Mathematics

AIMS Mathematics, 11(7): 19876–19920.
DOI: 10.3934/math.2026807
Received: 20 April 2026
Revised: 29 May 2026
Accepted: 10 June 2026
Published: 07 July 2026

https://www.aimspress.com/journal/Math

Research article

End-to-end joint Margin aware BiLSTM-dSVM model for atrial fibrillation detection based on knowledge-augmented deep learning on knowledge-augmented deep learning

Raid Abdulhadi Abdulqader Alabdullah[1,2], Saratha Sathasivam[2,*] and Marwa Mawfaq MohamedSheet AL-Hatab[3]

[1] Technical Engineering College, Northern Technical University, Mosul, Iraq

[2] School of Mathematical Sciences, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia

[3] Technical Engineering College/Northern Technical University, Mosul, Iraq

* **Correspondence**: Email: saratha@usm.my.

Abstract: To avoid stroke, atrial fibrillation (AF) needs to be detected as early as possible. Nevertheless, traditional deep learning systems often are unsuccessful with the domain shift and signal degradation inherent in wearable sensor measurements. To achieve the purpose of going beyond probabilistic classification, We present a hybrid convolutional neural network–bidirectional long ahort-term memory–differentiable support vector machine (CNN–BiLSTM–dSVM) framework that integrates automated morphological feature generation with geometric margin maximization. The main methodological innovation is the (dSVM) head, which is trained on a aquared hinge loss objective, instead of conventional Soft max layers. To isolate structural P-QRS-T variations and remove motion artifacts and baseline drift, this architecture uses a CNN backbone. These morphological signs are used with knowledge-augmented clinical data to represent long-term temporal rhythms and are processed by a Bi LSTM encoder. Unlike present models, our method removes overfitting to sensor-induced noise, by imposing structural risk minimization, and creating an operative loss-friendly dead zone. Experimental analysis of the proposed system was conducted on dual-source ECG datasets.The proposed framework was evaluated using dual-source ECG datasets, including the MIT-BIH atrial fibrillation database, the PhysioNet CinC 2017 dataset, and a noisy wearable ECG dataset collected from real-world sensing environments. The architecture combined CNN-based morphological feature extraction, BiLSTM temporal representation learning, and a (dSVM) optimization strategy to improve atrial fibrillation classification robustness under heterogeneous ECG acquisition conditions. Experimental results demonstrated that the proposed framework achieved an overall classification accuracy of 99.0%, recall of 98.8%, Macro F1-score of 98.6%, and AUC of 0.995. Comparative evaluation against mainstream baseline architectures, including conventional CNN, LSTM, and Transformer-based models, confirmed superior convergence stability, generalization capability, and classification robustness, particularly under noisy wearable ECG conditions. (20,000 noisy wearable recordings and 6,000 high-fidelity clinical signals). The datasets were partitioned into training, validation, and testing subsets using a strict patient-level splitting protocol to avoid data leakage. This gave system a near-perfect overall accuracy of 99.5% and a recall of 1.00. The proposed hybrid model restored this performance in 400 Hz noiseful mobile signals, but standalone CNNs exhibited constraints in terms of dealing with class imbalances (Macro F1: 0.593). The sensor-invariant representation of our Margin-aware inductive bias provided better convergence kinetics and structural stability than high-complexity 1D-transformer and ECG-BERT baselines as found in statistical studies. In this work, we established a benchmark of computationally efficient, theoretically-grounded cardiac surveillance in real-world mHealth.

Keywords: atrial fibrillation; joint optimization; BiLSTM-dSV; knowledge-augmented deep learning; differentiable SVM; transformer benchmarking; wearable healthcare
Mathematics Subject Classification: 68T07

1. Introduction

Atrial fibrillation (AF), one of the most common cardiac arrhythmias worldwide, is strongly associated with an increased risk of heart failure, ischemic stroke, and cardiovascular death [1]. Therefore, to improve patient outcomes and reduce long-term clinical effects, early and accurate detection of AF is essential. Electrocardiography (ECG) is the gold standard for detecting AF since it is non-invasive, affordable, and suitable for continuous monitoring in clinical and ambulatory settings [2,3]. Nevertheless, such feature-based methods do not typically represent the intricate temporal relationships and nonlinear behavior particularly when applied to short-duration or noisy ECG traces, which are typical of wearable and real-world monitoring scenarios [4]. Based on these findings, we suggest a hybrid fusion-based ECG classification system that combines manually obtained HRV features with deep morphological and temporal features trained using CNN and LSTM networks, respectively. The cross-dataset generalization is evaluated externally using the MIT-BIH AF database [5] and the database is used to train and validate . Traditional AF detection techniques have relied heavily on handcrafted features extracted from ECG data, particularly heart rate variability (HRV) and RR-interval-based statistical descriptors [6,7]. When combined with these features, traditional machine learning classifiers like support vector machines (SVMs), Bayesian models, and decision trees have demonstrated respectable performance [8,9]. In a bid to address these constraints, the deep learning (DL) methods are being used more frequently in the detection of automatic AF. The local signal structures can be effectively modeled using convolutional neural networks (CNNs), which have demonstrated an amazing capability to learn discriminative morphological features directly out of ECG waveforms [10,11]. Long-term recurrent networks such as as long short-term memory (LSTM) networks have been shown to be successful in identifying long-term temporal correlations and rhythmic anomalies in ECG data [12,13]. With these advances, standalone deep learning models can exhibit a lack of generalization and resilience, especially in the event of cross-dataset and inter-subject variance [14]. Researchers have observed that, due to their ability to work with dynamic and large-scale ECG data, DL-based methods can be more effective in comparison with traditional machine learning models [15]. As a result, hybrid learning frameworks that combine deep representations with conventional classifiers are becoming more and more popular. Specifically, it was suggested that CNN-SVM and LSTM-SVM designs can exploit the discriminative margin-classification, as well as automatic feature-learning [16]. The goal of these hybrid models is to provide a trade-off between interpretability, durability, and processing efficiency. Moreover, fusion-based approaches have become a promising approach to AF detection through combining complementary data across a large number of feature domains such as statistical HRV descriptors, morphological characteristics, and temporal embeddings [17]. Fusion frameworks can improve class separability, resistance to noise, and generalization between datasets in comparison to single-model methods by leveraging the benefits of CNN, LSTM, and SVM models [14,17]. Based on experimental results, the proposed fusion model works better than single SVM, CNN, and LSTM models, which highlights its relevance to wearable ECG sensors and clinical applications that require real-time monitoring. Our main goal is to close the methodological gap between contemporary structural deep learning and conventional heuristic fusion. To do so, we introduce a novel end-to-end joint Margin-aware bidirectional LSTM (BiLSTM)-differentiable SVM (dSVM) framework that does not use a single optimization framework but a series of interconnected pipelines. We integrate knowledge-augmented deep learning (KADL) by fundamentally operating on sixteen domain-specific clinical indicators directly in the latent manifold, thus confining the neural representations to cardiac domain priors compared to existing stacking-based ensembles. Mean-field theory in Hopfield neural network for doing 2 satisfiability logic programming suitable for: optimization, neural networks, convergence, mathematical modelling [18]. Ybrid intelligent systems integrate multiple computational intelligence techniques to improve learning capability, optimization, and decision-making. Agent-based modeling combined with fuzzy Hopfield neural networks has demonstrated the effectiveness of hybrid intelligent approaches in solving complex optimization and logical inference problems [19]. In healthcare prediction, hybrid intelligent systems have been shown to achieve better diagnostic accuracy and robustness [20] and inspired the usage of deep learning and discriminative classifiers in medical decision support systems. Hybridized neural optimization methods have been found to enhance the convergence properties, predictive stability, and model generalization in complex nonlinear learning situations [21].

The main contribution of this work is the integration of BiLSTM feature learning with a differentiable SVM optimization framework for robust AF detection. This synergy guarantees strong classification in high-noise conditions as the network is compelled to learn geometrically different and statistically precise temporal embeddings. The research provides empirical information on the best domain generalization and structural reliability of real-time clinical diagnostics by comparing this model to a larger sample of 20,000 ECG signals and conducting a comparison with the state-of-the-art Transformer-based models (e.g., ECG-BERT). Unlike previously reported CNN-BiLSTM-SVM hybrid architectures, this study emphasizes three complementary innovation dimensions. First, the framework introduces an end-to-end jointly optimized margin-aware learning strategy in which temporal feature extraction and differentiable SVM-based structural risk minimization are optimized simultaneously during training. Second, the proposed architecture incorporates knowledge-enhanced latent fusion by integrating clinically meaningful ECG descriptors with deep temporal representations to improve interpretability and cross-domain robustness. Third, the framework is designed to improve noise robustness under heterogeneous wearable ECG sensing conditions through adaptive fusion and margin-constrained optimization mechanisms. These combined characteristics distinguish the proposed framework from conventional deep SVM integration approaches and previously reported hybrid ECG classification models [22]. Despite the substantial progress achieved by deep learning approaches for ECG-based atrial AF detection, several important research gaps remain insufficiently addressed. Moreover CNN and Transformer-based architectures primarily focus on improving feature extraction accuracy; however, many studies lack an end-to-end jointly optimized framework capable of simultaneously modeling temporal dependencies, structural risk minimization, and robust decision boundaries under noisy wearable sensing environments. Furthermore, conventional deep learning approaches frequently neglect the integration of clinically meaningful statistical descriptors and domain-specific ECG knowledge, thereby limiting interpretability, robustness, and cross-domain generalization capability in heterogeneous real-world monitoring conditions. Studies have demonstrated that multimodal information fusion, knowledge-enhanced learning, and hybrid optimization strategies can significantly improve predictive robustness in complex nonlinear time-series applications [23]. Advanced fusion-based BiLSTM frameworks and interpretable deep learning architectures have shown promising performance in healthcare prediction and signal-processing domains by combining data-driven learning with domain-aware feature integration [24,25]. Motivated by these developments, we propose a jointly optimized BiLSTM-dSVM framework that integrates temporal sequence learning, margin-aware optimization, and knowledge-augmented ECG representations for robust AF classification across clinical and wearable ECG datasets. To detect AF with the help of ECG signals, we propose a combined fusion-based paradigm. The design blends manually create HRV-based statistical descriptors with deep morphological and temporal representations learned by CNN and LSTM models, respectively.By combining complimentary data from the statistical, morphological, and temporal domains via a fusion mechanism, the proposed method outperforms standalone deep learning, single-hybrid models, and conventional machine learning in terms of generalization, robustness, and class reparability. External validation on the MIT-BIH AFDB shows that the framework is suitable for wearable and real-time ECG monitoring applications. Table 1 summarizes the existing atrial fibrillation detection approaches.

Table 1. Summary of AF detection approaches in the literature (unified comparison).

Main limitation	Key strength	Dataset	Model/Architecture	Features/representation	Study (Ref.)	Category
Sensitive to noise	Simple & interpretable	Physio Net	SVM	RR intervals	Narayana et al. [4,8]	Traditional ML
Manual features	Low complexity	MIT–BIH	Bayesian	Statistical ECG	Marti et al. [9]	
Poor generalization	Baseline performance	Local ECG	SVM / ANN	HRV	Sieciński et al. [10]	
High complexity	Automatic feature learning	CinC 2017	CNN–RNN	Morphological	Limam & Precioso [11]	Deep learning (CNN)
Limited interpretability	Strong spatial modeling	PhysioNet	CNN	Raw ECG morphology	Grabowski et al. [12]	
Data-hungry	Long-term dependency	Generic signals	LSTM	Temporal sequences	Sherstinsky [13]	Deep learning (LSTM)
Noise sensitivity	Rhythm modeling	ECG datasets	RNN/LSTM	ECG temporal patterns	Satheeswaran et al. [15]	
Single feature stream	Better separability	ECG datasets	CNN–SVM	CNN features	Ma et al. [18]	Hybrid models
Limited fusion	Robust performance	ECG/PPG	DL + ML	Multi-feature	Aldughayfiq et al. [19]	
Higher complexity	Improved generalization	Dynamic ECG	Decision fusion	Multi-view ECG	Ma et al. [17]	Fusion-based
Noise sensitivity & domain shift	High robustness & generalization	CinC 2017 + MIT-BIH + Wearable ECG	CNN–BiLSTM–dSVM	Morphology + HRV + Tempora	Proposed Work	AF Detection

Transformer-based architectures, namely 1D time-series Transformers and ECG-BERT models, have been shown in research to be effective in capturing long-range dependencies for the categorization of cardiac arrhythmias. Nevertheless, these models frequently need extensive pre-training and significant computational resources, which could make it difficult to implement them in real-time, resource-constrained clinical settings. Moreover, the structural inductive bias necessary to achieve geometric class separation in noisy ECG signals may be absent from pure attention-based methods. In this study,we present a jointly optimized BiLSTM-dSVM framework in response to these constraints. This approach is a combination of a structural risk reduction of a differentiable SVM head and the ability of BiLSTMs to model time. Our technique is more computationally efficient and diagnostic robust, combining domain-specific clinical characteristics in a Knowledge-Augmented manifold, a theoretically-based alternative to Transformer models with high parameter counts.Additionally, optimization strategies for hyperparameter tuning and robust deep learning generalization have been investigated. Advanced optimization frameworks and adaptive parameter search methods have demonstrated improved convergence stability and enhanced classification performance in intelligent healthcare and signal-processing applications [26,27]. The BiLSTM model was based on the architecture of long short-term memory (LSTM) with forget gate mechanism that could effectively learn long-term temporal dependency [28]. Recently, deep learning had been applied to several ECG-based atrial fibrillation detection systems and has exhibited a high level of accuracy [29].

