



Article

Improving Early Prediction of Sudden Cardiac Death Risk via Hierarchical Feature Fusion

Xin Huang ¹, Guangle Jia ^{2,*}, Mengmeng Huang ¹, Xiaoyu He ¹, Yang Li ¹ and Mingfeng Jiang ^{1,*}

¹ School of Computer Science and Technology (School of Artificial Intelligence), Zhejiang Sci-Tech University, Hangzhou 310018, China; xhuang@zstu.edu.cn (X.H.)

² School of Information Science and Engineering, Harbin Institute of Technology, Weihai 264209, China

* Correspondence: jiaguangle2017@163.com (G.J.); m.jiang@zstu.edu.cn (M.J.)

Abstract

Sudden cardiac death (SCD) is a leading cause of mortality worldwide, with arrhythmia serving as a major precursor. Early and accurate prediction of SCD using non-invasive electrocardiogram (ECG) signals remains a critical clinical challenge, particularly due to the inherent asymmetric and non-stationary characteristics of ECG signals, which complicate feature extraction and model generalization. In this study, we propose a novel SCD prediction framework based on hierarchical feature fusion, designed to capture both non-stationary and asymmetrical patterns in ECG data across six distinct time intervals preceding the onset of ventricular fibrillation (VF). First, linear features are extracted from ECG signals using waveform detection methods; nonlinear features are derived from RR interval sequences via second-order detrended fluctuation analysis (DFA2); and multi-scale deep learning features are captured using a Temporal Convolutional Network-based sequence-to-vector (TCN-Seq2vec) model. These multi-scale deep learning features, along with linear and nonlinear features, are then hierarchically fused. Finally, two fully connected layers are employed as a classifier to estimate the probability of SCD occurrence. The proposed method is evaluated under an inter-patient paradigm using the Sudden Cardiac Death Holter (SCDH) Database and the Normal Sinus Rhythm (NSR) Database. This method achieves average prediction accuracies of 97.48% and 98.8% for the 60 and 30 min periods preceding SCD, respectively. The findings suggest that integrating traditional and deep learning features effectively enhances the discriminability of abnormal samples, thereby improving SCD prediction accuracy. Ablation studies confirm that multi-feature fusion significantly improves performance compared to single-modality models, and validation on the Creighton University Ventricular Tachyarrhythmia Database (CUDB) demonstrates strong generalization capability. This approach offers a reliable, long-horizon early warning tool for clinical SCD risk assessment.

Keywords: Sudden Cardiac Death (SCD); asymmetric; hierarchical feature fusion; electrocardiogram (ECG); Temporal convolutional network



Received: 6 September 2025

Revised: 5 October 2025

Accepted: 11 October 2025

Published: 15 October 2025

Citation: Huang, X.; Jia, G.; Huang, M.; He, X.; Li, Y.; Jiang, M. Improving Early Prediction of Sudden Cardiac Death Risk via Hierarchical Feature Fusion. *Symmetry* **2025**, *17*, 1738.

<https://doi.org/10.3390/sym17101738>

Copyright: © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Sudden cardiac death (SCD) constitutes a paramount global public health challenge, characterized by its unexpected nature and devastatingly high fatality rate [1]. The escalating prevalence of cardiovascular risk factors, coupled with demographic aging trends, underscores the urgent need for innovative strategies in preventive cardiology [2].

The electrocardiogram (ECG) serves as the fundamental tool for non-invasive cardiac electrical activity assessment. A critical challenge in ECG analysis stems from its inherent asymmetric nature, which in this context refers to the complex, nonlinear, and non-stationary characteristics of the signal. These statistical asymmetries manifest as dynamic, time-varying patterns—such as the differing dynamics between heart rate acceleration and deceleration, or subtle morphological changes under stress—that are not captured by linear or stationary models. The advancement of portable wearable single-lead ECG devices has revolutionized this field, enabling cost-effective, continuous, and real-time monitoring outside traditional clinical settings [3,4]. While standard 12-lead systems provide comprehensive spatial cardiac information, the practical advantages of single-lead devices—including unparalleled convenience for long-term use—make them ideal for widespread screening and ambulatory monitoring. Critically, essential biomarkers for SCD risk stratification, such as heart rate variability (HRV), can be effectively captured from high-quality single-lead signals [5].

The evolution of SCD prediction methodologies has traversed several distinct phases. Initial research was predominantly anchored in manual feature engineering, particularly from HRV time series, combined with conventional machine learning classifiers [6]. The subsequent paradigm integrated advanced signal processing techniques—such as wavelet transforms and empirical mode decomposition—to enhance feature quality and marginally extend the prediction horizon [7]. The most recent breakthrough has been ushered in by deep learning, which demonstrates a remarkable capacity to automatically learn discriminative patterns directly from raw or minimally processed physiological signals [8,9]. For instance, deep neural networks have now achieved cardiologist-level performance in arrhythmia detection from ambulatory ECGs [10].

Despite these progressive advancements, two pivotal challenges persistently impede clinical translation. First, the predominant focus of existing methods remains on short-term prediction windows, typically analyzing signals mere minutes before the onset of ventricular fibrillation. However, pathophysiological processes leading to SCD often initiate much earlier. This critical limitation severely restricts the time available for life-saving clinical interventions [11]. Second, current methodologies often operate in methodological silos: they either depend exclusively on handcrafted features grounded in physiological knowledge or rely purely on end-to-end deep learning representations. This dichotomy fails to fully exploit the complementary strengths of both approaches in characterizing the signal's asymmetric nature, ultimately constraining model generalization and reliability for long-horizon prediction [12].

To address these identified gaps, this paper proposes a novel hierarchical feature fusion framework for the early prediction of SCD risk. The principal contributions of this work are threefold:

- (1) First, we design a multi-level feature architecture that systematically and synergistically integrates complementary information from linear, nonlinear, and deep learning-based representations, thereby providing a multi-faceted characterization of the ECG signal's dynamics.
- (2) Second, we introduce a dedicated hierarchical fusion module to effectively combine multi-scale temporal contexts extracted by a Temporal Convolutional Network (TCN-Seq2vec)—which is particularly adept at learning from non-stationary sequences—with handcrafted clinical features, enabling a comprehensive modeling of asymmetric patterns across different temporal scales.
- (3) Third, we rigorously evaluate our model under a clinically relevant inter-patient paradigm, demonstrating significantly improved prediction accuracy throughout an

extended 60 min period preceding SCD onset, thereby offering a reliable tool for long-horizon risk stratification.

The remainder of this paper is structured as follows: Section 2 provides a structured review of related work. Section 3 elaborates on the proposed methodology. Section 4 presents the experimental results and a comparative analysis. Section 5 concludes this paper and Section 6 discusses potential future research directions.

2. Related Work

Traditional Feature Extraction and Machine Learning Approaches. Early SCD prediction research primarily focused on combining HRV feature extraction with traditional classifiers. Khazaei et al. [13] employed nonlinear methods such as incremental entropy for HRV feature extraction and achieved 95% accuracy 6 min before SCD using classifiers like Decision Trees and SVM. Ebrahimzadeh et al. [14] proposed a time local subset feature selection (TLSFS) method, extending the prediction window to 13 min pre-SCD with 84.28% accuracy. Shi et al. [15] used Ensemble Empirical Mode Decomposition (EEMD) for HRV decomposition and reached 94.7% accuracy 14 min pre-SCD. A common limitation of these early approaches was their reliance on manual feature engineering and specific classifiers, which often led to high computational costs and limited generalization capability.

Advanced Signal Processing and Feature Reduction Methods. Subsequent research integrated advanced signal processing techniques with statistical features and dimensionality reduction to improve performance. Shi et al. [16] combined Discrete Wavelet Transform (DWT) and Locality Preserving Projections (LPP) for feature processing, achieving 97.6% accuracy 14 min pre-SCD with only 5 features. Reddy et al. [17] employed Hilbert-Huang Transform (HHT) and DWT, extending prediction to 30 min pre-SCD. Centeno-Bautista et al. [18] used Complete Ensemble Empirical Mode Decomposition and achieved 97.28% accuracy 30 min pre-SCD. While these methods demonstrated improved performance, they faced challenges in handling the nonlinear, non-stationary, and asymmetrical nature of physiological signals. Statistical dimensionality reduction techniques often failed to preserve temporal dependencies and were inadequate for capturing dynamic, nonlinear relationships within the data.

Deep Learning and Multi-modal Fusion Approaches. Recent years have witnessed increased focus on deep learning and multi-modal feature fusion. Yang et al. [19] proposed multi-domain feature fusion, reaching 91.22% accuracy and predicting SCD 70 min in advance—a significant extension of the prediction horizon. Telangore et al. [20] developed a multi-modal deep learning model with 98.81% accuracy 30 min pre-SCD, though the model faced overfitting issues. Gao et al. [21] designed a specialized algorithm for low-signal-to-noise single-lead ECG, achieving 95.43% accuracy and proposing a Sudden Cardiac Death Index (SCDI) for risk quantification.

Research Gaps and Our Positioning. Despite these advances, current methods exhibit several limitations that our work aims to address. First, most deep learning approaches fail to effectively leverage the complementary information provided by handcrafted features and automatically learned representations. Second, while some studies have extended prediction windows, maintaining high accuracy over longer horizons remains challenging. Third, many models are developed under intra-patient paradigms, limiting their clinical applicability.

Unlike previous work, our approach systematically integrates multi-scale features through a dedicated hierarchical fusion module and is evaluated under a clinically relevant inter-patient paradigm. By combining the strengths of traditional feature engineering and deep learning, we achieve robust performance across extended prediction windows while addressing the generalization challenges of existing methods.

