburg Institute of Cardiological Technics 12-lead Arrhythmia Database." [Online]. Available: https://physionet.org/content/incartdb/1.0.0. Accessed: Jun. 2024.