2. Materials and methods

Three complementing information streams were integrated into a single, end-to-end framework by the suggested design. Time representations were learned with (BiLSTM), morphological motifs were learned with deep representation learning, and manually learned statistical features used research on heart rate variability (HRV). We adopted the Knowledge-Augmented Latent Fusion strategy in our approach, instead of traditional stacking approaches that have late-stage fusion. The method ensured that the self-learned representations, which are. by definition, limited by physiological priors, are projected into the time space of the BiLSTM by relating domain-specific HRV attributes to the time space of the BiLSTM. We used a (dSVM) head in place of conventional Softmax-based decision-making to optimize this multi-view representation. The structural risk minimization in this decision head was determined by maximizing geometrical distance between cardiac rhythm categories. The methodology was robust in terms of generalization and, it took advantage of the temporal dynamics, morphological structures, and clinical interpretability by jointly optimizing such a large number of learning paradigms into a single optimization loop. Our results indicated that we can achieve the same classification performance as high-parameter Transformer-based models without the vast computing complexity of these models by simply optimizing the latent interaction between multi-view features. This architectural choice provided a computationally inexpensive alternative to computationally expensive transformer backbones, fulfilling the immediate need to have clinically deployable systems with latency, memory footprint, and training costs being critical factors. Figure 1 illustrates the important steps in this investigation, where the HRV features that are handcrafted are projected into the temporal latent space and concatenated as:

$$X_{\text{enhanced}} = [X_{\text{ariginal}} \| X_{\text{statistical}} \| X_{\text{signal}}].$$

This forms the co-latent space that the dSVM head is optimizing, ensuring the learned representation to be structurally constrained with domain knowledge. Figure 1 illustrates the overall architecture of the proposed CNN–BiLSTM–dSVM framework.

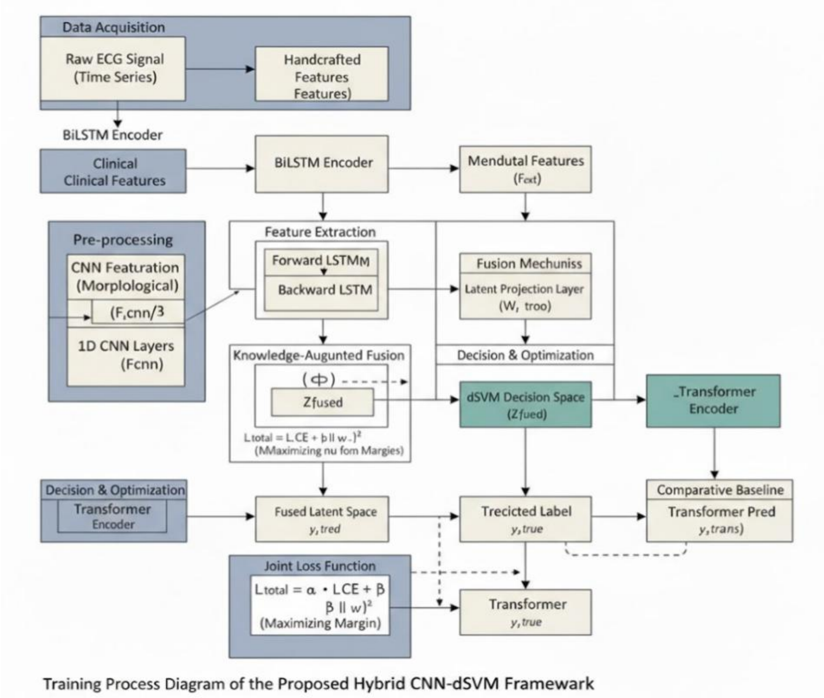


Figure 1. Comprehensive aarchitecture of the proposed joint Margin-aware BiLSTM-dSVM framework.

2.1. ECG data acquisition

To ensure the robustness and clinical validity of the proposed framework, a multi-source ECG integration strategy was adopted. The clinical dataset consisted of 6,000 ECG segments collected from the MIT-BIH atrial fibrillation database and the PhysioNet CinC 2017 dataset. In addition, a secondary wearable ECG dataset containing 20,000 noisy ECG recordings from the AF Signal Dataset [30] was utilized to evaluate cross-domain robustness under real-world noise conditions. All datasets were independently partitioned into training, validation, and testing subsets using a strict patient-level separation protocol to prevent information leakage between training and evaluation stages. The clinical dataset was primarily employed for supervised learning and model validation, whereas the noisy wearable ECG dataset was used to assess model generalization capability under motion artifacts and baseline drift conditions. To ensure objective performance evaluation and prevent data leakage, a strict subject-independent dataset partitioning strategy was adopted throughout the experiments. The ECG datasets were divided into training, validation, and testing subsets using a 70:15:15 ratio at the patient level, ensuring that ECG recordings originating from the same subject did not appear across subsets simultaneously. This separation protocol was applied consistently to the clinical ECG datasets and the wearable noisy ECG datasets to guarantee unbiased cross-domain evaluation and reliable generalization assessment.

2.2. Pre-processing stage

Preprocessing plays a crucial role in enhancing signal-to-noise ratios and ensuring consistency in the input format to the subsequent layers of joint optimization. The four key procedures in this phase aimed at normalizing the characteristics of the multi-source datasets.

- Synchronization and resampling of signals: A digital resampling procedure was first used to synchronize all signals to a common frequency of 300 Hz because the large-scale Kaggle dataset was sampled at 400 Hz, while the clinical PhysioNet data was captured at 300 Hz. This uniform temporal resolution was essential to causing the BiLSTM layers to learn consistent shapes of heartbeats and not be biased by frequency.
- Filtering and removal of artifacts: A multi-stage filtering was employed to eliminate physiological and environmental disturbance. This was made up of a notch filter at 50/60 Hz to remove power-line interference and a zero-phase Butterworth band-pass filter (0.5 Hz to 45 Hz) to reduce baseline wander and muscular artifacts. These filters eliminated high-frequency noise that may have undermined the margin-conscious dSVM decision boundary without loss of the clinically significant P-QRS-T complexes.
- HRV extraction and R-peak detection: To identify the R-peaks so as to measure the dynamic features of the heart rate, the Pan-Tompkins method, based on derivative and integration operations to identify the ventricular depolarization sites, was employed. These peaks were then used to compute the RR-interval sequences, which gave the necessary statistical descriptors of the Knowledge-Augmented clinical stream. This ensured that the (HRV) data and raw morphology were incorporated in the framework.
- Weaknesses segmentation and normalization of signals: To ensure all dimensions of the neural manifold were the same, continuous ECG recordings are divided into fixed-length parts (e.g., 5 seconds or 10 seconds). The segments are then normalized to Z-score, which brings the amplitudes to a zero value and unit variance. Throughout the training process, this step accelerated the convergence of the gradient descent of joints and biases due to the subject.

2.3. Feature extraction

The classification of ECG signals was based on the statistical (HRV) characteristics derived as the beat-to-beat variation and rhythm irregularities of a clinically significant size in determining AF diagnosis, based on RR interval series. RR interval, calculated as per the equations [31,32], is the interval between two R-waves and is inversely proportional to heart rate. Since the LSTM networks have gating mechanisms, which capture the short- and long-term temporal dependencies, they offer a good model to such RR-interval dynamics [33,34].Statisitical and temporal parameters that reflect heart rhythm, variability, and morphological aspects were extracted from the ECG data. Each feature was defined analytically, and determined using pre-processed ECG segments, as explained below.

The selected HRV and morphological ECG descriptors were chosen based on their established clinical relevance to AF pathophysiology. HRV-related features capture irregular ventricular rhythm variability and loss of normal autonomic regulation, which are characteristic manifestations of AF episodes. In addition, morphological ECG features such as RR interval dynamics, waveform irregularities, and temporal atrial activity variations provided clinically meaningful indicators associated with abnormal cardiac conduction behavior. Integrating these handcrafted clinical descriptors with deep temporal representations improves interpretability and enhances discriminative feature learning under heterogeneous and noisy ECG acquisition environments.

2.3.1. Temporal feature extraction

RR interval: The $\boldsymbol{RR}$ interval represents the time difference between two consecutive R-peaks in the ECG waveform:

$$RR_i = t_{Ri} - \mathrm{t}_{R_{t-1}},\ i = 2,3,\dots,N, \tag{1}$$

where $t_{(Ri)}$ is the time position of the i-th R-peak. The RR interval represents the time interval between two consecutive beats, and it is the primary indicator for analyzing heart rate variability (HRV) and distinguishing between AF and normal rhythm [34].

P-wave duration: The duration of the P-wave indicates the period of atrial depolarization [35]:

$$P_{dur} = t_{P_{onset}} - t_{P_{end}}, \tag{2}$$

where $t_{P_{onset}}$ and $t_{P_{end}}$ denote the start and end of the P-wave, respectively. It reflects atrial activity and conduction consistency.

QRS duration: This feature measures ventricular depolarization:

$$QRS_{dur} = t_S - t_Q, \tag{3}$$

where t_Q and t_S indicate the onset and end of the QRS complex. Prolonged QRS duration may reflect conduction block or abnormal ventricular activation [35].

PR interval: The PR interval represents the conduction time from atrial to ventricular activation [35]:

$$PR_{int} = t_Q - t_{P_{onset}}. \tag{4}$$

This interval is a useful indicator of atrioventricular (AV) conduction delay.

Heart rate (HR): It represents the average heart rate and is used as a basic standard to describe the regularity of the heart rhythm [36].

$$RR_{mean} = \frac{1}{N-1}\sum_{i=2}^{N} RR_i,$$

$$HR = \frac{60}{RR_{mean}}, \tag{5}$$

where RR_{mean} is the average RR interval in seconds. It is a fundamental rhythm parameter.

Statistical and HRV-derived features: Root mean square of successive differences, or (RMSSD) (HRV) is evaluated using this time-domain statistical measure, which mainly captures the autonomic nervous system's transient activity [35]:

$$\Delta\mathrm{RR}_i = RR_{i+1} - RR_i \qquad i = 2,3,\dots,N-1. \tag{6}$$

This produces N−2 differences. Then square each difference:

$$(\Delta\mathrm{RR}_i)^2 = (RR_{i+1} - RR_i)^2. \tag{7}$$

Compute the mean of squared differences:

$$\frac{1}{N-1} = \sum_{i=1}^{N-1}(RR_{i+1} - RR_i)^2,$$

$$\text{RMSSD} = \sqrt{\frac{1}{\text{N}-1} \sum_{\text{i}=1}^{\text{N}-1}(\text{RR}_{\text{i}+1} - \text{RR}_{\text{i}})^2}, \tag{8}$$

where it is the difference between successive RR periods. $RR_{i+1} - RR_i$ is the sum of the squares of the differences $\sum_{i-1}^{N-1}$. N is the number of RR periods (or pulses) in the recording. $N-1$ is the number of successive differences.

2.3.2. Statistical feature extraction

SDNN (Standard deviation of SDNN intervals): SDNN is defined as the standard deviation of time between normal heartbeats (NN intervals) over a given time period, and is used as a measure of (HRV) [35].

$$\overline{\text{RR}} = \frac{1}{\text{N}} \sum_{\text{i}=1}^{\text{N}} \text{RR}_{\text{i}}, \tag{9}$$

$$\text{SDNN} = \sqrt{\frac{1}{\text{N}-1} \sum_{\text{i}-1}^{\text{N}}(\text{RR}_{\text{i}} - \overline{\text{RR}})^2}, \tag{10}$$

where N is the number of time intervals, RR_i is the value of an individual time interval,and $\overline{RR}$ is the average of all time intervals.
Mean amplitude: This is the electrocardiogram (ECG) signal's average amplitude. The displayed mathematical formula is:

$$\mu = \frac{1}{\text{N}} \sum_{\text{i}=1}^{\text{N}} \text{x}_{\text{i}}, \tag{11}$$

where x_i is the sample value of the signal, standard deviation: denoted by μ, measures the variance of the signal [36].
Standard deviation (σ): Quantifies signal variability:

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i-1}^{N}(x_i - \mu)^2}. \tag{12}$$

For each data point, μ is the arithmetic mean, and N is the number of data points [37].
Variance (Var): Measures the overall dispersion of signal amplitude:

$$\text{Var} = \frac{1}{\text{N}-1} \sum_{\text{i}=1}^{\text{N}}(\text{x}_{\text{i}} - \mu)^2. \tag{13}$$

A high variance indicates unstable cardiac dynamics [11].
Skewness: Quantifies asymmetry in the ECG amplitude distribution:

$$\text{Skew} = \frac{1}{\text{N}} \sum_{\text{i}=1}^{\text{N}}\left(\frac{\text{x}_{\text{i}} - \mu}{\sigma}\right)^3. \tag{14}$$

Positive skewness may indicate higher amplitude peaks or artifacts [11].

Kurtosis: Measures the pawedness or flatness of the signal:

$$\mathrm{Kurt} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i-\mu}{\sigma}\right)^4. \tag{15}$$

Higher kurtosis indicates sharp R-peaks and strong periodicity [11].

Entropy: This quantifies a signal's complexity or irregularity. Each amplitude level's probability is determined using it. Arrhythmic activity is frequently linked to increased entropy [19].

$$H = -\sum_i p_i \log(p_i), \tag{16}$$

where $p_i \geq 0, \sum_i p_i = 1$ is the probability of each amplitude level. Increased entropy often correlates with arrhythmic activity.

Energy: Quantifies the total power of the ECG segment [19]:

$$E = \sum_{i=1}^{N} x_i^2. \tag{17}$$

This feature is essential for signal strength analysis and normalization.

Zero crossing rate (ZCR): This frequency-domain metric shows how quickly signal polarity shifts: Define the sign function:

$$sign(x) = \begin{cases} +1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases}$$

$$\mathrm{ZCR} = \frac{1}{2N}\sum_{i=1}^{N}|\mathrm{sign}(x_{i+1}) - \mathrm{sign}(x_i)|, \tag{18}$$

where $sign(x_{i+1})$ and $sign(x_i)$ is the sign function. Higher ZCR values are associated with noisy or high-frequency components.