3. Materials and Methods

The proposed framework (Figure 1) involves four sequential stages: signal preprocessing, feature extraction, feature fusion, and classification. Prior to feature extraction, raw ECG signals are denoised using discrete wavelet transform (DWT) to mitigate noise interference. Feature extraction is divided into three branches: (1) Linear feature extraction: Five linear features, including RR intervals, QRS complexes, and T waves, are extracted from the denoised ECG signals. (2) Nonlinear feature extraction: The scaling exponent α_1 is derived as a nonlinear feature from RR interval sequences of each 1 min ECG sample using second-order detrended fluctuation analysis (DFA-2). (3) Deep learning-based feature extraction: Multiscale deep representations are extracted from each 1 min ECG sample using a TCN based sequence to vector model (TCN-Seq2vec). These heterogeneous features are then hierarchically fused into a unified representation to enhance discriminative power. The classification stage employs a fully connected layer to map the fused features to probabilistic predictions across target categories.

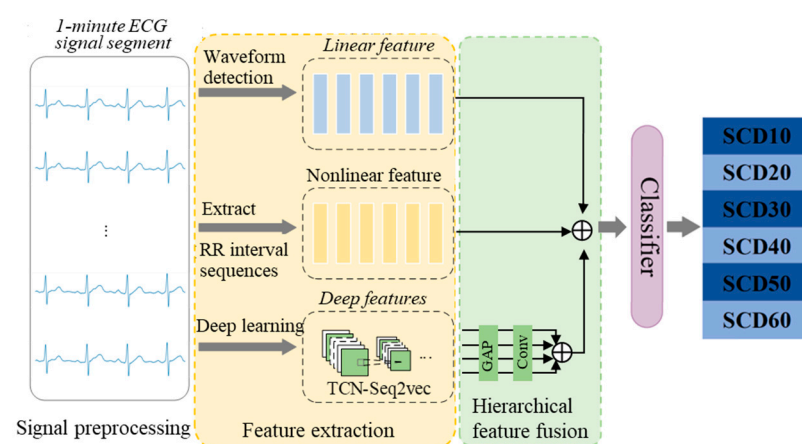


Figure 1. Schematic diagram of the proposed Sudden Cardiac Death (SCD) prediction approach.

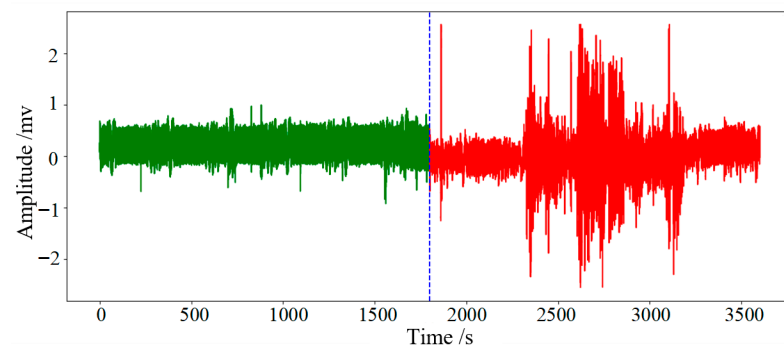
3.1. Dataset Description

This study utilizes two publicly available datasets: the Sudden Cardiac Death Holter Database (SCDH) [22], jointly established by MIT-BIH and the American Heart Association (AHA), and the Normal Sinus Rhythm (NSR) database [23] provided by MIT-BIH.

Table 1 provides a detailed display of the statistical information recorded in the database of the dataset used in this study. The SCDH database contains long-term ECG recordings from 23 participants, with durations ranging from several hours to 24 h per recording. These recordings capture the full progression from normal sinus rhythm to ventricular fibrillation (VF) onset, sampled at 250 Hz. As the SCDH data were collected from different hospital devices with inconsistent lead placements, this study utilizes the first lead of the provided dual-lead signals. Among these records, 20 subjects ultimately experienced cardiac arrest. Figure 2 displays representative ECG waveforms from the SCDH dataset. The green segment (left) shows the ECG waveform prior to VF onset, while the red segment (right) depicts the signal after VF onset. As is evident from the figure, the waveforms prior to the onset of VF exhibit a regular cardiac rhythm. Following the onset, the electrical activity rapidly devolves into a chaotic and irregular pattern with no discernible normal ECG components.

Table 1. Details of the data used in this work.

Database	Diagnosis	Sampling Rate	Subjects	
			Number of Subjects	Subjects Features
NSR	Normal	128 HZ	18	13 females (age 20–50) 5 males (age 26–45)
SCD	SCD	250 HZ	23	8 females, 13 males, 2 sex unknown (age 18–89)

**Figure 2.** ECG waveforms from the SCDH dataset. The green and red segments represent the signal prior to and after ventricular fibrillation (VF) onset, respectively.

The NSR database, serving as the control group, comprises ECG recordings from 18 healthy subjects sampled at 128 Hz. To ensure comparability and minimize analytical bias, the NSR data were resampled to 250 Hz using a Fast Fourier Transform (FFT)-based spectral resampling method, followed by normalization of both datasets. This was achieved by calculating the target length for each signal segment via the ratio of the sampling rates ($250/128 \approx 1.953125$). Specifically, for a standard 10 min ECG window, the original 76,800 data points ($10 \text{ min} \times 60 \text{ s} \times 128 \text{ Hz}$) were resampled to a new length of 150,000 points ($10 \text{ min} \times 60 \text{ s} \times 250 \text{ Hz}$). The method applies a Fast Fourier Transform (FFT) to the original signal, performs optimal interpolation in the frequency domain by zero-padding the spectrum to the target length, and then reconstructs the signal via the inverse FFT. This process ensures the precise conversion of the sampling rate while perfectly preserving the original temporal duration of the signal windows, which is critical for subsequent time-aligned feature extraction and comparative analysis.

For sudden cardiac death (SCD) prediction, only ECG data surrounding the VF onset are clinically relevant. Therefore, this study focuses on the 60 min ECG segments preceding VF onset. Table 2 details the 18 selected ECG recordings used in this study, including the precise timing of VF onset and the starting points of the extracted signal segments.

Table 2. Detailed record information in the SCDH dataset.

Record Name	Duration	The Starting Time of the Selected Segment	VF Onset Time
30	24:33:17	06:54:33	07:54:33
31	13:58:40	12:42:24	13:42:24
32	24:20:00	15:45:18	16:45:18
33	24:33:00	03:46:19	04:46:19
34	07:05:20	05:35:44	06:35:44
36	20:21:20	17:59:01	18:59:01
37	25:08:00	00:31:13	01:31:13
39	05:47:00	03:37:51	04:37:51

Table 2. Cont.

Record Name	Duration	The Starting Time of the Selected Segment	VF Onset Time
41	03:56:00	01:59:24	02:59:24
43	23:07:50	14:37:11	15:37:11
44	23:20:00	18:38:45	19:38:45
45	24:09:20	17:09:17	18:09:17
46	04:15:10	02:41:47	03:41:47
47	23:35:50	05:13:01	06:13:01
48	24:36:15	01:29:40	02:29:40
50	23:07:38	10:45:43	11:45:43
51	25:08:30	21:58:23	22:58:23
52	07:31:05	01:32:40	02:32:40

3.2. Signal Preprocessing

The ECG signal preprocessing phase involves three sequential steps: segmentation, denoising, and waveform detection. First, raw ECG signals are partitioned into discrete 1 min intervals. Subsequently, discrete wavelet transform (DWT) is applied to eliminate artifacts, including baseline wander and high-frequency noise. Finally, fiducial point detection identifies five key ECG waveform landmarks: Q-wave peak, R-wave peak, S-wave peak, T-wave peak, and T-wave end, enabling subsequent feature extraction.

To systematically investigate the temporal progression toward ventricular fibrillation (VF), we adopted a stratified segmentation approach. For SCD signals, the 60 min pre-VF period was divided into six consecutive 10 min intervals (0–10, 10–20, . . . , 50–60 min) as illustrated in Figure 3, with each interval subsequently segmented into 1 min samples. For NSR (normal sinus rhythm) signals, a 10 min ECG segment was randomly selected from each recording and partitioned into consecutive 1 min epochs. This design ensures identical NSR sample sizes across comparative datasets, thereby isolating model performance variations to temporal electrophysiological changes in pre-VF SCD signals. The standardized approach guarantees objective comparison of deterioration patterns preceding ventricular fibrillation [24].

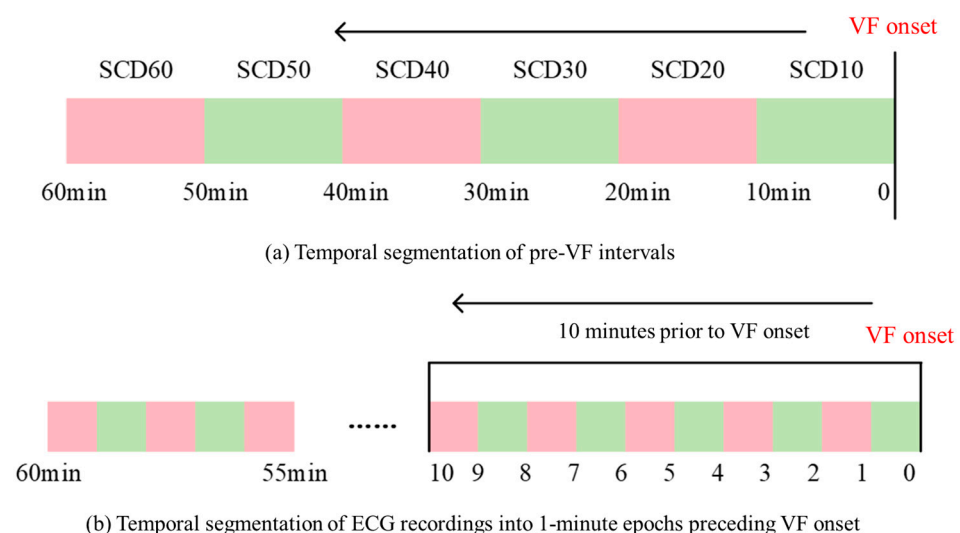


Figure 3. Our signal segmentation methodology.