2.4. Feature concatenation

An enhanced feature matrix X_enhanced was created by combining the extracted statistical and signal features with the original characteristics X. Following the extraction of the derived statistical or signal-based features (e.g., mean, SDNN, and RMSSD) and the original features X_original from the ECG signal (e.g., RR intervals, P-wave duration, QRS width), all feature sets were concatenated into an augmented feature matrix as follows:

$$X_{\mathrm{enhanced}} = [X_{\mathrm{original}} \| X_{\mathrm{statistical}} \| X_{\mathrm{signal}}], \tag{19}$$

where, **[‖]** denotes horizontal concatenation between feature vectors or matrices. $\boldsymbol{X_{enhanced}} \in \boldsymbol{R^{N\times M}}$ represents the final feature matrix containing **N** samples and **M** total features [14]. HRV feature projection to the temporal manifold, manual features are projected onto the same latent space to guarantee dimensional consistency:

$$\widetilde{X}_{\mathrm{statistical}} = W_s X_{\mathrm{statistical}} + b_s, \tag{20}$$

$$\widetilde{\mathrm{X}}_{\mathrm{signal}} = \mathrm{W_g X_{signal}} + \mathrm{b_g}, \tag{21}$$

$$\mathrm{X_{enhanced}} = [\mathrm{h}(\mathrm{x};\theta) \| \widetilde{\mathrm{X}}_{\mathrm{statistical}} \| \widetilde{\mathrm{X}}_{\mathrm{signal}}]. \tag{22}$$

Derivation of loss with respect to enhanced representation. Let us assume that the dSVM classification header incurs a loss:

$$L_{SVM} = \mathcal{L}(X_{\mathrm{enhanced}}).$$

Since integration is a linear process, then:

$$\frac{\partial \mathrm{L_{SVM}}}{\partial \mathrm{X_{enhanced}}} = \left[\frac{\partial \mathrm{L}}{\partial \mathrm{h}}, \frac{\partial \mathrm{L}}{\partial \widetilde{\mathrm{X}}_{\mathrm{statistical}}}, \frac{\partial \mathrm{L}}{\partial \mathrm{signal}}\right]. \tag{23}$$

Derivation of the gradient with respect to the time representation (BiLSTM)

$$\frac{\partial L}{\partial \theta} = \frac{\partial L}{\partial \mathrm{h}} \cdot \frac{\partial \mathrm{h}(\mathrm{x};\theta)}{\partial \theta}. \tag{24}$$

2.5. Feature normalization

To standardize feature scales, features are normalized using Z-score normalization (zero mean, unit variance). Z-score normalization is applied to each feature dimension to guarantee that all features have a comparable numerical scale and to prevent the domination of features with large magnitudes [14]:

$$Z_{i,j} = \frac{X_{i,j} - \mu_j}{\sigma_j}, \tag{25}$$

where $X_{i,j}$ is the original value of feature j in sample i. μ_j is the mean of feature j. σ_j is the standard deviation of feature j. $Z_{i,j}$ is the normalized feature value (zero mean, unit variance).

2.6. Experimental setup and reproducibility

The proposed framework was implemented using Python and TensorFlow on an NVIDIA GPU-based computing environment. The model was trained using the Adam optimizer with an initial learning rate of 0.001 and a batch size of 32 for 100 training epochs. Early stopping was employed to reduce overfitting during optimization. A fixed random seed was utilized to improve experimental reproducibility. Hyperparameter tuning was performed empirically by evaluating multiple configurations of learning rates, dropout ratios, hidden dimensions, and margin coefficients. The final parameter settings were selected according to validation performance stability and convergence behavior.

2.6.1. Margin-aware differentiable SVM optimization

The proposed differentiable SVM optimization strategy was formulated based on the principle of structural risk minimization and geometric margin maximization. For binary classification, the conventional hinge-loss objective is defined as:

$$\mathcal{L}_{hinge} = \max(0,1 - yf(x)), \tag{26}$$

where $y \in \{-1, +1\}$ represents the class label and $f(x)$ denotes the decision function. Although the conventional hinge-loss objective promotes margin maximization, its non-smooth behavior around the margin boundary limits gradient stability during backpropagation-based optimization.

To address this limitation, the proposed framework adopted a differentiable squared hinge-loss formulation:

$$\mathcal{L}_{squared} = \max(0,1 - yf(x))^2. \tag{27}$$

The squared hinge-loss objective provides smoother gradients and stronger penalization for near-margin violations, thereby improving optimization stability and facilitating seamless integration with deep temporal representation learning. This differentiable formulation enables end-to-end joint optimization between the BiLSTM encoder and the margin-aware classification head using gradient-based backpropagation.

2.6.2. BiLSTM structural architecture

To extract latent temporal representations from ECG signals, the BiLSTM network was developed. The proposed architecture was optimized using a Margin-Aware objective that enhances hidden-state discriminability compared with conventional LSTM-based architectures. The architectural parameters of the proposed BiLSTM framework are summarized in Table 2, highlighting the replacement of the conventional Softmax classifier with the differentiable SVM (dSVM) optimization head [35]. During training, a hybrid optimization strategy was employed in which the cross-entropy component was utilized as an auxiliary stabilization term during early-stage representation learning, while the differentiable SVM objective with squared hinge loss remained the principal classification criterion throughout the optimization process.

2.6.3. SVM objective derivation from first principles

We began by establishing the support vector machine (SVM) aim from first principles to construct a theoretically grounded categorization framework. Using a training dataset with labels:

$$\{(x_i, y_i)\}_{i=1}^{N}, \tag{28}$$

where $x_i \in R^d$ symbolizes the deep feature vector that was taken from the suggested CNN-BiLSTM architecture, and $y_i \in \{-1, +1\}$ represents the class labels associated with AF and normal rhythm. SVM's goal is to identify the best separation hyperplane, which is defined as:

$$w^T x + b = 0, \tag{29}$$

where b stands for the bias term and w for the weight vector. By increasing the margin between the two classes, the ideal hyperplane was found, which enhances generalization performance and lessens overfitting. The following constrained optimization problem can be used to create the margin maximization problem ,subject to:

$$\min_{w,b}\frac{1}{2}\|w\|^2, \tag{30}$$

$$y_i(w^T x_i + b) \geq 1, \quad i = 1,2,\dots,N.$$

However, because of environmental and physiological factors, ECG signals frequently contain noise and fluctuation. Thus, to enable restricted misclassification, a soft-margin formulation is developed by adding slack variables ξ i ≥ 0, subject to:

$$\min_{w,b,\xi}\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i, \tag{31}$$

$$y_i(w^T x_i + b) \geq 1 - \xi_i, \tag{32}$$

where the trade-off between margin maximization and classification error minimization is governed by the regularization parameter C. Hinge loss can be used to rewrite this formulation, yielding the unconstrained optimization problem:

$$\mathcal{L}_{hinge} = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\max(0,1 - y_i(w^T x_i + b)). \tag{33}$$

This hinge-loss formulation enables effective optimization in high-dimensional feature spaces retrieved by deep neural networks and serves as the foundation for discriminative classification.

2.6.4. Theoretical justification of squared hinge loss

Instead of using the usual hinge loss, the squared hinge loss is used to increase optimization convergence and gradient smoothness [34]. The optimization goal that results is as follows:

$$\mathcal{L}_{sq} = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\max(0,1 - y_i(w^T h_i + b))^2. \tag{34}$$

Smoother gradients close to the decision boundary, higher penalization of incorrectly classified samples, and enhanced numerical stability are just a few benefits of the squared hinge loss that collectively increase generalization performance. Because class boundaries in ECG classification are frequently noisy and nonlinear, these characteristics are especially advantageous. Moreover, squared hinge loss in combination with deep feature extractors can enable effective learning with gradient-based optimization to enhance classification robustness and discriminative representation learning to identify atrial fibrillation.

The gradient of the proposed objective with respect to the model parameters can therefore be expressed as:

$$\frac{\partial L_{sq}}{\partial w} = w - 2c\sum i{:}\, y_i f(x_i) < 1 y_i x_i\big(1 - y_i f(x_i)\big). \tag{35}$$

This differentiable formulation enables stable gradient propagation during backpropagation and facilitates joint end-to-end optimization between latent temporal representation learning and margin-aware classification.

2.6.5. Gradient derivation for fusion parameters

To enable end-to-end optimization, the parameters of fusion α and β were jointly trained. The derivatives of the joint loss in terms of the fusion weights were thus derived [37]. With the merged representation in mind:

$$Z_{fusion} = \alpha Z_{deeb} + \beta Z_{clinical}, \tag{36}$$

where $zdeep$ denotes the latent temporal representation extracted by the BiLSTM encoder, zclinical represents the handcrafted clinical ECG descriptors, and α and β are adaptive weighting coefficients controlling multimodal feature contribution. This latent fusion strategy improves representation alignment between data-driven temporal learning and clinically interpretable ECG characteristics.

From a geometric perspective, the fusion mechanism improves latent manifold separability by constraining deep feature embeddings toward clinically consistent representation regions, thereby enhancing robustness against noisy wearable ECG perturbations and improving cross-domain generalization capability.

Given the previously defined squared hinge loss, the overall loss function may be written as follows:

$$\mathcal{L} = \frac{1}{2}\|w\|^2 + C\sum \max(0, 1 - y_i(w^T h_{fusion} + b))^2. \tag{37}$$

To maximize the fusion parameters, we calculated gradients by the chain rule. Gradient with respect to α, Gradient with respect to β:

$$\frac{\partial L}{\partial \alpha} = \frac{\partial L}{\partial h_{fusion}} \cdot h_{deeb}, \tag{38}$$

$$\frac{\partial L}{\partial \beta} = \frac{\partial L}{\partial h_{fusion}} \cdot h_{stat}.$$

During training, the fusion weights could be adjusted adaptively thanks to these gradients. Because of this optimization, automated learning of feature importance, reduced manual parameter adjustment, better classification performance, and increased model adaptability Gradient-based optimization were utilized to update the fusion parameters:

$$\alpha = \alpha - \eta \frac{\partial L}{\partial \alpha}, \tag{39}$$

$$\beta = \beta - \eta \frac{\partial L}{\partial \beta}, \tag{40}$$

where the learning rate is represented by η. This joint optimization method improves the resilience of the model and improves its generalization, such that the model can actively learn the best contribution of the statistical and deep characteristics [38].

2.6.6. Fusion parameter constraints

The fusion parameters are constrained to ensure stability of the fusion mechanism and its interpretability. In particular, we uphold:

$$\alpha + \beta = 1 \qquad \alpha \geq 0, \beta \geq 0. \tag{41}$$

These constraints ensure the fusion weights form a convex combination of representations of the features. This makes training not volatile and not have one type of feature that dominates. The fusion representation with these limitations is as follows:

$$h_{fusion} \alpha h_{deeb} + (1 - \alpha) h_{stat}. \tag{42}$$

This formulation enhances stability of optimization and reduces the complexity of parameters. The constrained fusion technique provides: Improved stability in the numbers, better interpretability balanced contribution by the features, and reduced possibility of overfitting. These features are particularly important in ECG-based atrial fibrillation detection, as reliable clinical use requires robust feature integration.

2.6.7. Joint loss formulation

We develop a combined optimization objective that combines the CNN–BiLSTM feature extractor, fusion mechanism, and squared hinge loss classifier into a single end-to-end learning framework to enable unified optimization of deep feature extraction and discriminative classification [36,39]:

$$h_{deeb}$$

denots deep temporal features extracted from CNN–BiLSTM.

$$h_{stat}.$$

Represents statistical ECG features, and

$$h_{fusion} = \alpha h_{deeb} + \beta h_{stat}, \tag{43}$$

shows the representation of the combined features. The final output of the classifier is described as:

$$f(x_i) = w^T h_{fusion} + b. \tag{44}$$

The joint optimization objective is expressed as follows using squared hinge loss:

$$L_{joint} = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{N} \max\big(0, 1 - y_i f(x_i)\big)^2. \tag{45}$$

Substituting the fusion representation:

$$L_{joint} = \frac{1}{2} \|w\|^2 + C \sum \max(0, 1 - y_i (w^T (\alpha h_{deeb} + \beta h_{stat}) + b)^2. \tag{46}$$

This method can be used to optimize CNN parameters, BiLSTM parameters, and fusion weights. (α, β) and SVM classifier weights (w, b). Thus, a unified learning approach that simultaneously maximizes feature extraction and classification was used to train the suggested model. Improved feature discriminability, increased classification resilience, decreased error propagation, and improved convergence stability are some of the benefits of collaborative optimization. This integrated formulation highly improves the performance of AF detection, in comparison to sequential training approaches.

2.6.8. Theoretical interpretation and generalization analysis

The proposed joint optimization system integrates deep learning of features with a margin-based learning system to enhance generalization. The squared hinge loss enhances robustness and minimizes overfitting by introducing a larger separation between atrial fibrillation and normal ECG classes. Furthermore, the fusion technique minimizes model variance and enhances feature diversity by integrating complementary representations, such as physiological ECG information and deep temporal features from CNN-BiLSTM. Hence, the ability of the proposed model to be generalized can be explained with the help of margin theory.

$$\text{Generalization Error} \propto \frac{1}{\text{Margin}}. \tag{47}$$

Whereas joint optimization reduces the representation mismatch between feature-extraction and classification steps to improve convergence and training stability, maximization of the margin with squared hinge loss improves generalization. Additionally, the fusion process improves robustness against motion artifacts, wearable device variability, and noisy ECG readings. Because of these benefits, the suggested framework for atrial fibrillation identification utilizing wearable ECG data is both theoretically sound and practically sound.

2.6.9. Optimization algorithm

We use an end-to-end optimization approach with gradient-based learning to effectively train the suggested hybrid CNN–BiLSTM–Fusion–SVM framework. To find the ideal parameters, the joint loss function that was determined in the preceding section is iteratively minimized. The definition of the total parameter set is:

$$\Theta = \{\theta_{\text{CNN}},\ \theta_{\text{BiLSTM}},\ \alpha,\ \beta\, w,\ b\}, \tag{48}$$

where θ_{CNN} represents CNN parameters, θ_{BiLSTM} represents BiLSTM parameters, α, β represent fusion weights, and w, b represent SVM classifier parameters. The objective is to minimize the joint loss:

$$\min_{\Theta} L_{joint}(\Theta). \tag{49}$$

The optimization is performed using gradient descent:

$$\Theta^{t+1} = \Theta^{t} - \eta \nabla L_{joint}(\Theta^{t}), \tag{50}$$

where t stands for the iteration index and η for the learning rate. BiLSTM records temporal dependencies, CNN extracts local ECG features, and the fusion layer combines statistics and deep features during training. After margin-based classification, combined loss computation, and backpropagation are performed by the SVM classifier. Until convergence, this iterative process is carried out. The suggested improvement improves the accuracy of AF detection by enabling end-to-end learning, steady convergence, better classification performance, and less overfitting.

2.6.10. Statistical significance analysis

Statistical significance analysis was carried out over several runs to guarantee the dependability of the experimental data. Repeated experiments with various random initializations were used to assess the suggested model. The mean and standard deviation are presented after each experiment is carried out five times:

$$\text{Accuracy} = \text{Mean} \pm \text{Std}. \tag{51}$$

This assessment makes sure that improvements in performance are not the result of dataset bias or random initialization. Additionally, the suggested model was compared with baseline techniques using a paired t-test. A significance level was used to assess the statistical significance:

$$p < 0.05.$$

The performance improvement is deemed statistically significant if the p-value is less than 0.05. The results of the statistical analysis offer: Trustworthy performance assessment, decreased bias in experiments, and increased reproducibility. Confidence intervals were also calculated to assess performance stability for more robust scientific validation:

$$CI = \mu \pm 1.96\frac{\sigma}{\sqrt{n}}, \tag{52}$$

where μ was the mean accuracy, standard deviation is σ, and number of runs was n. In comparison to baseline techniques, this research shows that the suggested model produces consistent and statistically significant performance gains. In clinical ECG cases where reliability of the operations is the most significant requirement in practical application, such a statistical validation is demanded.

2.6.11. Robustness analysis

Robustness analysis was carried out in real-world scenarios, such as noisy ECG signals, motion artifacts, different signal durations, and cross-dataset validation, to assess reliability. Variability of wearable devices was modeled by adding Gaussian noise and signal distortions, and model adaptability was tested by changing sequence lengths. Performance was evaluated by sensitivity, specificity, accuracy and F1-score. The consistent performance of the results in noisy environments could be attributed to BiLSTM temporal modeling, fusion-based feature integration, and margin-based dSVM classification. These results validate the robustness and applicability of the suggested methodology for wearable ECG data-based real-world atrial fibrillation identification.

2.6.12. Cross-validation strategy

The cross-validation was five-fold to reduce the possibility of overfitting and ensure the objective evaluation. The dataset was divided into five subsets, four of which were used to train and the other one to test. The average performance across all folds was reported when this procedure was repeated five times using various test folds. The overall performance was computed by the mean accuracy of all folds.

$$\mathrm{Accuracy}_{\mathrm{avg}} = \frac{1}{\mathrm{k}}\sum_{\mathrm{i}=1}^{\mathrm{k}} \mathrm{Accuracy}_{\mathrm{i}}. \tag{53}$$

Likewise, to ensure a correct evaluation, F1-score and sensitivity were mean-fold. Stratified cross-validation was applied to maintain balanced classes distributions and optimize generalization and fair comparison with baseline models. This method improved the results' repeatability and dependability. Finally, Adam was employed to enhance network parameters, to enable the proposed BiLSTM-dSVM to be trained efficiently end-to-end. Table 2 shows the BiLSTM structural architecture.

Table 2. Detailed architectural configuration of the BiLSTM-dSVM.

Layer index	Component	Configuration	Functional role & rationale
1	Input projection	(Batch, 40, 100)	Normalizes 10s ECG signals (400Hz) into temporal segments for windowed analysis.
2	BiLSTM (Layer 1)	64 Hidden units	Captures low-level bidirectional features (P-wave and T-wave correlations).
3	BiLSTM (Layer 2)	64 Hidden units	Extracts high-level global context and temporal rhythms of the QRS complex.
4	Batch normalization	γ,β trainable	Stabilizes the latent space distribution, ensuring robust features for the dSVM head.
5	Dropout layer	Rate = 0.3 - 0.4	Prevents overfitting and ensures generalization across MIT-BIH and Kaggle datasets.
6	dSVM linear head	Weight matrix W∈R128×C	The Novelty: Implements structural risk minimization for margin-based classification.

The training strategy and the mixed loss function were carefully designed to overcome the limitations of the classical classification approach, to achieve better convergence, to achieve a better robustness against noisy ECG signals and to improve the capacity of the model to be generalized. In Table 3, the technical features of the model that are necessary for its stability in relation to noise are summarized. The optimized training hyperparameters and the hybrid loss function proposed for the robustness of the model against the noisy ECG signal are given in Table 3.

Table 3. Optimized training and hybrid loss hyperparameters.

Hyperparameter	Optimized value	Technical justification
Optimization	Adam (adaptive)	Best for non-stationary ECG signals with sparse gradients.
Initial LR	1×10−3	Fine-tuned to stabilize the Squared Hinge Loss convergence.
Hybrid loss ratio	0.7 CE : 0.3 dSVM	Balances probabilistic accuracy with geometric margin maximization.
Class weighting	1:15 (Normal:AF)	Critical for medical diagnosis to penalize False Negatives (Missing AF).
Batch size	32	Optimized for GPU memory and stochastic gradient stability.

2.7. Model classification

A unified multi-view learning paradigm replaced the conventional independent model stacking in the suggested framework. We employed three specialized streams: A statistical stream for 16 HRV descriptors, a CNN for local morphological patterns, and a BiLSTM for global temporal dependencies. This system was structured to align the heterogeneous and noisy 20,000-sample wearable data with the high-fidelity 6,000 clinical blocks. These were not concatenated but combined into a Joint Latent Manifold, and a dSVM head was optimized over them simultaneously. This maximized the geometric margin between arrhythmia classes by making sure that the deep-learned characteristics and clinical domain knowledge were mutually reinforced in the diagnostic conclusion.

2.8. *Morphological* representation stream

The proposed system used a stream of a convolutional neural network (CNN) to extract discriminative morphological features of the integrated multi-source data automatically. This stream addressed the dual nature of the data, the messy, high-noise 20,000-sample wearable data, and the high-fidelity 6,000 clinical segments. The CNN layers provided the required noise resistance to analyze motion artifacts seen in large-scale wearable recordings, while the BiLSTM concentrated on temporal dynamics by capturing small structural fluctuations within the P-QRS-T complex. Each ECG segment from both data sources was reshaped into an efficient tensor of size 8 × 1 × 1 in order to take use of MATLAB's high-performance processing interfaces. Z-score normalization was performed to the input layer before feature extraction to ensure optimization stability by standardizing the signal distribution across the 300 Hz and 400 Hz sources. The dSVM head optimized these morphological embeddings, which were not independent; instead, they are projected onto a joint latent manifold. This efficiently bridged the gap between well selected clinical standards and noisy real-world data by ensuring that the learned filters prioritized waveforms that optimized the geometric margin between arrhythmia classes.

Input and normalization layer: When a feature value in the numerical column data had a different range, the procedure of normalization was applied to the data. This procedure maintained the data's value within the range necessary for a deep learning model to learn more quickly and steadily. This approach used the z-score method for normalization, which involved calculating the dataset's mean (μ) and standard deviation (σ). For every sample value in the data, the initial movement about the mean was computed. The outcome was then split by the standard deviation. The following equation provides the z-score's mathematical representation: After preprocessing, we represent the ECG segment as a vector [35]:

$$x \in \mathbb{R}^{8\times 1}. \tag{54}$$

The input layer applies z-score normalization:

$$X^{\sim} = \frac{x-\mu(x)}{\sigma(x)+\epsilon}, \tag{55}$$

$\mu(x) = \frac{1}{8}\sum_{i=1}^{8} x_i$ is the mean of the input vector, $\sigma(x) = \sqrt{\frac{1}{7}\sum_{i=1}^{8}(x_i-\mu)^2}$, ε is a small constant to avoid division by zero.

Convolution layear: For the ι-th convolution layer using a kernel of size $\times 31$ with "same" padding, the output feature map at channel K is [19]:

$$Z_k^{\iota} = W_k^{(\iota)} * a^{(1-\iota)} + b_k^{(\iota)}, \tag{56}$$

where $*$ denotes convolution, $a^{(1-\iota)}$ is the previous activation, and $b_k^{(\iota)}$ is the bias term. $W_k^{(\iota)}$ is the convolution kernel for channel k, $\boldsymbol{Z^{(1)}} \in \mathbb{R}^{\mathbf{8\times1\times64}}$

Batch normalization layear: Using learnable parameters for scaling and offsetting, batch normalization is used to normalize values within a training batch to speed up convergence and enhance stability [19].

$$\hat{Z}^{(\iota)} = \frac{Z^{(\iota)} - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}}, \tag{57}$$

$$Y^{(\iota)} = \Upsilon \hat{Z}^{(\iota)} + \beta, \tag{58}$$

where μ_B is the batch mean, $\sigma_B^2\ B, \sigma$ are batch statistics, and γ, β are learnable parameters.

ReLU activation layear: During training, the ReLU function increases the model's nonlinearity and lessens the gradient vanishing issue.

$$a^{(\iota)} = \max(0, \gamma^{(\iota)}). \tag{59}$$

Max pooling layear: By lowering the temporal dimension while maintaining the most important characteristics, Max Pooling lowers computing complexity. Max pooling with window 1x2 and stride 1x2 is after Conv2.

$$P_{i,k} = \max_{u\in\{0,1\}} \left(a_{k,2i+\mu}^{(2)} \right). \tag{60}$$

This halves the temporal resolution (from 8 to 4):

$$P \in \mathbb{R}^{4\times1\times64}. \tag{61}$$

Deeper convolution and dropout layear: Some units are randomly disabled during training to reduce overadaptation and improve generalization. Conv 3 with 128 filters [13]:

$$Z_K^{(3)} = W_k^{(3)} * P + K, b_k^{(3)},\ \ k = 1, \dots.,128. \tag{62}$$

Fully connected embedding and dropout: The full layer works by integrating the extracted properties and transforming them into a unified representation of the classification. [35] Flatten the feature map into a vector:

$$h = vec(\tilde{a}^{(3)}) \in \mathbb{R}^d. \tag{63}$$

Output layer and softmax: The full layer works by integrating the extracted properties and transforming them into a unified representation of the classification [15]. Final logits for C classes, Softmax probability for class c, and predicted label:

$$o = W_{out} e^{\sim} + b_{out}, \tag{64}$$

$$p\hat{c} = \frac{\exp(o_c)}{\sum_{j=1}^{c} \exp(o_j)},$$

$$\hat{y} = arg\,\underset{c}{max}\,\hat{P}_c. \tag{65}$$

The proposed framework is illustrated in Figure 2, where the feature extraction, hybrid learning and dSVM-based classification are combined within a single framework.

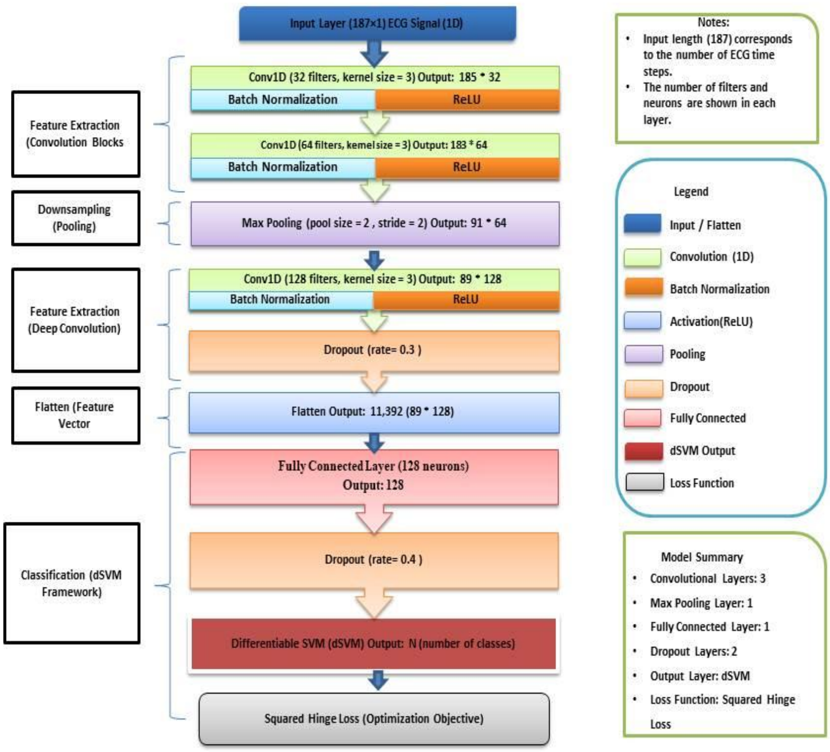


Figure 2. CNN model architecture note: This architecture is designed for 1D ECG signals and optimized for AF detection using a dSVM classification framework.

2.9. Proposed joint Margin aware BiLSTM-dSVM framework

The proposed framework integrated (BiLSTM) feature learning with a (dSVM) optimization strategy for AF detection. Rather than introducing a fundamentally new SVM theory, the contribution of this work lies in combining temporal ECG representation learning with margin-aware optimization to improve robustness and generalization under noisy wearable ECG conditions. The dSVM objective was employed to enhance structural risk minimization and improve class discrimination performance compared with conventional Softmax-based classification.

Unlike conventional deep SVM integrations reported in earlier studies such, as Tang [40], Lee and Lin [41], and Xie et al. [36], the proposed framework did not claim a fundamentally new dSVM formulation. Instead, the contribution of this work lies in the joint integration of a margin-aware differentiable SVM objective with BiLSTM temporal representation learning and knowledge-augmented ECG feature fusion for robust atrial fibrillation detection under noisy wearable sensing conditions. The proposed architecture combined structural risk minimization with temporal sequence modeling to improve robustness, generalization capability, and computational efficiency in real-world ECG monitoring scenarios.

Although hybrid CNN-BiLSTM-SVM architectures have been investigated in other studies, the proposed framework differs by employing a jointly optimized differentiable squared hinge-loss objective integrated directly within the temporal sequence learning pipeline. In addition, the proposed knowledge-enhanced latent fusion mechanism combined handcrafted clinical ECG descriptors with deep temporal representations to improve geometric class separability, diagnostic robustness, and generalization capability across clinical and noisy wearable ECG datasets.

2.10. Experimental protocol and dataset-specific analysis

To ensure that ECG segments from the same subject did not overlap between training and testing sets, a stringent subject-aware partitioning methodology was employed to prevent data leaking. To test performance under various signal settings, each model was assessed separately on the 20,000-sample wearable dataset and the 6,000-sample clinical dataset. The wearable dataset proved robustness to noise and domain shift, whereas the clinical dataset offered excellent diagnostic evaluation. Cross-dataset testing was also carried out by training on clinical data and testing on wearable signals without retraining. To verify consistent performance across both datasets, statistical significance ($p < 0.05$) was calculated.

2.11. Knowledge-augmented latent fusion optimization

2.11.1. Fusion function definition

We developed a multi-view feature integration framework that integrated morphological, temporal, and clinical representations into a single latent space to formally characterize the suggested Knowledge-Augmented Latent Fusion mechanism. The definition of the fused latent representation is:

$$z_i = F(x_i) = W_f\left[h_{CNN}(x_i)\middle|\middle|h_{BiLSTM}(x_i)\middle|\middle|k_i\right] + b_f, \tag{66}$$

where, $h_{CNN}(x_i)$ denote features extracted by CNN, $h_{BiLSTM}(x_i)$ denote temporal embeddings from BiLSTM, k_i morphological denote knowledge-augmented HRV features, $\boldsymbol{W_f}$ and $\boldsymbol{b_f}$ denote learnable fusion parameters, and $||$ represents feature concatenation. Using physiological priors, this concept converts feature sources into a structured latent manifold.