The ECG signals from the SCDH dataset exhibit significant baseline wander, with some recordings additionally contaminated by high-frequency noise interference. In contrast, NSR dataset recordings demonstrate substantially lower noise levels. Given these

characteristics, we employ DWT [25] for signal denoising due to its superior performance in analyzing local signal features and processing non-stationary biosignals. The denoising procedure consisted of three key steps: (1) an 8-level decomposition is performed using Daubechies8 (db8) wavelet basis functions, selected for their morphological similarity to characteristic ECG waveforms. (2) the approximation coefficients at level 8 and detail coefficients at level 1 are zeroed to eliminate baseline drift and high-frequency noise, respectively. (3) the ECG signal was reconstructed using the modified wavelet coefficients. As demonstrated in Figure 4, this approach effectively removed noise interference while preserving the clinically relevant morphological features of the original ECG signal.

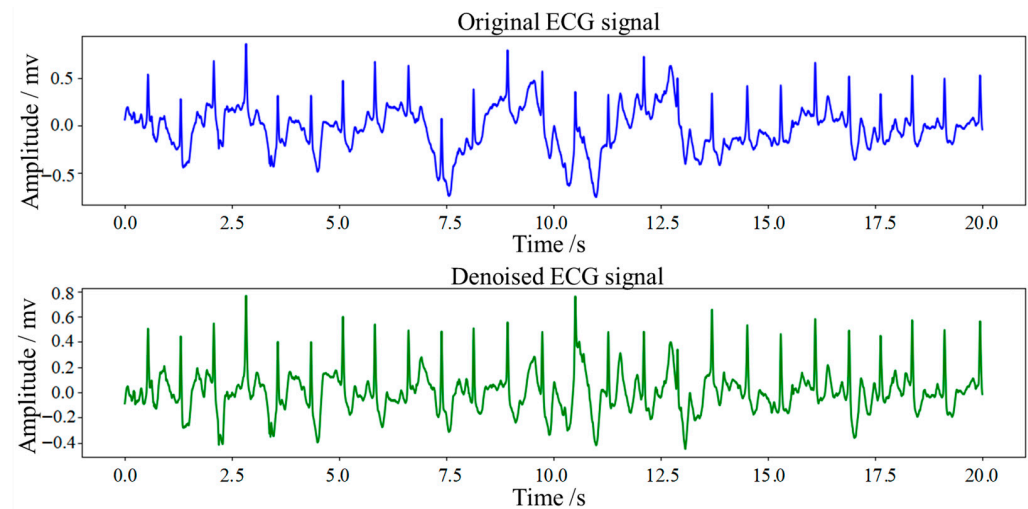


Figure 4. Original and denoised ECG signals.

The R-wave peak positions were identified using the Pan-Tompkins algorithm [26], with immediate calculation of the preceding RR interval upon each R-wave detection. Here, the RR interval is defined as the temporal distance between the peak of the R wave in the current cardiac beat and the corresponding peak in an adjacent beat, reflecting the periodicity of ECG signals. Centered on each identified R-wave peak, we establish a search window spanning ± 0.25 RR intervals to localize the Q and S points, which are determined as the first local minima occurring before and after the R-wave peak within this window, respectively. Following QRS complex detection, we implement the area-based T-wave detection algorithm proposed by Zhang et al. [27] for precise T-wave peak localization. The algorithm operates through the following procedure:

- (1) The search intervals k_a and k_b are delineated within the RR interval according to Equations (1) and (2) to ensure complete inclusion of the T-wave end while preventing overlap with adjacent waveforms.

$$k_a = \begin{cases} R_i + [0.15RR_i] + 37 & \text{if } RR_i < 220 \\ R_i + 70 & \text{if } RR_i \geq 220 \end{cases} \quad (1)$$

$$k_b = \begin{cases} R_i + [0.7RR_i] - 9 & \text{if } RR_i < 220 \\ R_i + [0.2RR_i] + 101 & \text{if } RR_i \geq 220 \end{cases} \quad (2)$$

where R_i is the i -th R-wave peak location in the ECG signal.

- (2) For each temporal point k , the area integral within the sliding window was computed using Equation (3), with the window width fixed at 9 sample points.

$$A_k = \sum_{j=k-w+1}^k (s_j - \bar{s}_k) \quad (3)$$

where s_j represents the j -th sampling point of the ECG signal, and \bar{s}_k represents the signal mean within the smooth window centered on k .

- (3) The T-wave end was identified as the sample point k within search intervals $[k_a, k_b]$ where the amplitude A_k reached its extremum (either maximum or minimum depending on T-wave polarity). Subsequently, the T-wave peak was localized by detecting the first local minimum while retrograde searching from the identified endpoint.

The fiducial point detection effect of ECG signal is shown in Figure 5. It can be seen from the figure that the detection effect is significant, and each key waveform point is clearly marked, indicating that the above method can effectively identify important features.

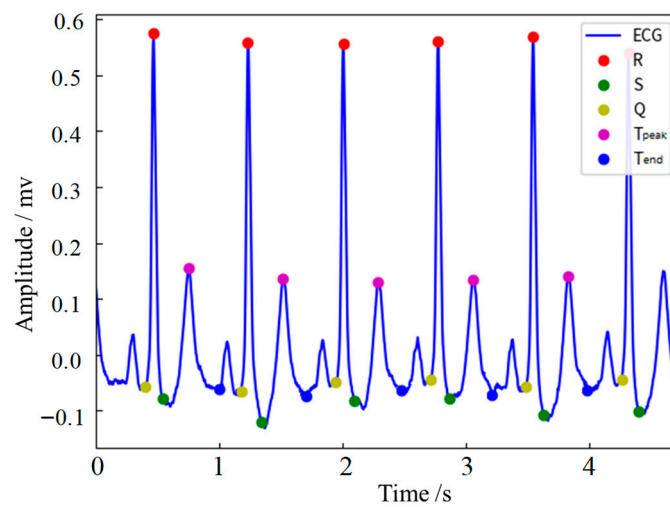


Figure 5. Fiducial point localization in ECG recordings.

3.3. Feature Extraction

3.3.1. Linear Feature Extraction

In studies of SCD risk prediction, linear features derived from ECG signals hold substantial clinical value due to their ease of measurement and immediate applicability in clinical settings. Commonly used linear features include RR intervals, QRS complexes, and T-wave parameters. For the i -th cardiac cycle ($i = 1, 2, 3, \dots, N$), the characteristic fiducial points are defined as follows: $R[i]$ denotes the R-wave peak position, $Q[i]$ represents the Q-wave peak, $S[i]$ indicates the S-wave peak, $T_{peak}[i]$ corresponds to the T-wave peak, and $T_{peak}[i]$ marks the T-wave end point.

- (1) The preceding RR interval is mathematically defined as the time interval between the current R-wave peak and the previous R-wave peak, and the calculation formula is shown in Equation (4):

$$RR[i] = \frac{R[i] - R[i-1]}{f_s} \quad (4)$$

Here, f_s is the sampling rate.

- (2) The QRS duration is defined as the time interval between the Q-wave peak and the S-wave peak, and the calculation formula is shown in Equation (5):

$$QRS[i] = \frac{S[i] - Q[i]}{f_s} \quad (5)$$

- (3) The QT interval is defined as the time from the beginning of the QRS wave to the end of the T wave, and the QTc interval refers to the QT interval corrected based on heart rate [28]. The calculation formula is shown in Equation (6):

$$QTc[i] = \frac{T_{\text{end}}[i] - Q[i]}{f_s} \cdot \sqrt{RR[i]} \quad (6)$$

- (4) The Tp-Te interval is formally defined as the temporal duration between the T-wave peak and T-wave end within the same cardiac cycle. The calculation formula is shown in Equation (7):

$$Tp - Te[i] = \frac{T_{\text{end}}[i] - T_{\text{peak}}[i]}{f_s} \quad (7)$$

- (5) The amplitude of T-wave is defined as the voltage difference between the T-wave peak and T-wave end, and the calculation formula is shown in Equation (8):

$$T_{\text{amp}}[i] = |s(T_{\text{peak}}[i]) - s(T_{\text{end}}[i])| \quad (8)$$

Variability in cardiac cycle counts across 1 min sample segments leads to inconsistent feature quantities (RR intervals, QRS durations, QTc intervals, etc.) extracted from different samples, which would cause compatibility errors in direct model training. To standardize dimensionality, we compute the mean value for each feature within every sample segment.

3.3.2. Nonlinear Feature Extraction

Detrended fluctuation analysis (DFA) [29] is a computational method for quantifying self-similarity and long-range dependence in time series data. This technique characterizes fluctuation properties by systematically removing trend components, demonstrating particular effectiveness in analyzing signals with long-term correlations. The second-order detrended fluctuation analysis (DFA2) [30] represents an enhanced version that employs quadratic polynomial fitting to eliminate nonlinear trends. This advanced method provides more precise analysis of datasets exhibiting complex fluctuation patterns. In the DFA2 method, scaling exponents α_1 and α_2 are commonly extracted to quantify long-range correlation properties of time series across different scales. Specifically, α_1 characterizes the fluctuation behavior and self-similarity of the time series at smaller scales, making it suitable for dynamic analysis over shorter time intervals, whereas α_2 describes the fluctuation behavior and self-similarity at larger scales, thereby revealing long-term dependencies. The extraction procedure for the DFA2 α_1 feature is as follows:

- (1) RR interval extraction and preprocessing: The RR interval sequence is derived from each 1 min ECG signal segment, followed by outlier detection. Any identified outliers are replaced with the average of their immediate neighboring values.
- (2) Cumulative sum transformation: The RR interval sequence $[r_1, r_2, \dots, r_n]$ is converted into an integrated series $Y(k)$ by computing the cumulative sum of deviations from the expected value, as expressed in Equation (9):

$$Y(k) = \sum_{i=1}^k (r_i - \bar{r}), k = 1, 2, \dots, n \quad (9)$$

Here \bar{r} is the mean of the RR interval.