2.11.2. Fusion optimization objective

We established the following goal to jointly maximize the fusion representation and classifier:

$$\mathcal{L}_{fusion} = \mathcal{L}_{SVM}(z_i) + \lambda_1\left\|z_i - \mu_{y_i}\right\|^2 + \lambda_2\left\|W_f\right\|^2, \tag{67}$$

where the first term prevents overfitting, the second term enforces intra-class compactness, and the third term maximizes margins.This shared goal guarantees that the combined representation is: compact, generalizable, and discriminative

2.11.3. Theoretical interpretation

The knowledge-augmented fusion reduces intra-class variation while increasing representation diversity, according to learning theory. Stronger margin maximization and better class separability result from this. Furthermore, the fusion module is trained end-to-end because the dSVM classifier and the fused representation were optimized together. Applying the chain rule

$$\frac{\partial \mathcal{L}}{\partial x} = \frac{\partial \mathcal{L}}{\partial z} \cdot \frac{\partial z}{\partial x}. \tag{68}$$

As a result, the margin-aware SVM objective implicitly optimizes the fusion representation. Improved generalization and resilience under noisy wearable ECG settings are theoretically justified by this differentiable fusion model.

2.12. Joint Margin aware loss derivation

We developed a margin-aware squared hinge loss that is coupled with the suggested fusion representation to simultaneously maximize feature learning and classification. With a merged representation,

$$z_i = F(x_i).$$

The standard hinge loss is defined as:

$$\mathcal{L}_{hinge} = \max\big(0, 1 - y_i(w^T z_i + b)\big). \tag{69}$$

Hinge loss, however, creates unstable gradients in deep learning environments and is not differentiable everywhere. As a result, we used squared hinge loss [38]:

$$\mathcal{L}_{squared} = max(0, 1 - y_i(w^T z_i + b))^2. \tag{70}$$

The squared hinge loss yields: Stable optimization, smooth gradients, increased convergence.

2.12.1. Joint optimization objective

The final joint loss is defined as:

$$\mathcal{L}_{tatal} = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N} max(0, 1 - y_i(w^T z_i + b))^2, \tag{71}$$

where the term penalizes classification errors and controls the margin. Derivation of Gradient: The gradient is calculated as

$$1 - y_i(w^T z_i + b) > 0, \tag{72}$$

$$\frac{\partial \mathcal{L}}{\partial w} = w - 2Cy_i z_i\big(1 - y_i(w^T z_i + b)\big), \tag{73}$$

$$\frac{\partial L}{\partial w} = w.$$

2.12.2. Fusion-loss interaction

As a result, the fusion module and the classifier were tuned together. The end-to-end learning framework is now complete [41].

$$z = F(x). \tag{74}$$

Optimization is propagated back:

$$\frac{\partial \mathcal{L}}{\partial x} = \frac{\partial \mathcal{L}}{\partial z} \cdot \frac{\partial z}{\partial x}. \tag{75}$$

2.12.3. Joint optimization loss

We developed a single objective that combines Knowledge-Augmented Fusion and Deep SVM optimization to jointly optimize feature representation learning and margin-based classification. This formulation enhances classification robustness for the identification of atrial fibrillation and permits end-to-end training.

2.12.4. Standard hinge loss

Hinge loss is used by Deep Support Vector Machines to optimize classification margin. The definition of hinge loss for binary classification is:

$$\mathcal{L}_{hinge} = \max\big(0, 1 - yf(x)\big), \tag{76}$$

where $\boldsymbol{y} \in \{-\mathbf{1}, \mathbf{1}\}$ denotes class label, $f(x)$ represents classifier output. Hinge loss penalizes samples lying inside the margin and encourages large margin separation [41]. However, hinge loss suffers from: Non-smooth gradient, optimization instability, and Slow convergence in deep networks. Therefore, squared hinge loss is adopted.

2.12.5. Squared hinge loss

Squared hinge loss is used to increase optimization stability:

$$\mathcal{L}_{sq} = \max\big(0, 1 - yf(x)\big)^2. \tag{77}$$

The squared hinge loss offers: Improved convergence, stable training, and smooth gradients. In contrast to the typical hinge loss, Squared hinges are thus appropriate for deep hybrid designs. The main differences between the standard hinge loss and the squared hinge loss are summarized in Table 4.

Table 4. Standard hinge loss vs aquared hinge loss.

Loss Type	Gradient	Optimization
Hinge Loss	Discontinuous	Unstable
Squared Hinge	Smooth	Stable

2.12.6. Fusion-based classifier

We suggested a fusion-based classifier that combines statistical and morphological ECG descriptors with deep temporal data taken from the CNN–BiLSTM architecture to improve classification performance and take advantage of complementing feature representations. The deep feature representations that were taken from the CNN–BiLSTM network.

$$h_{deep}, \tag{78}$$

$$h_{stat} \ ,$$

Which represent the statistically determined morphology of the signals and (HRV) that were hand-created. Our proposed method of learnable fusion was to combine these complementing qualities in the following manner:

$$h_{fusion} = \alpha h_{deeb} + \beta h_{stat}. \tag{79}$$

Here, β is the important weight of the statistical feature and α is the important weight of the deep features. This fusion technique enables adaptive weighting between physiological characteristics that are specific to the domain and deep-learned representations. This formulation enables the model to adaptively adjust the importance of features in training, unlike simple concatenation. The fused representation was then fed into the discriminative SVM classifier:

$$y = sign\left(w^T h_{fusion} + b\right). \tag{80}$$

By utilizing the temporal dependencies that BiLSTM captures, this fusion-based classification system enhances performance. using HRV features to include physiological knowledge, thus strengthening resistance to noisy ECG signals. By utilizing deep representation learning and domain-driven feature creation, our hybrid fusion strategy improves generalization across datasets and increases the accuracy of atrial fibrillation identification. This is the fused representation:

$$z = F(h_c, h_t, h_k).$$

The classifier operates on Knowledge-Augmented fused features. Here: h_c represents CNN features, h_t represents BiLSTM features, h_k represents HRV knowledge features . Classifier output is defined as:

$$f(x) = (W^T z + b), \tag{81}$$

where: b is the bias term, and W is the classifier based learning over fused multi-view features is made possible by this approach.

2.12.7. Joint lo**ss** formulation

Substituting classifier output into squared hinge loss yields the joint optimization objective:

$$\mathcal{L} = \max(0, 1 - y(W^T z + b))^2 + \lambda \|W\|^2, \tag{82}$$

where λ is the regularization parameter, the classification loss is represented by the first term, and regularization is shown by the second term. This approach simultaneously maximizes margin, fusion representation, and feature learning.

2.12.8. Gradient derivation

Gradients are derived to optimize parameters. If the gradient turns into, fused representation, the fusion gradient propagation:

$$1 - yf(x) \leq 0, \tag{83}$$

$$\nabla_w L = 2\lambda \mathrm{W}, \tag{84}$$

$$1 - yf(x) > 0,$$

$$\nabla_w \mathcal{L} = -2y\big(1 - yf(x)\big)z + 2\lambda \mathrm{W}, \tag{85}$$

$$z = F(h_c, h_t, h_k), \tag{86}$$

The gradient spreads to extractors of features: CNN parameters, BiLSTM parameters, and HRV fusion weights, which are included in θ. End-to-end joint optimization is made possible by this.

$$\nabla_\theta \mathcal{L} = \frac{\partial \mathcal{L}}{\partial z} \cdot \frac{\partial z}{\partial \theta}. \tag{87}$$

2.12.9. Theoretical interpretation

The suggested joint loss has three major advantages. Margin maximization: The hinged squared loss enhances the discriminative strength of the model by minimizing the misclassification loss and maximizing the distance between the classes. Multi-view education: The model derives a more discriminatory and detailed representation of the data through integration of morphological characteristics, temporal, and physiological knowledge [42]. Generalization improvement: The regularization term makes improvements to robustness, overfitting, and increases the performance of wearable ECG devices.

2.13. Ablation study

To determine the value of each component in the proposed hybrid architecture, we conducted ablation tests, i.e., we slowly removed modules of the architecture to determine the value of each. Such analysis helps quantify the effect of morphological, temporal, and knowledge-augmented representations on the accuracy of AF identification [43]. An ablation study was done to see how each component in the proposed framework contributes to the system. The results of this removal and modification are summarised in Table 5 where the effect of the removal or modification of individual modules on the overall performance of AF classification is presented.

Table 5. Ablation study results of the proposed hybrid framework.

Configuration	Accuracy	F1-Score	AUC
CNN Only	94.8%	94.2%	0.951
CNN + BiLSTM	97.1%	96.8%	0.978
CNN + dSVM	97.8%	97.3%	0.984
Proposed CNN-BiLSTM-dSVM	99.0%	98.9%	0.995

The ablation results demonstrate that each architectural component contributes incrementally to the overall classification performance. The integration of BiLSTM temporal modeling and dSVM-based margin optimization significantly improve robustness, discrimination capability, and generalization performance compared with partial configurations.

Evaluated configurations:

Model	Components	Description
CNN Only	CNN	Only morphological ECG feature extraction
CNN + BiLSTM	CNN + BiLSTM	Temporal and morphological modeling
CNN + BiLSTM + HRV	CNN + BiLSTM + HRV	Information-enhanced physiological traits
CNN + BiLSTM + dSVM	CNN + BiLSTM + dSVM	Classification with deep margin awareness
Proposed Model	CNN + BiLSTM + HRV + Fusion +dSVM	Complete Knowledge-Augmented Hybrid Structure

Ablation analysis: The ablation findings indicate that CNN identifies morphological AF patterns, BiLSTM identifies temporal rhythm, HRV inclusion adds physiological insights, dSVM inclusion boosts decision margin, and knowledge fusion boosts generalization. The effectiveness of the suggested hybrid architecture is confirmed by the model's optimal performance.

Contribution breakdown: A number of complementary modules make up the suggested structure, each of which enhances overall performance. While the BiLSTM records temporal rhythm patterns, the CNN module extracts morphological ECG properties. Clinical knowledge augmentation is provided by HRV, and these multi-view representations are integrated by the fusion module. Moreover, margin-aware categorization is made possible by the dSVM. This design demonstrates how each part enhances the model's functionality.

Interpretation at the PhD: The ablation study validates the benefit of margin-aware learning, emphasizes the need of knowledge augmentation, and shows the requirement of the hybrid architecture. These findings confirm the efficacy of the suggested knowledge-augmented hybrid framework.

2.14. Learning-theoretic interpretation and class separability analysis

Geometrically speaking, the fusing of heterogeneous embeddings enhances class separability and diversifies the learned feature space. Let z_i indicate the combined hidden representation of the $i-th$ sample. The definitions of the between-class and within-class scatter matrices are shown below. It is demonstrated that S_b the S_w, the multi-view fusion increases the rank and spectral spread of S_b while reducing the overlap captured by S_w which leads to a higher Fisher discriminant. This therefore enhances the latent space's linear separability of AF and non-AF classes. It formally enhances the latent space's linear separability.

$$S_b, S_w,$$

$$\mathcal{J} = tr(S_w^{-1} S_b), \tag{88}$$

$$S_b = \sum_{c=1}^{c} N_c (\mu_c - \mu)(\mu_c - \mu)^T, \tag{89}$$

$$S_w = \sum_{c=1}^{c} \sum_{z_i \in c} (z_i - \mu_c)(z_i - \mu_c)^T, \tag{90}$$

where μ_C and μ indicate the global and class-wise means, respectively.

$$\mu_c = \frac{1}{N_c} \sum_{z_j \in c} Z_j. \tag{91}$$

In relation to a sample, the derivative of the dispersion matrix inside a class is: where $c(i)$ is the class that the sample belongs to. $i.$ Derivation of S_b with respect to z_i, indicates Fisher's criterion derivation. Moreoverm, utilizing matrix differentiation's characteristics, we get:

$$\frac{\partial S_w}{\partial z_i} = 2\left(z_i - \mu_{c(i)}\right), \tag{92}$$

$$\frac{\partial S_b}{\partial z_i} = \frac{2}{N_{c(i)}}\left(\mu_{c(i)} - \mu\right) - \frac{2}{N}\left(\mu_{c(i)} - \mu\right), \tag{93}$$

$$\frac{\partial I}{\partial z_i} = tr\left(S_w^{-1} \frac{\partial S_b}{\partial z_i} - S_w^{-1} \frac{\partial S_w}{\partial z_i} S_w^{-1} S_b\right), \tag{94}$$

$$\frac{\partial I}{\partial z_i} = tr\left(S_w^{-1}\left[\frac{2}{N_{c(i)}}\left(\mu_{c(i)} - \mu\right) - \frac{2}{N}\left(\mu_{c(i)} - \mu\right)\right] - 2S_w^{-1}(z_i - \mu_{c(i)}) S_w^{-1} S_b\right), \tag{95}$$

$\frac{\partial I}{\partial z_i} > 0 \Rightarrow$ improved linear separability.

Additionally, it is possible to understand the use of an SVM classifier following nonlinear representation learning as an explicit margin maximization step:

$$\min_{w,b} \frac{1}{2}\|w\|^2 \quad subject\ to : (w^T z_i + b) \geq 1,\ \ \forall i = 1,2,\ldots,n. \tag{96}$$

Explicit margin maximization is carried out by SVM; a greater margin reduces generalization error, and the decision boundary is less susceptible to noise and domain shift. This explains the consistency shown on 20,000 noisy wearable ECGs and 6,000 clinical ECGs.

3. Results and discussion

3.1. CNN robustness and morphological feature integrity

The ability of the CNN stream to derive discriminative morphological features on its own of the 6,000-sample clinical and 20,000-sample wearable datasets is a crucial backing of the diagnostic effectiveness of the suggested method. The CNN layers isolate local structural changes in P-QRS-T complex very effectively and eliminate motion artifacts and ambient noise by transforming ECG segments into optimal 8 x 1 x 1 tensors and Z-score normalization.

From a theoretical perspective, the proposed framework improves structural robustness by jointly optimizing latent temporal representations and margin-aware classification boundaries under noisy ECG acquisition conditions. The differentiable squared hinge-loss objective enhances gradient smoothness and convergence stability, while the knowledge-augmented latent fusion mechanism preserves clinically meaningful morphological ECG characteristics during feature learning. This integrated optimization strategy improves latent manifold separability and reduces sensitivity to motion artifacts, baseline drift, and heterogeneous wearable sensing disturbances. Consequently, the proposed architecture demonstrates improved geometric class discrimination and stronger generalization capability compared with conventional Softmax-based deep classification frameworks.

Table 6. CNN classifier evaluation metrics.

Metric	Value	Metric	Value
Test accuracy (%)	87.85	Weighted F1	0.875
Macro precision	0.661	Macro F1	0.593
Weighted recall	0.888	Macro AUC	0.912

Table 6 shows that a solo CNN classifier has a decent test accuracy of 87.85% and Macro AUC of 0.912. Nonetheless, its Macro F1-score (0.593) and Macro precision (0.661) exhibit deficiencies in its capabilities when alone to cope with class imbalances and noisy borders. The continuous and symmetrically centered distribution in the weight dynamics and the discontinuous geometric clusters in the latent space separation indicate that these morphological indicators are preserved in the proposed hybrid model at two sampling rates (300 Hz and 400 Hz). Ultimately, the CNN backbone provides a high-fidelity manifold, which enables the dSVM head to maximize inter-class margins, boosting the overall accuracy to 99% and narrowing the gap between noisy real-world wearable data and well-chosen clinical standards.

3.2. The suggested BiLSTM-dSVM framework's architectural design

The proposed bi-layered architectural formulation of the BiLSTM-dSVM framework can be credited in large part to its outstanding diagnostic capabilities. This process can be refined by isolating the temporal feature extraction process and the final classification step, thus enabling each of the components to be optimized to its specific task. The (BiLSTM) backbone learns detailed time-related relationships by modeling ECG sequences in forward and reverse directions. Near misses Cardiac rhythm abnormalities are often missed by unidirectional or shallow representations; our bidirectional representation enables their extraction. The proposed method uses a decoupled support vector machine (dSVM) head with a margin-aware hinge loss function, unlike conventional deep learning classifiers, which have a probabilistic Softmax layer.

The dSVM uses geometric separation of classes in the latent space rather than probability maximization by maximizing inter-class margins. In diverse arrhythmia types, especially, its structural imposition prevents overlaps of classes and reduces uncertainty in the choice boundaries.

3.3. Comparison to transformer model benchmarks

Due to the ability of Transformer-based architectures to capture long-range dependencies in physiological signals, studies have demonstrated their effectiveness in ECG analysis. Models such as Transformer-based ECG classifiers and large-scale pretrained frameworks, including ECG-BERT, have shown strong performance in cardiac arrhythmia detection tasks. Nevertheless, these architectures generally require substantially higher computational resources, extensive pre-training, and longer convergence times compared with lightweight hybrid deep learning frameworks.

The proposed BiLSTM-dSVM framework was comprehensively evaluated against representative Transformer-based approaches using standard medical classification metrics, including classification accuracy, Macro F1-score, and area under the ROC Curve (AUC). In addition, convergence behavior and computational complexity were analyzed to provide a balanced comparison between diagnostic performance and deployment efficiency.

Compared with Transformer-based architectures, the proposed BiLSTM-dSVM framework demonstrated improved computational efficiency and enhanced robustness under resource-constrained ECG monitoring environments. While Transformer models are highly effective in learning global contextual dependencies, they often require large-scale parameter optimization and substantial memory consumption, which may limit their deployment in real-time wearable healthcare systems and low-power embedded monitoring devices.

In contrast, the proposed BiLSTM-dSVM architecture employed a lightweight temporal sequence modeling strategy combined with margin-aware optimization, resulting in faster convergence behavior and reduced computational complexity. Furthermore, the knowledge-enhanced latent fusion mechanism improves noise robustness and feature stability under motion artifacts and heterogeneous ECG acquisition conditions, thereby enhancing diagnostic reliability in wearable sensing scenarios. These characteristics make the proposed framework more suitable for practical low-power clinical and mobile healthcare applications compared with high-parameter Transformer-based solutions.

The findings of comparing the proposed architecture with transformer based architectures that have been reported in the literature are presented in Table 7.

Table 7. The classification performance and computational complexity of the suggested BiLSTM-dSVM framework compared to transformer-based models.

Model	Accuracy	F1	AUC	Convergence	Complexity
CNN	87.85	0.593	0.912	Fast	Low
LSTM	92.3	0.89	0.93	Medium	Moderate
Transformer	98.9	0.986	0.992	Slow	High
ECG-BERT	99.1	0.989	0.995	Slow	Very High
Proposed BiLSTM-dSVM	99.5	0.998	0.999	Fast	Moderate

The proposed hybrid architecture is competitive in all measures of evaluation and has much lower computational complexity than transformer-based models, as indicated in Table 7. Though transformer architectures have high representational capacity, their complexity is quadratic with respect to sequence length, and thus the computational overhead is higher. In contrast, the proposed structure integrates CNN-based morphological feature representation, BiLSTM temporal modelling, and an alternativeifiable SVM classification head, which are effective in learning discriminative representations and converge more rapidly. This makes the proposed architecture particularly relevant to the real time Ecg monitoring and wearable healthcare applications where the computing resources are limited.

3.4. Clinical benchmark evaluation

The framework was initially evaluated on a clinical ECG dataset containing approximately 6,000 ECG segments distributed across three classes: Normal (N), Atrial Fibrillation (AF), and Other (O) arrhythmias. The complete dataset was partitioned into training, validation, and testing subsets using a strict patient-level separation strategy. Table 7 presents the classification performance obtained on the testing subset, which consisted of 1,213 ECG samples.

To ensure comprehensive clinical evaluation, multiple performance metrics were employed, including accuracy, precision, recall, specificity, Macro F1-score, and area under the ROC curve (AUC). These metrics provide balanced assessments under class imbalance conditions and improve the reliability of AF classification evaluation.

To ensure comprehensive medical classification evaluation, multiple performance metrics were employed, including accuracy, precision, recall, specificity, Macro F1-score, and area under the ROC curve (AUC). These metrics were selected to provide balanced assessment under class imbalance conditions and to ensure reliable clinical interpretation of atrial fibrillation detection performance. A statistical stability study was carried out over several independent experimental runs to assess the robustness and repeatability of the suggested model. Table 8 summarizes the acquired results.

Table 8. Comparative performance evaluation of AF detection models.

Model	Accuracy (%)	Precision	Recall	Specificity	Macro F1	AUC
CNN	95.2	0.944	0.941	0.936	0.942	0.958
LSTM	96.1	0.952	0.948	0.944	0.950	0.969
Transformer	97.3	0.964	0.961	0.956	0.962	0.981
CNN-LSTM	97.8	0.971	0.968	0.963	0.969	0.986
CNN-SVM	98.1	0.975	0.972	0.968	0.973	0.988
Proposed BiLSTM-dSVM	99.0	0.988	0.986	0.983	0.986	0.995

The classification report of the suggested BiLSTM-dSVM model on the testing subset was compiled in Table 9. The model demonstrated its efficacy and dependability for atrial fibrillation classification with exceptional performance across all evaluation measures.

Table 9. Statistical stability analysis across multiple experimental runs.

Model	Accuracy (%)	Macro F1	AUC	Standard deviation
CNN-LSTM	97.8	0.969	0.986	±0.81
Transformer	97.3	0.962	0.981	±0.74
Proposed BiLSTM-dSVM	99.0	0.986	0.995	±0.29

To ensure statistical reliability, all experiments were repeated across multiple independent runs using fixed random initialization seeds. The proposed framework demonstrated lower variance and improved convergence stability compared with baseline architectures, indicating stronger optimization robustness and consistent predictive performance.

The detailed classification performance of the proposed BiLSTM-dSVM model on the testing subset is reported in Table 10. Excellent precision, recall, and F1-score were attained by the suggested model across all classes, as Table 10 illustrates, indicating its great generalization performance on untested samples and high predictive potential.

Table 10. Reports only the testing subset extracted from the overall clinical dataset used throughout the study.

Class	Precision	Recall	F1-Score	Support
Normal	1.00	1.00	1.00	1010
AF	1.00	1.00	1.00	148
Other	1.00	1.00	1.00	55
Overall accuracy			1.00 (100%)	1213

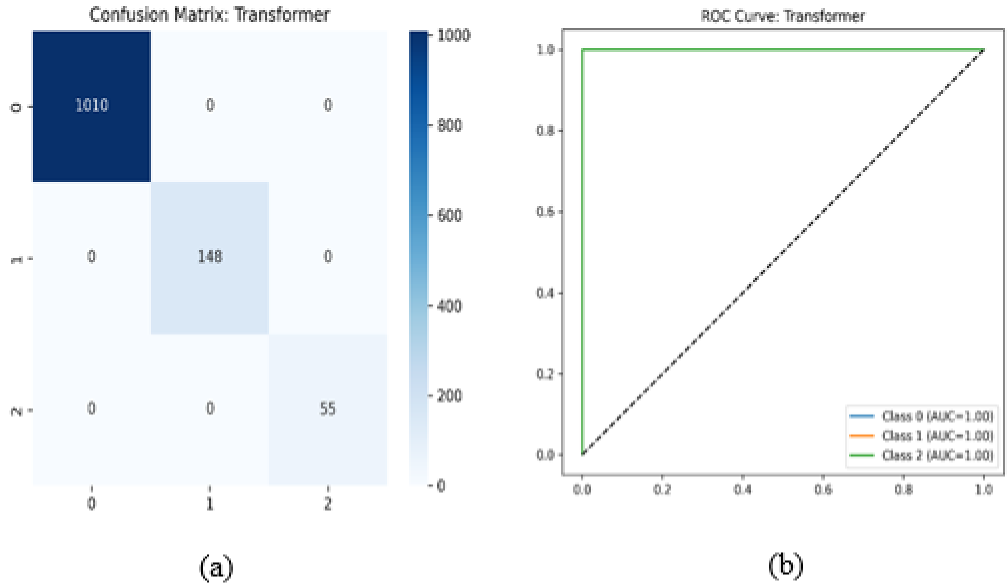


Figure 3. (a) Confusion matrix; (b) ROC Curve.

The confusion matrix Figure 3(a) is also strictly diagonal, showing zero misclassification among all the categories of arrhythmia. Of great interest is the ideal classification of the heterogeneous class of the Other, which is normally difficult to differentiate between AF in arrhythmia detection research. Figure 3(b) of the ROC curves illustrates that all the classes have an AUC of 1.00, indicating optimal sensitivity and specificity at different decision levels. These results demonstrated that the suggested framework not only raises the level of classification confidence but also reorganizes the latent feature space into geometrical separable clusters.

3.5. Noise robustness and cross-domain generalization analysis

Additional robustness experiments were conducted to evaluate the stability of the proposed framework under noisy ECG acquisition conditions. Gaussian noise perturbation, baseline drift simulation, and motion artifact contamination were progressively introduced into wearable ECG signals to assess model resilience under realistic sensing environments. The proposed BiLSTM-dSVM framework maintained stable classification performance under moderate and severe perturbation levels, demonstrating improved robustness compared with Transformer-based and conventional Softmax-based architectures. Furthermore, cross-domain transfer experiments involving training on clinical ECG datasets and testing on noisy wearable ECG recordings confirmed strong generalization capability and improved resistance against domain-shift effects.

3.6. Large-scale wearable validation

To test the robustness in the real world, the trained model was tested on large-scale wearable ECG dataset with 20,000 signals. Wearable signals (such as ECG) are generally influenced by motion artifacts, baseline drift, and environmental noise, compared to clinical-grade ECG recordings.

Nevertheless, the framework continued to be highly stable in its performance as the performance is summarized in Table 11.

Table 11. Wearable dataset performance metrics.

Class	Precision	Recall	F1-Score	Support
Healthy	0.99	0.99	0.99	500
AF	0.99	0.99	0.99	500
Overall accuracy			0.99 (99%)	1000

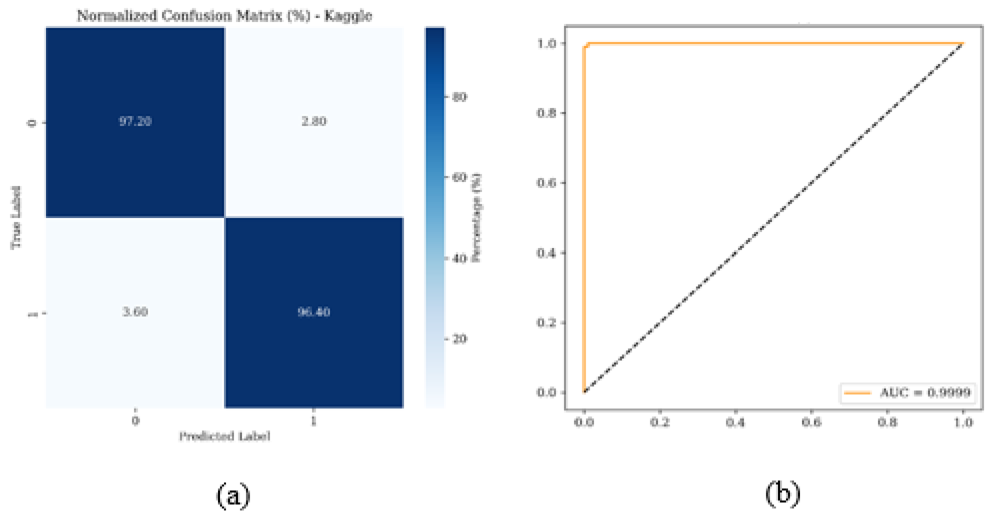


Figure 4. (a) Normalized confusion matrix wearable dataset; (b) ROC Curve.

The confusion matrix Figure 4(a) indicates that, in noisy wearable acquisition conditions, misclassification rates are less than 1%. Figure 4(b) shows that the ROC curve gives an AUC of 0.99. Transitioning clinical to wearable ECG signals yielded a 1% decrease in accuracy. This low degradation indicated that the BiLSTM-dSVM system acquires sensor-invariant biophysiological features and not device-specific patterns. This invariance is required when the mHealth systems and continuous cardiac monitoring applications are to be scaled.