- (3) The integrated series $Y(k)$ is divided into windows of length s (where s ranges between 4 and 16 RR intervals). The scale range of 4 to 16 RR intervals was selected for the DFA2 analysis based on a combination of physiological and practical considerations.

Physiologically, this range (corresponding to approximately 4–16 s for a typical heart rate) captures the short-term correlation properties of heart rate dynamics, which are influenced by autonomic nervous system regulation and have been shown to be altered in pathological states leading to SCD. From a methodological standpoint, this range is well suited for the analysis of 1 min ECG segments, as the upper limit of 16 ensures a sufficient number of data windows for a robust calculation of the scaling exponent, while the lower limit of 4 excludes very short-scale noise.

In conventional DFA2, non-overlapping windows are typically used, with each window independently fitted by a second-order polynomial to remove local trends. However, for shorter data segments, the limited number of non-overlapping windows may lead to increased statistical uncertainty. To address this, a maximum-overlap windowing strategy is adopted in this study to enhance the number of analyzable subintervals. Specifically, for a given window length s , each new window starts one sample point after the previous window's starting position, resulting in an overlap length of $s - 1$ between adjacent windows.

- (4) Within each subinterval, the local trend $y(i)$ is estimated via second-order polynomial fitting. The detrended fluctuation $F(v, s)$ is then computed as the root-mean-square deviation of the integrated series from the fitted trend, as given by Equation (10):

$$F(v, s) = \sqrt{\frac{1}{s} \sum_{i=1}^s [Y(v + i - 1) - y(i)]^2} \quad (10)$$

where v denotes the starting index of the window. The fluctuation function $F(s)$ is obtained by averaging $F(v, s)$ across all windows. Finally, the scaling exponent α_1 is derived through log-log regression analysis.

Figure 6a,b present the visualization results of scale exponent α_1 derived from RR interval sequences using DFA2 and the box plots of DFA2 α_1 features for different sample categories across six datasets, respectively.

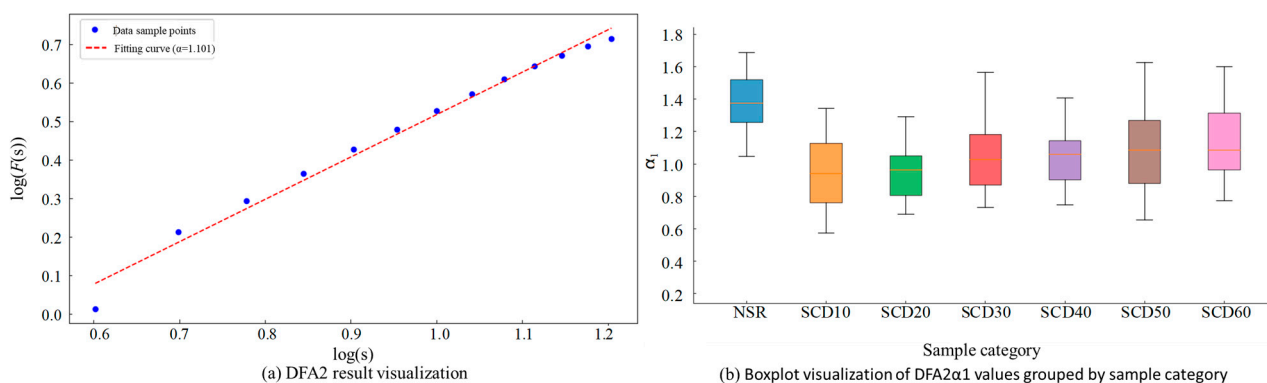


Figure 6. Nonlinear feature extraction.

As evidenced by the figure, the α_1 values of NSR samples are distributed in the range of 1.35 ± 0.2 , while those of SCD samples fall within 0.95 ± 1.5 , demonstrating a significant difference in DFA2 α_1 values between normal ECG signals and SCD signals, where lower values indicate higher SCD risk.

The successful discrimination achieved by the DFA2 α_1 feature underscores its utility in quantifying the nonlinear, long-range correlations within the heart rate. These complex correlation patterns are a direct manifestation of the dynamic asymmetries in the autonomic nervous system's regulation of the heart. Specifically, a decrease in α_1 towards 0.5 (uncorrelated noise) suggests a breakdown in the healthy, fractal-like temporal structure of the heartbeat—a key aspect of the pathological asymmetry that our model aims to capture.

3.3.3. Deep Learning-Based Feature Extraction

We employ a Temporal Convolutional Network-based sequence-to-vector model (TCN-Seq2vec) for deep feature extraction, whose architecture is illustrated in Figure 7a. The model is designed to hierarchically capture multi-scale temporal patterns from ECG signals and condense the entire sequence into an informative fixed-length vector. This is achieved through the synergistic operation of multiple Temporal Convolutional Blocks (TCNBlocks) for feature extraction and a Temporal Attention Mechanism for dynamic feature aggregation.

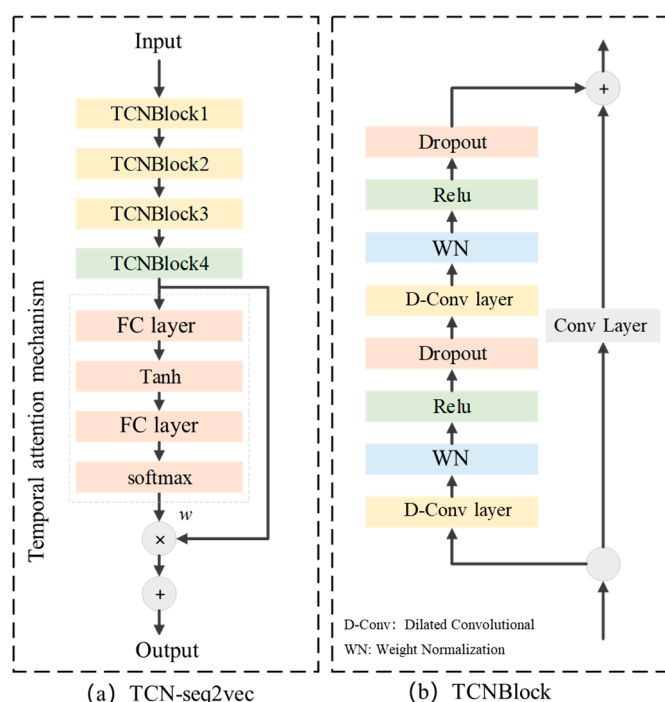


Figure 7. Architecture of the TCN-Seq2vec Model.

The specific structure and parameters of the model are as follows:

(1) Temporal Convolutional Block (TCNBlock). As the fundamental building unit, each TCNBlock is designed to capture temporal dependencies at a specific scale. The use of dilated convolutions is crucial for exponentially expanding the receptive field without losing resolution or increasing computational cost excessively. This allows the network to integrate information from both immediate and distant time points in the ECG signal, which is essential for capturing both transient arrhythmic events and longer-term heart rate trends. Specifically, each block implements:

Dilated Convolutions: Using a fixed kernel size of 3 with dilation rates following a geometric progression (1, 2, 4) across consecutive blocks, enabling exponential receptive field expansion from approximately 5 to 85 time steps while maintaining computational efficiency.

Normalization and Regularization: Each convolutional layer is followed by weight normalization, ReLU activation, and dropout regularization (rate = 0.5) to ensure training stability and prevent overfitting.

Residual Connections: Identity or 1×1 projection shortcuts are incorporated to mitigate gradient vanishing, formulated as:

$$Output = ReLU(Residual(x) + Block(x))$$

where $Block(x)$ represents the sequential processing through the two normalized convolutional layers.

(2) Temporal Attention Mechanism. This module dynamically identifies and emphasizes clinically salient segments (e.g., specific arrhythmic beats or ischemic morphologies) by computing adaptive weights for different time steps, thereby producing a context-aware vector representation where critical periods contribute more significantly. Its implementation involves three dedicated components:

Feature Reduction. The high-dimensional feature maps from the final TCNBlock (128 channels) are first processed by a dedicated TCNBlock configured with a kernel size of 3, a dilation rate of 1, and 64 output channels. This step reduces the feature dimensionality, serving as an information bottleneck that focuses the subsequent computation on the most salient patterns. The resulting tensor is then permuted from the shape (batch, channels, timesteps) to (batch, timesteps, channels) to align the temporal dimension for attention scoring.

Attention Weight Computation. The permuted features are fed into a two-layer fully connected network to compute the importance of each time step. The first linear layer ($64 \rightarrow 32$ dimensions) employs a Tanh activation function to capture non-linear interactions, while the second linear layer ($32 \rightarrow 1$ dimension, without bias) produces a scalar importance score for each time step. These raw scores are then normalized across the temporal dimension via a Softmax function to generate the final attention weights α_i , which sum to one.

Context-Aware Aggregation. The final output is computed as a weighted sum of the original (reduced) feature sequence, formally expressed as $z = \sum_{i=1}^T \alpha_i h_i$, where h_i is the feature vector at the i -th time step. This operation aggregates the entire variable-length sequence into a single, fixed-dimensional vector z , whose representation is dominated by the features from time steps deemed most critical for SCD risk, thereby enhancing the model's discriminative power and potential clinical interpretability.

(3) TCN-Seq2vec Model. As shown in Figure 7a, the model first processes the input sequence through multiple TCNBlocks with increasing dilation rates (1, 2, 4) and channel numbers (32, 64, 128). This design creates a hierarchical feature pyramid, where lower layers capture fine-grained, short-term patterns (e.g., spike shapes) and higher layers capture coarse-grained, long-term trends (e.g., heart rate instability). The outputs from these blocks are then fused. Finally, the Temporal Attention Mechanism is applied to this multi-scale representation, transforming the variable-length sequence into a discriminative fixed-length vector for the final classification, effectively summarizing the entire 1 min ECG episode.

Crucially, the combination of multi-scale temporal convolution and dynamic attention empowers the TCN-Seq2vec model to effectively handle the asymmetric, non-stationary characteristics of ECG signals. It does not assume signal stability but instead learns to identify and weigh the importance of critical, often short-lived, pathological episodes within the longer context, which is essential for early SCD prediction.

3.4. Hierarchical Feature Fusion and Classification

A central hypothesis of this work is that the effective integration of complementary information from heterogeneous features—ranging from clinically interpretable handcrafted features to high-dimensional deep representations—is pivotal for robust and accurate SCD prediction. This is because the pathological processes leading to SCD manifest as complex asymmetries across multiple aspects of the ECG signal. While existing methods often rely on a single type of feature or employ simple concatenation, they fail to capture the intrinsic, hierarchical nature of cardiac electrophysiological patterns that manifest across different temporal scales. For instance, handcrafted nonlinear features (like DFA²) explicitly quantify one form of dynamic asymmetry in heart rate, while deep learning features implicitly learn another from raw waveform morphology. A simple concatenation cannot fully model the

interactions between these different characterizations of asymmetry. To bridge this gap, we propose a dedicated Hierarchical Feature Fusion Module, the core of which is illustrated in Figure 8.

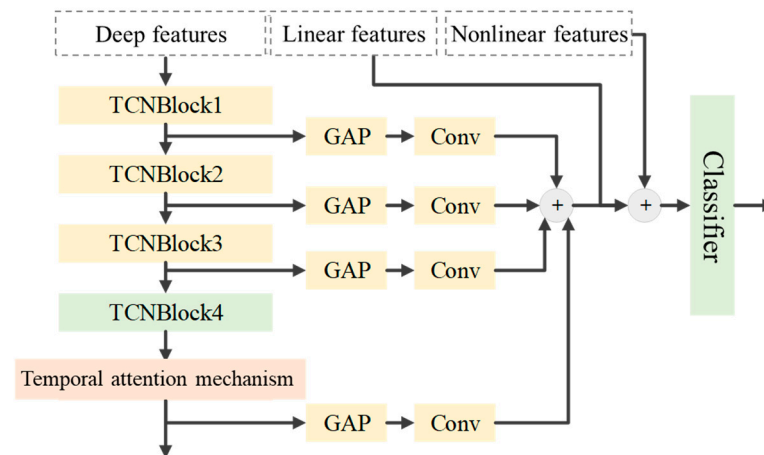


Figure 8. Hierarchical Feature Fusion Module and the Classifier.

The design of this module is driven by the need to preserve and synergize multi-scale contextual information. The process begins by harvesting deep feature maps from successive TCNBlocks. Each of these blocks inherently captures temporal dependencies at a unique scale; shallow blocks extract local, short-term patterns (e.g., morphological nuances of individual beats), while deeper blocks model more global, long-term context (e.g., heart rate trend over minutes). To convert these variable-length feature maps into a fixed-dimensional yet informative representation, we apply Global Average Pooling (GAP) after each TCNBlock. This operation not only standardizes the feature dimensions but also enhances translation invariance and mitigates overfitting by reducing the number of parameters.

The resulting vectors (32-, 64-, and 128-dimensional) encapsulate the hierarchical characteristics of the input signal. To facilitate a coherent fusion, these multi-scale vectors are then projected into a unified 64-dimensional semantic space using dedicated 1×1 convolutional layers. This crucial step serves two purposes: firstly, it achieves dimensional alignment for subsequent processing, and secondly, it performs a learnable, non-linear transformation that allows the model to calibrate and weight the contribution of each feature scale adaptively. This is a significant advancement over naive concatenation, as it enables feature interaction and refinement before the final fusion.

The final step involves the concatenation of the aligned multi-scale deep features with the handcrafted linear and nonlinear features. This creates a comprehensive and enriched representation that leverages both the abstract, high-level patterns learned by the TCN and the domain-specific, clinically grounded knowledge embedded in the handcrafted features. This fused feature vector is then passed through a two-layer fully connected classifier with ReLU activation and dropout regularization for the final SCD risk prediction. The entire hierarchical fusion strategy ensures that the model's decision is informed by a rich tapestry of information spanning from local signal details to global physiological trends, thereby significantly enhancing its explanatory power and predictive reliability.

4. Experiments and Results

4.1. Implementation Details and Evaluation Metrics

4.1.1. Implementation Details

The software and hardware environments used in this study are as follows: the operating system is Ubuntu 20.04.6 LTS (Canonical, London, UK), the NVIDIA graphics card is NVIDIA GeForce RTX 4090 (NVIDIA Inc., Santa Clara, CA, USA), the deep learning framework is PyTorch 2.1.0 (Meta Inc., Menlo Park, CA, USA), and the parallel computing platform is CUDA 12.1 (NVIDIA Inc., Santa Clara, CA, USA). During the model training process, the cross-entropy loss function and the Adaptive Moment Estimation (Adam) optimizer were adopted. The initial learning rate was set to 0.001, and it was reduced by a factor of 0.1 after every 5 training epochs. The batch size of the model was set to 32, and the maximum number of training epochs was set to 40 to ensure the stability of the model.

For linear feature extraction, the process involved computing 5-dimensional hand-crafted features from ECG signals, including time-domain, statistical, and morphological characteristics. These features were normalized using StandardScaler to ensure zero mean and unit variance. Key hyperparameters included a 15,000-point analysis window with no overlap and physiologically plausible heart rate constraints.

For the nonlinear feature extraction, Detrended Fluctuation Analysis (DFA) was employed to capture long-range correlations in heart rate variability, yielding a 1-dimensional scaling exponent. The DFA implementation required careful parameter selection including logarithmic scaling ranges (4–16 for short-term and 16–64 for long-term fluctuations), 50% segment overlap, and first-order detrending. Additional non-linear measures such as sample entropy were considered with embedding dimension 2 and tolerance threshold 0.2.

For the TCN-Seq2vec deep feature extraction module, a temporal convolutional network with multi-scale processing automatically learned hierarchical representations from raw ECG signals. The TCN architecture employed progressively increasing channels (32, 64, 128) with exponential dilation factors (1, 2, 4). A temporal attention mechanism with reduction to 64 channels and tanh activation highlighted clinically relevant segments. Comprehensive regularization included spatial dropout (0.5), feature dropout (0.3), and attention dropout (0.2) to prevent overfitting.

4.1.2. Evaluation Metrics

To comprehensively evaluate the performance of the proposed method, precision (Pre), recall (Rec), accuracy (Acc), and F1-score (F1) were selected as the evaluation metrics. Their calculation formulas are shown in Equations (11) to (14):

$$\text{Pre} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% \quad (11)$$

$$\text{Rec} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\% \quad (12)$$

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100\% \quad (13)$$

$$\text{F}_1 = 2 \times \frac{(\text{Pre} \times \text{Rec})}{(\text{Pre} + \text{Rec})} \quad (14)$$

Here, True Positive (TP) and True Negative (TN) represent the number of correctly classified positive samples and negative samples, respectively; False Positive (FP) and False Negative (FN) represent the number of incorrectly classified positive samples and negative samples, respectively.

4.2. Experimental Result

To evaluate the generalization capability of the proposed method, we conducted 10 independent repeated experiments for each of the six datasets that we obtained during the signal preprocessing (SCD10, SCD20, SCD30, SCD40, SCD50, SCD60). In each experiment, the dataset was randomly partitioned into training and testing sets following an inter-patient paradigm, with specific record numbers detailed in Table 3 (where each record number corresponds to a distinct patient). To systematically present the findings, Tables 4–9 provide detailed experimental results for six datasets.

Table 3. Record Identifiers for Training and Test Sets.

Experiment ID	Training Set (Record No.)	Test Set (Record No.)
1	30, 32, 34, 36, 37, 41, 43, 44, 45, 46, 47, 51, 52	31, 33, 39, 48, 50
2	30, 31, 33, 34, 36, 39, 41, 43, 44, 45, 48, 50, 51	32, 37, 46, 47, 52
3	30, 31, 32, 33, 34, 36, 37, 41, 43, 44, 45, 46, 48	39, 47, 50, 51, 52
4	30, 33, 37, 39, 41, 43, 45, 46, 47, 48, 50, 51, 52	31, 32, 34, 36, 44
5	31, 32, 33, 34, 39, 43, 44, 45, 46, 47, 48, 51, 52	30, 36, 37, 41, 50
6	30, 31, 32, 34, 36, 37, 41, 43, 44, 46, 48, 50, 51	33, 39, 45, 47, 52
7	30, 31, 32, 33, 36, 39, 43, 44, 45, 47, 48, 51, 52	34, 37, 41, 46, 50
8	30, 31, 32, 33, 37, 41, 45, 46, 47, 48, 50, 51, 52	34, 36, 39, 43, 44
9	30, 31, 32, 33, 34, 36, 37, 39, 44, 45, 46, 47, 48	41, 43, 50, 51, 52
10	30, 31, 32, 33, 34, 36, 37, 41, 43, 45, 47, 48, 51	39, 44, 46, 50, 52

Table 4. Experimental Results on the SCD10 Dataset.