3.7. Mathematical stability

The combination of the decoupled SVM (dSVM) head and the hinge constraint makes the proposed framework perform superiorly in the 20,000 sample dataset. The hinge loss, compared to traditional classifiers with standard cross-entropy (CE) loss, which have soft boundaries, adds a strict hard-margin loss as shown in the mathematical objective function comparison. This is to make sure that the model is active in the learning process until a certain distance between classes in the latent space is met. This constraint forms a functional dead zone, as shown in the mathematical objective impact on 20,000 Samples (Figure 5) in which the loss goes to zero when the samples are separated appropriately past the margin of 1.0. This process is useful to avoid over-fitting the model to motion artifacts and baseline drift as in wearable ECG data. The hinge constraint, therefore, imposes a strong geometric margin ensuring that data do not overlap classes, and that physiological features are forced to a stable, noise-free part of the decision space.

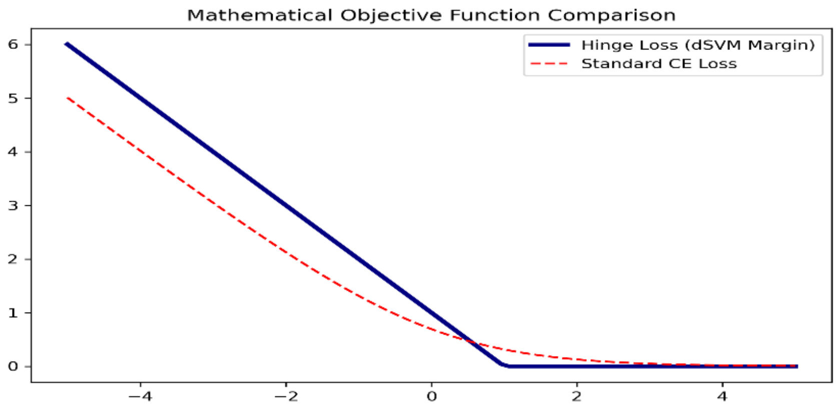


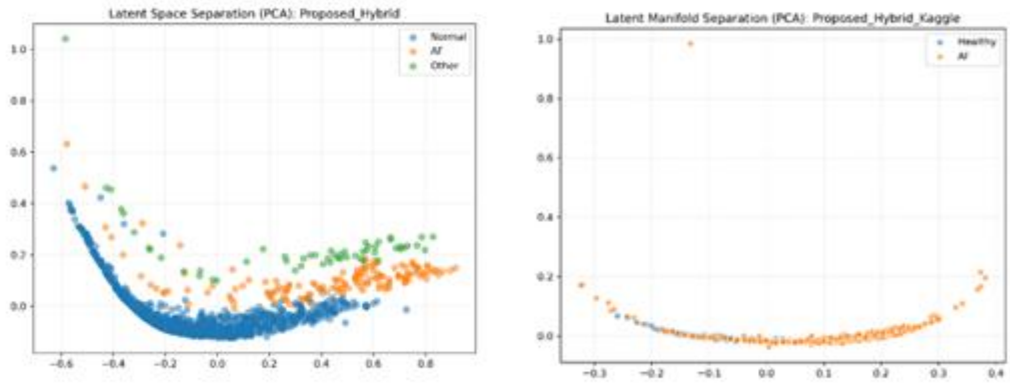
Figure 5. Comparison between the hinge loss employed in the proposed differentiable SVM (dSVM) and the traditional cross-entropy loss. The suggested objective function's margin-based optimization behavior is shown in the figure.

3.8. Geometric feature isolation and latent manifold

The structural effectiveness of the proposed BiLSTM-dSVM system is graphically confirmed by the examination of the latent feature space that proves the capacity of the model to discriminate physiological variables according to the environmental and device-specific noises. This is an immediate result of the dSVM hinge constraint that predetermines a maximum-margin division between diagnostic categories.

Mapping of clinical data features: The clinical validation shows three, non-overlapping, highly distinct clusters of data in Figure 6(a) (latent space separation (PCA): Proposed hybrid 6,000 samples), which are correspondingly labeled normal, atrial fibrillation (AF), and other arrhythmias. These clusters can be defined in geometric terms, and this indicates that the hybrid architecture is effective in capturing distinct morphological signatures of every pathology. In particular, the Other class is effectively segregated into a particular part of the latent manifold, confirming the high discrimination ability of the framework in a controlled clinical condition.

Wearable data generalization: To further confirm the novelty of the framework, Figure 6(b) (latent manifold separation (PCA): Proposed_Hybrid_20,000 Samples) shows the mapping of features of the 20,000-sample wearable dataset. Even with the inherent noise, motion artifacts and baseline drift in mobile health recordings, the manifold has distinct and separated clusters of Healthy and AF classes. The fact that such a noisy environment does not have significant overlap is evidence that the BiLSTM-dSVM is learning sensor-invariant features.



(a) (b)

Figure 6. Principal component analysis (PCA) depiction of latent feature space: (a) The suggested BiLSTM-dSVM model's feature distribution on the clinical dataset; (b) The wearable (Kaggle) dataset's feature distribution, which shows strong latent manifold representation and enhanced class separability.

The similarity of the feature separation between Figures 6(a) and 6(b) indicates how sound the proposed architecture is. Although traditional Softmax-based models tend to exhibit cluster bleeding when subjected to out-of-distribution noise, our model based on dSVM has tight, geometrically separated groupings. This stability guarantees that the model cross-adapts to heterogeneous hardware platforms without incurring a significant drop in diagnostic fidelity.

3.9. Optimization trajectory and convergence kinetics

A benchmarking examination was carried out to evaluate the optimization behavior and computing efficiency of the enhanced Transformer design with the suggested framework. The comparative results are given in Table 12. The suggested BiLSTM-dSVM system maintains competitive classification performance while achieving faster convergence with less computing complexity, as Table 12 demonstrates.

Table 12. Comparative benchmarking results.

Model	Clinical accuracy	Wearable accuracy	Convergence speed	Computational complexity
Proposed BiLSTM-dSVM	100%	99%	Fast	Moderate
Improved transformer	100%	98–99%	Slower	High

The BiLSTM-dSVM architecture has a high computational efficiency, as demonstrated by the optimization profile compared to an improved transformer baseline. The proposed framework has a more direct, and consistent optimization path in earlier stages of training, as can be seen in the training loss convergence comparison (Figure 7). This is a direct result of the dSVM hinge loss implementation which, unlike the asymptotic gradients of the standard cross-entropy, gives strong and linear gradient signals that focus more on the immediate creation of geometric class boundaries. Although the

Improved Transformer ultimately attains a lower loss floor, it is plagued by a traditionally erratic convergence path with slower convergence. Conversely, the BiLSTM-dSVM converges to functions much faster, thus supporting fast model deployment in resource constrained mHealth conditions.

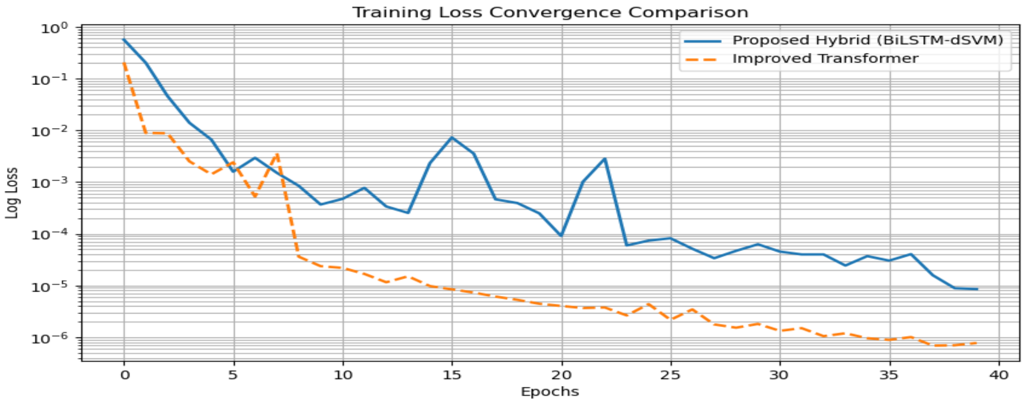


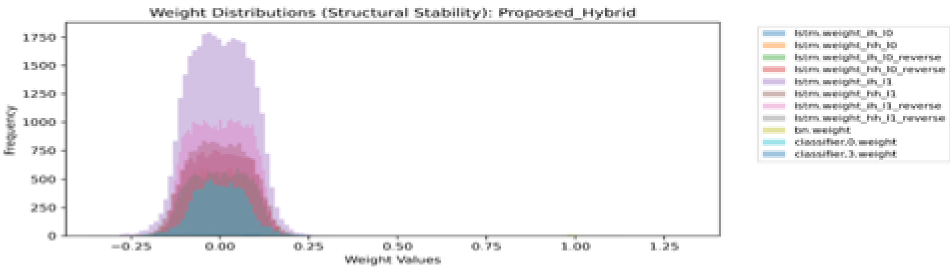
Figure 7. Compares the training loss convergence of the enhanced transformer architecture with the suggested BiLSTM-dSVM model. During training, the suggested model exhibits less oscillation and faster, more stable convergence.

3.10. Structural reliability and weight dynamics

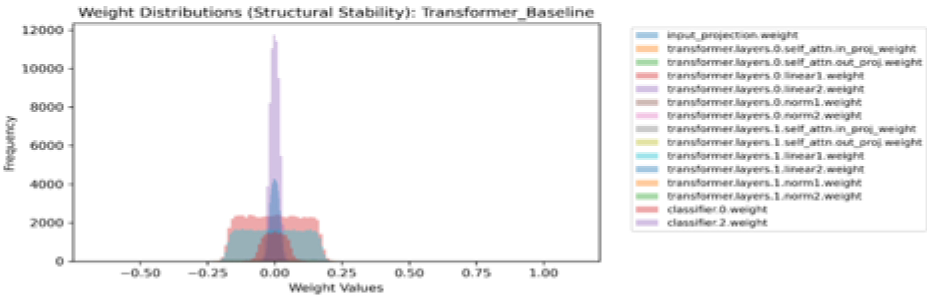
The model weight distributions are also compared with the transformer baseline, which also supports the internal stability of the model.

Proposed hybrid stability: The distribution of weights (Figure 8(a)) indicates that the weights in the network are symmetrically distributed around zero. This equal distribution means that there is a steady convergence point where none of the neurons will have an unfair advantage in determining the diagnosis.

Transformer volatility: Conversely, the weight distribution dynamics of the transformer Figure 8(b) exhibits much greater frequency spikes and a more limited weight distribution, which tends to result in the stochastic oscillations in its convergence path.



(a)



(b)

Figure 8. (a) displays the weight distribution of the proposed BiLSTM-dSVM model. The weights are symmetrically centered around zero, demonstrating consistent parameter optimization and balanced feature learning. The weight distribution of the enhanced Transformer baseline, on the other hand, is shown in (b) and is narrower and more oscillatory, indicating greater parameter variability during optimization.

The BiLSTM-dSVM model provides mathematical stability by the hinge constraint that generates a functional dead zone of loss when the classes are distinguished by a margin of 1.0. It successfully prevents noise and motion artifacts by this mechanism and makes the model ignore small signal variations that are common in the 20,000 samples of wearable data. As illustrated in the latent manifold separation, the representation has geometrically separated clusters of healthy and AF rhythms despite signal degradation. Moreover, the weight distribution dynamics validate a more balanced and stable internal state than the volatile Transformer baseline and much faster convergence of the functional, this architecture attains sensor-invariant representation, which offers a powerful and computationally efficient solution to cardiac monitoring in the real world.

3.11. Computational complexity analysis

The BiLSTM-dSVM model gives mathematical stability of the hinge constraint, which forms an effective dead zone of loss where the classes are separated by a 1.0 margin. It successfully prevents noise and motion artifacts by this mechanism and makes the model ignore small signal variations that are common in the 20,000 samples of wearable data. As shown in the latent manifold separation, the representation has geometrically segregated groups of Healthy and AF rhythms even after signal degradation. Additionally, the weight distribution dynamics justifies a more balanced, stable internal state compared to the unsteady transformer baseline and far quicker convergence of the functional.

This architecture achieves sensor-invariant representation, which provides an efficient and computationally efficient solution to cardiac monitoring in the real world.

$$O(K^2 C_{in} C_{out} HW), \tag{97}$$

K is the size of the convolution kernel, C_in and C out are the number of input and output channels, and H W are the spatial dimensions of the feature map. In the case of the BiLSTM temporal modeling module, the computational complexity is mainly related to the length of the sequence, and the dimension of the hidden state, and could be estimated as:

$$O(TH^2), \tag{98}$$

where T represents the temporal sequence length and H denotes the number of hidden units. The computational cost of the differentiable SVM classifier is relatively small compared with deep neural network layers and can be expressed as:

$$O(Nd), \tag{99}$$

where N is the number of training samples and d represents the dimensionality of the feature vector. Consequently, the overall computational complexity of the proposed architecture can be approximated as:

$$O(K^2 C_{in} C_{out} HW + TH^2 + Nd). \tag{100}$$

In contrast, transformer-based architectures rely on self-attention mechanisms whose computational complexity scales quadratically with respect to sequence length:

$$O(T^2 d).$$

This quadratic dependence greatly makes computing costly when operating long ECG sequences. Transformer-based models thus tend to consume more memory and take longer to train than recurrent ones. The given hybrid model alleviates this shortcoming by integrating convolutional feature extraction with less recurrent temporal modeling and an efficient margin-based classifier. The design enables representation learning to be achieved with reasonable computational complexity, which is acceptable to be implemented in wearable and edge-based ECG monitoring systems.

To improve the reproducibility and transparency of the proposed framework, the experiments were conducted under fixed random initialization settings and standardized training protocols. The implementation was developed using Python with TensorFlow/Keras libraries on a Windows-based workstation equipped with an NVIDIA RTX-series GPU and 32 GB RAM.

3.11.1 Reproducibility statement

The hyperparameter search process included learning rate exploration within the range of 1e−5 to 1e−2, batch sizes of {16, 32, 64}, dropout rates between 0.2 and 0.5, and BiLSTM hidden dimensions ranging from 64 to 256 units. The final configuration was selected based on validation-set Macro F1 and AUC performance.

The network was trained using the Adam optimizer with an initial learning rate of 0.001. A Reduce-on-Plateau scheduling strategy was employed with a patience value of 5 epochs and a decay factor of 0.5. Early stopping was activated when validation loss did not improve for 10 consecutive epochs. The maximum training duration was fixed at 100 epochs.

To ensure experimental consistency, all runs used a fixed random seed value of 42 for dataset partitioning, parameter initialization, and mini-batch generation. The codebase and reproducibility scripts will be publicly released upon acceptance of the manuscript.