Experiment ID	TP	FN	FP	TN	Rec	Pre	Acc	F ₁
1	50	0	0	50	100.00%	100.00%	100.00%	100.00%
2	50	0	1	49	99.00%	99.02%	99.00%	99.00%
3	50	0	1	49	99.00%	99.02%	99.00%	99.00%
4	50	0	1	49	99.00%	99.02%	99.00%	99.00%
5	49	1	0	50	99.00%	99.02%	99.00%	99.00%
6	50	0	1	49	99.00%	99.02%	99.00%	99.00%
7	49	1	1	49	98.00%	98.00%	98.00%	98.00%
8	50	0	1	49	99.00%	99.02%	99.00%	99.00%
9	49	1	0	50	99.00%	99.02%	99.00%	99.00%
10	49	1	0	50	99.00%	99.02%	99.00%	99.00%
Mean	—	—	—	—	99.00%	99.02%	99.00%	99.00%

Table 5. Experimental Results on the SCD20 Dataset.

Experiment ID	TN	FP	FN	TP	Rec	Pre	Acc	F ₁
1	50	0	0	50	100.00%	100.00%	100.00%	100.00%
2	50	0	0	50	100.00%	100.00%	100.00%	100.00%
3	50	0	3	47	97.00%	97.17%	97.00%	97.00%
4	50	0	1	49	99.00%	99.02%	99.00%	99.00%
5	50	0	1	49	99.00%	99.02%	99.00%	99.00%
6	50	0	3	47	97.00%	97.17%	97.00%	97.00%
7	50	0	0	50	100.00%	100.00%	100.00%	100.00%
8	50	0	0	50	100.00%	100.00%	100.00%	100.00%
9	49	1	0	50	99.00%	99.02%	99.00%	99.00%
10	50	0	1	49	99.00%	99.02%	99.00%	99.00%
Mean	—	—	—	—	99.00%	99.04%	99.00%	99.00%

Table 6. Experimental Results on the SCD30 Dataset.

Experiment ID	TN	FP	FN	TP	Rec	Pre	Acc	F ₁
1	49	1	0	50	99.00%	99.02%	99.00%	99.00%
2	50	0	1	49	99.00%	99.02%	99.00%	99.00%
3	49	1	1	49	98.00%	98.00%	98.00%	98.00%
4	50	0	3	47	97.00%	97.17%	97.00%	97.00%
5	49	1	1	49	98.00%	98.00%	98.00%	98.00%
6	50	0	1	49	99.00%	99.02%	99.00%	99.00%
7	49	1	1	49	98.00%	98.00%	98.00%	98.00%
8	50	0	1	49	99.00%	99.02%	99.00%	99.00%
9	49	1	1	49	98.00%	98.00%	98.00%	98.00%
10	49	1	0	50	99.00%	99.02%	99.00%	99.00%
Mean	—	—	—	—	98.40%	98.43%	98.40%	98.40%

Table 7. Experimental Results on the SCD40 Dataset.

Experiment ID	TN	FP	FN	TP	Rec	Pre	Acc	F ₁
1	47	3	0	50	97.00%	97.17%	97.00%	97.00%
2	50	0	1	49	99.00%	99.02%	99.00%	99.00%
3	50	0	8	42	92.00%	93.10%	92.00%	91.95%
4	50	0	2	48	98.00%	98.08%	98.00%	98.00%
5	47	3	1	49	96.00%	96.07%	96.00%	96.00%
6	50	0	2	48	98.00%	98.08%	98.00%	98.00%
7	47	3	1	49	96.00%	96.07%	96.00%	96.00%
8	50	0	3	47	97.00%	97.17%	97.00%	97.00%
9	50	0	2	48	98.00%	98.08%	98.00%	98.00%
10	50	0	1	49	99.00%	99.02%	99.00%	99.00%
Mean	—	—	—	—	97.00%	97.19%	97.00%	97.00%

Table 8. Experimental Results on the SCD50 Dataset.

Experiment ID	TN	FP	FN	TP	Rec	Pre	Acc	F ₁
1	50	0	5	45	95.00%	95.45%	95.00%	94.99%
2	48	2	1	49	97.00%	97.02%	97.00%	97.00%
3	47	3	1	49	96.00%	96.07%	96.00%	96.00%
4	50	0	4	46	96.00%	96.30%	96.00%	95.99%
5	50	0	4	46	96.00%	96.30%	96.00%	95.99%
6	50	0	3	47	97.00%	97.17%	97.00%	97.00%
7	48	2	1	49	97.00%	97.02%	97.00%	97.00%
8	50	0	2	48	98.00%	98.08%	98.00%	98.00%
9	50	0	4	46	96.00%	96.30%	96.00%	95.99%
10	48	2	1	49	97.00%	97.02%	97.00%	97.00%
Mean	—	—	—	—	96.50%	96.67%	96.50%	96.50%

As shown in Table 4, the SCD10 model achieved consistently high performance metrics (Rec = 99.00%, Pre = 99.02%, Acc = 99.00%, F1 = 99.00%) during the 0–10 min window preceding ventricular fibrillation (VF) onset, demonstrating excellent stability across all experimental trials.

Table 5 reveals that the SCD20 model maintained equally outstanding performance (Rec = 99.00%, Pre = 99.04%, Acc = 99.00%, F1 = 99.00%) in the 10–20 min pre-VF window, matching the SCD10 model’s effectiveness.

The results in Table 6 indicate a slight performance degradation for the SCD30 model (Rec = 98.40%, Pre = 98.43%, Acc = 98.40%, F1 = 98.40%) during the 20–30 min pre-VF window, though it retained high accuracy and stability compared to earlier time windows.

Performance trends show further decline in the 30–40 min pre-VF window (Table 7), where the SCD40 model yielded metrics of Rec = 97.00%, Pre = 97.19%, Acc = 97.00%, F1 = 97.00%. Notably, Experiments 3 and 5 demonstrated significantly lower performance, suggesting reduced model stability compared to shorter prediction windows.

Table 9. Experimental Results on the SCD60 Dataset.

Experiment ID	TN	FP	FN	TP	Rec	Pre	Acc	F1
1	50	0	7	43	93.00%	93.86%	93.00%	92.97%
2	50	0	2	48	98.00%	98.08%	98.00%	98.00%
3	41	9	0	50	91.00%	92.37%	91.00%	90.93%
4	50	0	5	45	95.00%	95.45%	95.00%	94.99%
5	50	0	5	45	95.00%	95.45%	95.00%	94.99%
6	50	0	6	44	94.00%	94.64%	94.00%	93.98%
7	50	0	5	45	95.00%	95.45%	95.00%	94.99%
8	50	0	3	47	97.00%	97.17%	97.00%	97.00%
9	46	4	0	50	96.00%	96.30%	96.00%	95.99%
10	47	3	1	49	96.00%	96.07%	96.00%	96.00%
Mean	—	—	—	—	95.00%	95.48%	95.00%	94.98%

This downward trend continues in the 40–50 min pre-VF window (Table 8), with the SCD50 model achieving Rec = 96.50%, Pre = 96.67%, Acc = 96.50%, F1 = 96.50%. The increased frequency of false positives (FPs) across multiple experiments indicates deteriorating model reliability.

The poorest performance occurs in the 50–60 min pre-VF window (Table 9), where the SCD60 model's metrics drop to Rec = 95.00%, Pre = 95.48%, Acc = 95.00%, F1 = 94.98%. The substantial increase in both false negatives (FNs) and FPs confirms significantly reduced model stability at this extended prediction horizon.

The SCD10 and SCD20 models demonstrated exceptional performance across all evaluation metrics, maintaining both high accuracy ($\geq 99\%$) and remarkable stability. These results indicate that our proposed method can reliably predict SCD events within the critical 20 min window preceding onset.

While the SCD30 and SCD40 models showed modest performance degradation compared to their shorter-term counterparts, they still maintained accuracy levels above 97%. The observed decline became more pronounced in the SCD50 and SCD60 models, particularly for SCD60, potentially due to diminished feature discriminability between SCD and NSR samples in these extended timeframes.

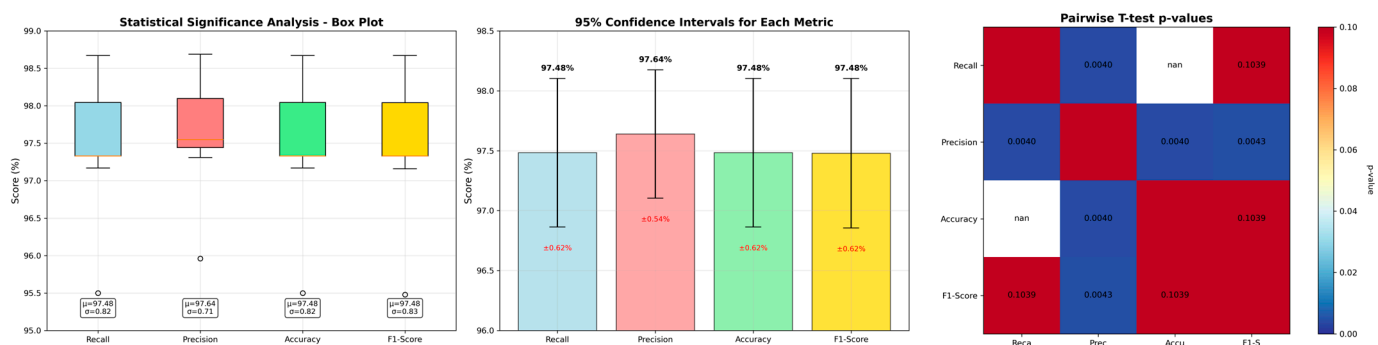
Our analysis reveals an important temporal pattern: SCD signals occurring closer to VF onset exhibit more distinct characteristics compared to normal ECG, while those further removed (50–60 min pre-onset) demonstrate greater similarity to normal sinus rhythm. This finding explains the observed performance gradient across time windows.