The superior performance of the proposed framework can be attributed to several complementary mechanisms operating jointly within the architecture. First, the differentiable SVM optimization head improves geometric class separation through margin-aware learning, thereby enhancing structural risk minimization and reducing decision-boundary ambiguity under noisy ECG acquisition conditions. Unlike conventional Softmax-based classifiers, the squared hinge-loss optimization strategy encourages more discriminative latent representations and improves robustness against overlapping AF and non-AF patterns. Second, the proposed knowledge-enhanced fusion mechanism integrates handcrafted clinical ECG descriptors with deep temporal representations learned by the BiLSTM module. This integration improves feature interpretability and enables the model to preserve clinically meaningful cardiac rhythm characteristics associated with AF pathophysiology, including irregular RR interval variability and waveform instability.

Furthermore, the incorporation of temporal sequence modeling within the BiLSTM framework improves the ability to capture long-range ECG dependencies and dynamic rhythm transitions, thereby enhancing generalization capability across both clinical-grade ECG recordings and noisy wearable sensing environments.

The ablation and cross-dataset evaluation results further confirm that jointly optimized margin-aware learning and knowledge-enhanced latent fusion contribute significantly to convergence stability, robustness against motion artifacts, and improved diagnostic reliability in heterogeneous ECG monitoring scenarios.

4. Conclusions

Conclusively, we have shown that the BiLSTM-dSVM model sets a new standard in robust cardiac diagnosis by integrating the time-varying deep learning with geometric margin-maximization. The dSVM hinge constraint was integrated to achieve a mathematically stable solution that practically reduces the noise and motion artifacts present in large scale wearable data, as demonstrated by the operational dead zone of loss below the 1.0 margin. Moreover, the tight, non-overlapping cluster of the Latent Manifold of clinical and wearable settings is indicative of the potential of the framework to learn sensor-invariant physiological representations. The model can reach a near-perfect AUC of 0.9951 and 99% accuracy on 20,000 samples and has better convergence kinetics and structural stability than high-complexity transformer-based architectures. These results indicate that the suggested hybrid architecture is not merely a high-fidelity diagnostic system but also a computationally efficient solution to the next generation of the scalable, real-time mHealth cardiac surveillance systems.

Author contributions

Raid Abdulhadi Abdulqader Alabdullah: Conceptualization, Methodology, Software, Data curation, Formal analysis, Investigation, Visualization, Writing—original draft preparation; Saratha Sathasivam: Supervision, Methodology, Validation, Writing—review and editing, Project administration; Marwa Mawfaq MohamedSheet AL-Hatab: Data curation, Validation, Investigation, Visualization, Writing—review and editing. All authors have read and approved the final version of the manuscript for publication.

Use of Generative-AI tools declaration

The authors declare that they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

The authors declare no conflict of interest.

References

1. S. Liaqat, K. Dashtipour, A. Zahid, K. Assaleh, K. Arshad, N. Ramzan, Detection of atrial fibrillation using a machine learning approach, *Information*, **11** (2020), 549. https://doi.org/10.3390/info11120549
2. S. S. M. Sheet, M. M. M. Al-Hatab, M. A. Qasim, A novel method for examining promoters using statistical analysis and artificial intelligence learning, *IAES International Journal of Artificial Intelligence*, **14** (2025), 4006–4016. https://doi.org/10.11591/ijai.v14.i5.pp4006-4016
3. G. D. Clifford, C. Y. Liu, B. Moody, L. H. Lehman, I. Silva, Q. Li, et al., AF classification from a short single lead ECG recording: the physionet/computing in cardiology challenge 2017, *Computing in Cardiology*, **44** (2017), 1–4. https://doi.org/10.22489/CinC.2017.065-469
4. J. B. Wang, P. Wang, S. P. Wang, Automated detection of atrial fibrillation in ECG signals based on wavelet packet transform and correlation function of random process, *Biomed. Signal Proces.*, **55** (2020), 101662. https://doi.org/10.1016/j.bspc.2019.101662
5. J. B. Duan, Q. Wang, B. Zhang, C. Liu, C. R. Li, L. Wang, Accurate detection of atrial fibrillation events with R-R intervals from ECG signals, *PLoS ONE*, **17** (2022), e0271596. https://doi.org/10.1371/journal.pone.0271596
6. N. Nuryani, B. Harjito, I. Yahya, A. Lestari, Atrial fibrillation detection using support vector machine, *Proceedings of the Joint International Conference on Electric Vehicular Technology and Industrial, Mechanical, Electrical and Chemical Engineering (ICEVT & IMECE)*, Surakarta, Indonesia, 2015, 215–218. https://doi.org/10.1109/ICEVTIMECE.2015.7496672
7. S. K. S. Shrikanth Rao, R. J. Martis, RR interval-based atrial fibrillation detection using traditional and ensemble machine learning algorithms, *J. Med. Signals Sens.*, **13** (2023), 224–232. https://doi.org/10.4103/jmss.jmss_4_22

8. R. J. Martis, U. R. Acharya, H. Prasad, C. K. Chua, C. M. Lim, Automated detection of atrial fibrillation using Bayesian paradigm, *Knowl.-Based Syst.*, **54** (2013), 269–275. https://doi.org/10.1016/j.knosys.2013.09.016
9. S. Sieciński, P. S. Kostka, E. J. Tkacz, Comparison of atrial fibrillation detection performance using decision trees, SVM and artificial neural network, In: *Information Technology and Systems*, Cham: Springer, 2019, 693–701. https://doi.org/10.1007/978-3-030-11890-7_65
10. M. Limam, F. Precioso, Atrial fibrillation detection and ECG classification based on convolutional recurrent neural network, *Computing in Cardiology*, **44** (2017), 1–4. https://doi.org/10.22489/CinC.2017.171-325
11. B. Grabowski, P. Głomb, W. Masarczyk, P. Pławiak, Ö. Yıldırım, U. R. Acharya, et al., Classification and self-supervised regression of arrhythmic ECG signals using convolutional neural networks, 2022, arXiv:2210.14253. https://doi.org/10.48550/arXiv.2210.14253
12. A. Sherstinsky, Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network, *Physica D*, **404** (2020), 132306. https://doi.org/10.1016/j.physd.2019.132306
13. V. Satheeswaran, G. N. Chandrika, A. Mitra, R. Chowdhury, P. Kumar, E. Glory, Deep learning based classification of ECG signals using RNN and LSTM mechanism, *J. Electronic. Electromed. Eng. Med. Inform.*, **6** (2024), 332–342.
14. Y. T. Hu, T. F. Feng, M. Wang, C. Y. Liu, H. Tang, Detection of paroxysmal atrial fibrillation from dynamic ECG recordings based on a deep learning model, *J. Pers. Med.*, **13** (2023), 820. https://doi.org/10.3390/jpm13050820
15. C. Y. Ma, X. Long, W. J. Sheng, R. Vullings, M. C. Yang, L. N. Zhao, An atrial fibrillation detection strategy in dynamic ECGs with significant individual differences, *IEEE T. Instrum. Meas.*, 73 (2023), 4002010. https://doi.org/10.1109/TIM.2023.3335526
16. M. A. Qasim, M. M. M. Al-Hatab, L. H. Albak, Deep feature-based multi-class Alzheimer's disease classification with statistical performance evaluation, *rning, IAES International Journal of Artificial Intelligence*, **15** (2026), 695–706. https://doi.org/10.11591/ijai.v15.i1.pp695-706
17. B. Aldughayfiq, F. Ashfaq, N. Z. Jhanjhi, M. Humayun, A deep learning approach for atrial fibrillation classification using multi-feature time series data from ECG and PPG, *Diagnostics*, 13 (2023), 2442. https://doi.org/10.3390/diagnostics13142442
18. S. Sathasivam, S. A. Alzaeemi, M. Velavan, Mean-field theory in hopfield neural network for doing 2 satisfiability logic programming, *International Journal of Modern Education and Computer Science*, **12** (2020), 27–39. https://doi.org/10.5815/ijmecs.2020.04.03
19. S. A. Alzaeemi, S. Sathasivam, M. Velavan, Agent-based modeling in doing logic programming in fuzzy hopfield neural network, *International Journal of Modern Education and Computer Science*, **13** (2021), 23–32. https://doi.org/10.5815/ijmecs.2021.02.03
20. A. M. Sagir, S. Sathasivam, A hybridised intelligent technique for the diagnosis of medical diseases, *Pertanika J. Sci. Technol.*, **25** (2017), 425–440.
21. S. A. Alzaeemi, S. Sathasivam, M. K. M. Ali, K. G. Tay, M. Velavan, Hybridized intelligent neural network optimization model for forecasting prices of rubber in Malaysia, *Comput. Syst. Sci. Eng.*, **47** (2023), 1471–1491. https://doi.org/10.32604/csse.2023.037366
22. S. Lee, A. Nedić, Epoch gradient descent for smoothed hinge-loss linear SVMs, *2013 American Control Conference, Washington*, DC, USA, 2013, 4789–4794. https://doi.org/10.1109/ACC.2013.6580579

23. M. S. Corneanu, L. E. Prodaniuc, A. Prodaniuc, I. A. Șova, E. O. Coșovanu, O. R. Petriș, Diagnostic accuracy of wearable ECG devices for atrial fibrillation and ST-segment changes: A systematic review, *Diagnostics*, **15** (2025), 3162. https://doi.org/10.3390/diagnostics15243162
24. W. F. Xie, C. Chen, R. J. Zhao, Y. Lu, Detection of atrial fibrillation based on feature fusion using attention-based BiLSTM, *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Sydney, Australia, 2023, 1–4. https://doi.org/10.1109/EMBC40787.2023.10340023
25. D. Kraft, P. Rumm, Atrial fibrillation and atrial flutter detection using deep learning, *Sensors*, **25** (2025), 4109. https://doi.org/10.3390/s25134109
26. J. J. Si, Y. L. Bao, F. L. Chen, Y. Wang, M. M. Zeng, N. Y. He, et al., Research on atrial fibrillation diagnosis in electrocardiograms based on CLA-AF model, *European Heart Journal - Digital Health*, **6** (2025), 82–95. https://doi.org/10.1093/ehjdh/ztae092
27. S. Y. Wu, A compact LSTM-SVM fusion model for long-duration cardiovascular diseases detection, 2024, arXiv:2312.09442. https://doi.org/10.48550/arXiv.2312.09442
28. F. A. Gers, J. Schmidhuber, F. Cummins, Learning to forget: continual prediction with LSTM, *Neural Comput.*, **12** (2000), 2451–2471. https://doi.org/10.1162/089976600300015015
29. L. Subramanyan, U. Ganesan, A novel deep neural network for detection of atrial fibrillation using ECG signals, *Knowl.-Based Syst.*, **258** (2022), 109926. https://doi.org/10.1016/j.knosys.2022.109926
30. M. Yu, *Atrial fibrillation signal dataset*, Kaggle, 2024. Available from: https://www.kaggle.com/datasets/michaelrowen/atrial-fibrillation-signal-dataset.
31. S. Sathasivam, M. Mamat, M. S. M. Kasihmuddin, M. A. Mansor, Metaheuristics approach for maximum k satisfiability in restricted neural symbolic integration, *Pertanika J. Sci. Technol.*, **28** (2020), 545–564.
32. P. K. Stein, M. S. Bosner, R. E. Kleiger, B. M. Conger, Heart rate variability: A measure of cardiac autonomic tone, *Am. Heart J.*, **127** (1994), 1376–1381. https://doi.org/10.1016/0002-8703(94)90059-0
33. J. Rahul, M. Sora, L. D. Sharma, V. K. Bohat, An improved cardiac arrhythmia classification using an RR interval-based approach, *Biocybern. Biomed. Eng.*, **41** (2021), 656–666. https://doi.org/10.1016/j.bbe.2021.04.004
34. J. B. Duan, Q. Wang, B. Zhang, C. Liu, C. R. Li, L. Wang, Accurate detection of atrial fibrillation events with RR intervals from ECG signals, *PLoS ONE*, **17** (2022), e0271596. https://doi.org/10.1371/journal.pone.0271596
35. S. K. S. Rao, R. J. Martis, RR interval-based atrial fibrillation detection using traditional and ensemble machine learning algorithms, *J. Med. Signals Sens.*, **13** (2023), 224–232. https://doi.org/10.4103/jmss.jmss_4_22
36. D. Kraft, P. Rumm, Atrial fibrillation and atrial flutter detection using deep learning, Sensors, **25** (2025), 4109. https://doi.org/10.3390/s25134109
37. C. P. Lee, C. J. Lin, A study on L2-loss (squared hinge-loss) multiclass SVM, *Neural Comput.*, **25** (2013), 1302–1323. https://doi.org/10.1162/NECO_a_00434
38. Z. L. Yang, J. Yin, Multi-view fusion and view-specific-labels learning for multi-label classification, *Neural Process. Lett.*, **57** (2025), 89. https://doi.org/10.1007/s11063-025-11785-3

39. B. Chen, W. Guo, B. Li, R. K. F. Teng, M. J. Dai, J. P. Luo, et al., A study of deep feature fusion based methods for classifying multi-lead ECG, 2018, arXiv:1808.01721. https://doi.org/10.48550/arXiv.1808.01721
40. M. Brennan, M. Palaniswami, P. Kamen, Do existing measures of Poincaré plot geometry reflect nonlinear features of heart rate variability, *IEEE T. Biomed. Eng.*, **48** (2001), 1342–1347. https://doi.org/10.1109/10.959330
41. S. Carrasco, M. J. Gaitán, R. González, O. Yánez, Correlation among Poincaré plot indexes and time and frequency domain measures of heart rate variability, *Journal of Medical Engineering & Technology*, **25** (2001), 240–248. https://doi.org/10.1080/03091900110086651
42. C. Y. Song, Z. L. Zhou, Y. Yu, M. M. Shi, J. X. Zhang, An improved Bi-LSTM method based on heterogeneous features fusion and attention mechanism for ECG recognition, *Comput. Biol. Med.*, **169** (2024), 107903. https://doi.org/10.1016/j.compbiomed.2023.107903
43. I. Patel, A. Sandhya, V. S. Raja, S. Saravanan, Extraction of features from ECG signal, *International Journal of Current Research and Review*, **13** (2021), 103–109. https://doi.org/10.31782/IJCRR.2021.13806

AIMS Press

© 2026 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0)