As summarized in Table 10, the proposed risk prediction model achieved an average accuracy of 97.48% across 10 independent trials, with individual experiment accuracy ranging from 95.50% to 98.67%. The consistent performance (Rec, Pre, and F1 all $>97\%$) confirms balanced and stable identification of positive cases. The relatively lower accuracy (95.50%) in Experiment 3 may be attributed to reduced feature discriminability in its particular test set composition.

Table 10. Performance Metrics of the SCD Risk Prediction Model.

Experiment ID	Rec	Pre	Acc	F1
1	97.33%	97.58%	97.33%	97.33%
2	98.67%	98.69%	98.67%	98.67%
3	95.50%	95.96%	95.50%	95.48%
4	97.33%	97.51%	97.33%	97.33%
5	97.17%	97.31%	97.17%	97.16%
6	97.33%	97.52%	97.33%	97.33%
7	97.33%	97.42%	97.33%	97.33%
8	98.33%	98.41%	98.33%	98.33%
9	97.67%	97.79%	97.67%	97.66%
10	98.17%	98.20%	98.17%	98.17%
Mean	97.48%	97.64%	97.48%	97.48%

As shown in Figure 9, statistical analysis was conducted to assess significance between performance metrics. The model achieved an average recall of 97.48% (SD = 0.82%), precision of 97.64% (SD = 0.71%), and F1-score of 97.48% (SD = 0.83%). Pairwise t-tests revealed that precision was statistically significantly higher than both recall ($p = 0.004$) and F1-score ($p = 0.004$), while no significant difference was found between recall and F1-score ($p = 0.104$). The lower standard deviation observed for precision (0.71%) compared to recall (0.82%) and F1-score (0.83%) suggests more consistent performance in positive predictive value across experimental trials. All metrics demonstrated low variability (SD < 0.85%), indicating stable model performance.

**Figure 9.** Statistical analysis and p -value heatmap.

These comprehensive results demonstrate that our method provides reliable predictions in most experimental scenarios, exhibiting both good generalizability and robustness for clinical SCD risk assessment.

Table 11 presents a comprehensive comparison between our proposed method and existing approaches in the literature. The results demonstrate significant advancements in both prediction accuracy and clinical applicability. The comparative analysis follows the standard practice in the field, benchmarking reported performances on the common task of SCD prediction from short-term ECG signals. While direct implementation of all cited methods with an identical preprocessing pipeline is not feasible, the comparison remains valid as it reflects the performance achieved by different methodological paradigms on standard clinical prediction tasks using public datasets. The consistent and significant performance advantage of our proposed method, as established through our rigorous internal validation, indicates that its superiority is primarily attributable to the model architecture rather than minor variations in data preparation.

Table 11. Comparison of the proposed method with other studies.

Authors	Methods	Signals	Data Duration	Acc
Tseng et al. [31]	CNN	ECG	5 min before SCD	88.00%
Chen et al. [32]	Clustering	HRV	5 min before SCD	98.40%
Khazaei et al. [13]	KNN	HRV	6 min before SCD	95.00%
Ebrahimzadeh et al. [14]	ME	HRV	13 min before SCD	82.85%
Shi et al. [16]	KNN	ECG	14 min before SCD	97.60%
Abrishami et al. [33]	LSTM	ECG	15 min before SCD	90.00%
Gao et al. [21]	SVM	ECG	30 min before SCD	95.43%
Saragih, Isa [34]	1D-CNN	ECG	30 min before SCD	95.89%
Centeno-Bautista et al. [35]	SVM	ECG	30 min before SCD	97.28%
Jablo et al. [36]	SVM/KNN	ECG	60 min before SCD	94.42%
Ours	TCN	ECG, HRV	30 min before SCD 60 min before SCD	98.80% 97.48%

Tseng et al. [31] developed a CNN-based model utilizing 2D short-time Fourier transform or continuous wavelet transform (CWT) for ECG feature extraction, achieving 88% accuracy in SCD prediction 5 min before onset. Chen et al. [32] employed phase space reconstruction and fuzzy C-means clustering for HRV analysis, reaching 98.40% accuracy at the 5 min prediction window.

Several studies have explored nonlinear analysis techniques: Khazaei et al. [13] applied increment entropy and recurrence quantification analysis to HRV signals, obtaining 95% accuracy 6 min pre-SCD. Ebrahimzadeh et al. [14] combined time-domain, frequency-domain, time-frequency, and nonlinear features with TLSFS-based feature selection, achieving 82.85% accuracy 13 min before SCD onset.

Advanced signal processing methods have shown promising results: Shi et al. [16] implemented discrete wavelet transform (DWT) and locality preserving projections (LPP) with sophisticated feature ranking techniques, attaining 97.6% accuracy 14 min pre-SCD using only 5 LPP features. Centeno-Bautista et al. [35] combined complete ensemble empirical mode decomposition (CEEMD) with statistical analysis and SVM classification, reaching 97.28% accuracy 30 min before SCD.

Longer-term prediction studies include Gao et al. [21], who developed a specialized algorithm for low-SNR single-lead ECG signals incorporating 12-dimensional features (including ventricular late potentials and T-wave alternans), achieving 93.22% accuracy 30 min pre-SCD (improving to 95.43% with NSR controls). Abrishami et al. [33] proposed a Deep Bidirectional LSTM (DBLSTM) network for ECG interval segmentation, achieved 90% accuracy in T-Wave segmentation for ECG analysis. This architecture employs stacked bidirectional LSTM layers to process sequences in both forward and backward directions. Saragih and Isa [34] proposed an SCA prediction system that integrates Wavelet Packet Transform with a CNN to identify subtle patterns in ECG recordings. Evaluated through 10-fold cross-validation on one-minute segments taken 30 min before onset, the model attained an accuracy of 95.89%. Jablo et al. [36] employed empirical mode decomposition (EMD) with nonlinear feature extraction and SVM/KNN classification, obtaining 94.42% accuracy 60 min before SCD onset.

Notably, our method requires only three carefully selected features to achieve superior performance (97.48% accuracy) at the 60 min prediction window, while offering three key advantages:

- (1) Providing substantially longer clinical intervention time (60 min vs. typically <30 min in literature)
- (2) Maintaining lower computational complexity through efficient feature selection
- (3) Delivering more consistent performance across extended prediction windows

This comparative analysis demonstrates that our approach represents a significant improvement over existing methods in terms of both early prediction capability and practical clinical implementation.

To further validate the proposed method, we performed additional evaluation using the Creighton University Ventricular Tachyarrhythmia Database (CUDB), which contains 35 eight-minute ECG recordings. Based on expert annotations, we identified 7 subjects with documented ventricular fibrillation (VF), ventricular tachycardia (VT), or atrial fibrillation (AF) (Record Numbers: cu01, cu02, cu03, cu09, cu16, cu18, cu21).

Following the same experimental protocol, we conducted 10 repeated random trials using an inter-patient paradigm for dataset partitioning. The training set was composed of 13 patients from the NSR dataset and 18 patients from the SCD dataset, while the test set included the remaining 5 patients from the NSR dataset along with 5 patients from CUDB (Record Numbers: cu01, cu02, cu03, cu09, cu16).

As demonstrated in Table 12, our method achieved consistently strong performance on this independent validation dataset. These results indicate that the proposed approach maintains excellent generalization capability when applied to different patient populations and clinical datasets. The robust performance across multiple validation scenarios underscores the method's potential for reliable clinical implementation in SCD risk assessment.

Table 12. Performance of the Proposed Method on the CUDB Dataset.

Experiment ID	TN	FP	FN	TP	Rec	Pre	Acc	F1
1	50	0	2	14	87.50%	100.00%	96.97%	93.33%
2	49	1	2	14	87.50%	93.33%	95.45%	90.32%
3	50	0	1	15	93.75%	100.00%	98.48%	96.77%
4	50	0	1	15	93.75%	100.00%	98.48%	96.77%
5	50	0	0	16	100.00%	100.00%	100.00%	100.00%
6	49	1	2	14	87.50%	93.33%	95.45%	90.32%
7	48	2	3	13	81.25%	86.67%	92.42%	83.87%
8	50	0	0	16	100.00%	100.00%	100.00%	100.00%
9	50	0	1	15	93.75%	100.00%	98.48%	96.77%
10	49	1	1	15	93.75%	93.75%	96.97%	93.75%
Mean	—	—	—	—	91.88%	96.71%	97.27%	94.19%

4.3. Ablation Experiment

To further validate the effectiveness of the proposed method, we performed ablation studies by decomposing the complete model into four sub-models: Sub-model A extracting linear features from ECG signals; Sub-model B deriving nonlinear features (DFA2 features) from RR interval sequences; Sub-model C employing the TCN-Seq2vec model to extract deep learning features from ECG signals; and Sub-model D integrating all features through the fusion module. The ablation results are summarized in Table 13.

The experimental results demonstrate that Sub-model A achieves stable performance across all metrics with a recall (Rec) approaching 90%, indicating its sensitivity to SCD-related features. Sub-model B shows improved performance over Sub-model A in all evaluation metrics, suggesting enhanced discriminative capability for SCD characteristics. Notably, Sub-model C outperforms both preceding sub-models, validating the superior effectiveness of deep learning-based feature extraction. The complete model (Sub-model D), which integrates all three feature types, achieves optimal performance with Rec, Pre, Acc, and F1 scores all exceeding 97%, while maintaining low false positive and false negative rates. This performance gap is not attributable to the mere presence of multiple features, but specifically to the hierarchical fusion mechanism. The module's ability to align, calibrate, and synergistically combine multi-scale deep features with handcrafted features—rather

than treating them indifferently—is the key driver behind this enhancement. These results conclusively demonstrate that our hierarchical fusion strategy is uniquely effective in unlocking the full, complementary potential of heterogeneous features, thereby achieving state-of-the-art predictive performance for SCD risk.

Table 13. Results of the Ablation Study (The checkmark (✓) indicates the inclusion of a component, while the dash (—) denotes its exclusion).

Model Variants				Rec	Pre	Acc	F ₁
A	B	C	D				
✓	—	—	—	89.93%	93.72%	89.93%	91.64%
—	✓	—	—	90.53%	94.03%	90.53%	92.12%
—	—	✓	—	94.03%	95.55%	94.03%	94.76%
✓	✓	—	—	94.75%	95.91%	94.75%	95.31%
✓	—	✓	—	94.52%	95.77%	94.52%	95.12%
—	✓	✓	—	94.82%	95.98%	94.82%	95.37%
✓	✓	✓	—	95.57%	96.58%	95.57%	96.05%
✓	✓	✓	✓	97.48%	97.64%	97.48%	97.48%

4.4. Complexity Analysis of the Proposed Method

To thoroughly evaluate the computational efficiency and practical deployment potential of our proposed model, we conducted a detailed analysis of its time and space complexity. The analysis indicates that the model achieves a favorable balance between high performance and computational cost, making it suitable for potential clinical applications.

Time complexity is measured in terms of the number of Floating Point Operations (FLOPs) required for a single forward pass of one input sample. Space complexity is quantified by the total number of learnable parameters. The detailed breakdown is presented in Table 14.

Table 14. Complexity Analysis of the Proposed TCN-Seq2Vec Model.

Module	Time Complexity (FLOPs)	Space Complexity (Parameters)	Computational Load
TCN Block #1 (1 → 32)	1.44 M	2.9 K	0.2%
TCN Block #2 (32 → 64)	92.16 M	18.4 K	13.1%
TCN Block #3 (64 → 128)	368.64 M	73.7 K	52.5%
Temporal Attention Module	150.00 M	52.2 K	21.4%
Feature Projection Layers	0.58 M	12.4 K	0.1%
Classifier	0.03 M	33.3 K	<0.1%
Total	~612.85 M	~192.9 K	100%

The analysis in Table 14 yields several key observations:

Localized Computational Burden: The computational bottleneck is concentrated in the TCN module, with the third TCN block alone accounting for over 52% of the total FLOPs. This is a direct result of its large number of output channels (128) operating on the full-length sequence ($L = 15,000$). This design is an intentional trade-off to extract high-level, abstract features from the ECG signal, which is crucial for achieving high accuracy in SCD risk prediction.

High Parameter Efficiency: Despite the significant computational load, the model maintains a relatively modest parameter count of approximately 192.9 K parameters. This demonstrates high parameter efficiency, which is largely attributable to the weight-sharing mechanism of one-dimensional convolutions. This efficiency allows the model to effectively

process long sequences without an explosion in the number of parameters, reducing the risk of overfitting.

Feasible Deployment Profile: With a total complexity of approximately 612.85 MFLOPs per sample, the model is computationally feasible for modern GPU hardware. When processed in batches (e.g., Batch Size = 16), real-time or near-real-time inference can be readily achieved. This level of complexity is justifiable given the critical nature of the SCD prediction task and supports the model's potential for future clinical integration.

In summary, the proposed model strikes a balance between performance and efficiency by strategically allocating computational resources to its core feature extractor (the TCN), resulting in a robust yet practically deployable architecture for SCD risk assessment.

5. Conclusions

To address the challenges of effective feature extraction and significant inter-individual variability in existing sudden cardiac death (SCD) prediction research, we propose a novel SCD risk prediction framework based on multi-level hierarchical feature fusion. Our approach is designed to comprehensively model the inherent non-stationary and asymmetrical characteristics of ECG signals. The methodology comprises four key components: (1) conventional waveform detection for extracting linear features from ECG signals, (2) DFA2-based analysis of RR interval time series to capture nonlinear characteristics, (3) TCN-Seq2vec modeling for multi-scale deep learning feature extraction from raw ECG data, and (4) integrated fusion of these complementary feature sets for final classification. This comprehensive approach achieves reliable SCD risk prediction with a 60 min early warning window, demonstrating superior performance compared to conventional single-modality methods.

6. Discussion and Future Work

This study presents a deep learning framework for sudden cardiac death prediction using single-lead ECG. The proposed model demonstrates robust performance, achieving a balanced F1-score of 97.48% with statistically significant precision (97.64%, $p < 0.01$) across 10 independent trials. While these results highlight the clinical potential for wearable-based monitoring, several limitations warrant discussion and outline directions for future research. First, the model's evaluation on a single-center dataset of limited size necessitates further validation on larger, multi-center cohorts. We therefore explicitly state that evaluating the proposed framework on contemporary datasets such as MUSIC (MUerte subita en insuficiencia cardiaca) will be a primary objective of our immediate future research. Second, the current analysis focused on a clinically relevant operating point and we will consider comprehensive ROC and AUC analysis in our subsequent studies for extended threshold exploration. Third, we acknowledge practical constraints including the reliance on single-lead recordings and the computational demands for real-time monitoring. Future work will address these through multi-lead integration where feasible, and model optimization techniques (e.g., pruning, quantization) to enhance deployment feasibility on wearable devices.

By transparently addressing these limitations while building upon the demonstrated strengths, this work establishes a foundation for reliable SCD risk prediction and its eventual clinical translation.

Author Contributions: Conceptualization, X.H. (Xin Huang), X.H. (Xiaoyu He) and M.J.; Methodology, X.H. (Xin Huang), M.H., Y.L. and M.J.; Software, M.H.; Formal analysis, X.H. (Xiaoyu He); Investigation, X.H. (Xin Huang), M.H. and X.H. (Xiaoyu He); Data curation, X.H. (Xin Huang) and Y.L.; Writing—original draft, X.H. (Xin Huang) and M.H.; Writing—review & editing, X.H. (Xin Huang), G.J., Y.L. and M.J.; Visualization, X.H. (Xiaoyu He); Supervision, X.H. (Xiaoyu He), Y.L.

and M.J.; Funding acquisition, G.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science Foundation of Zhejiang Sci-Tech University (ZSTU) (Grant No. 22232337-Y), the National Key Research and Development Program of China (2023YFE0205600), the National Natural Science Foundation of China (62272415), the Special Support Program for High-level Talents of Zhejiang Province (2023R5216), and the Key Research and Development Program of Ningxia Province (2023BEG02065).

Data Availability Statement: The datasets used in this study are from the MIT-BIH Database available online at <https://archive.physionet.org/physiobank/database/sddb/> (accessed on 4 October 2025); <https://archive.physionet.org/physiobank/database/nsrdb/> (accessed on 4 October 2025).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zipes, D.P.; Wellens, H.J.J. Sudden cardiac death. *Circulation* **1998**, *98*, 2334–2351. [CrossRef]
2. Roth, G.A.; Mensah, G.A.; Johnson, C.O.; Addolorato, G.; Ammirati, E.; Baddour, L.M.; Barengo, N.C.; Beaton, A.Z.; Benjamin, E.J.; Benziger, C.P.; et al. Global burden of cardiovascular diseases and risk factors, 1990–2019: Update from the GBD 2019 study. *J. Am. Coll. Cardiol.* **2020**, *76*, 2982–3021. [CrossRef] [PubMed]
3. Linh, V.T.N.; Han, S.; Koh, E.; Kim, S.; Jung, H.S.; Koo, J. Advances in wearable electronics for monitoring human organs: Bridging external and internal health assessments. *Biomaterials* **2025**, *314*, 122865. [CrossRef] [PubMed]
4. Jena, N.; Singh, P.; Chandramohan, D.; Garapati, H.N.; Gummadi, J.; Mylavarapu, M.; Shaik, B.F.; Nanjundappa, A.; Apala, D.R.; Toquica, C.; et al. Wearable Technology in Cardiology: Advancements, Applications, and Future Prospects. *Rev. Cardiovasc. Med.* **2025**, *26*, 39025. [CrossRef] [PubMed]
5. Abdou, A.; Krishnan, S. Horizons in single-lead ECG analysis from devices to data. *Front. Signal Process.* **2022**, *2*, 866047. [CrossRef]
6. Razak, S.F.A.; Ismail, S.N.M.S.; Yogarayan, S.; Abdullah, M.F.A.; Kamis, N.H.; Aziz, A.A. Comparative study of machine learning algorithms in classifying HRV for the driver's physiological condition. *Civ. Eng. J.* **2023**, *9*, 2272–2285. [CrossRef]
7. Sessa, F.; Anna, V.; Messina, G.; Cibelli, G.; Monda, V.; Marsala, G.; Ruberto, M.; Biondi, A.; Cascio, O.; Bertozzi, G.; et al. Heart rate variability as predictive factor for sudden cardiac death. *Aging* **2018**, *10*, 166. [CrossRef]
8. Kaspal, R.; Alsadoon, A.; Prasad, P.W.C.; Al-Saiyd, N.A.; Nguyen, T.Q.V.; Pham, D.T.H. A novel approach for early prediction of sudden cardiac death (SCD) using hybrid deep learning. *Multimed. Tools Appl.* **2021**, *80*, 8063–8090. [CrossRef]
9. Barker, J.; Li, X.; Khavandi, S.; Koeckerling, D.; Mavilakandy, A.; Pepper, C.; Bountziouka, V.; Chen, L.; Kotb, A.; Antoun, I.; et al. Machine learning in sudden cardiac death risk prediction: A systematic review. *Europace* **2022**, *24*, 1777–1787. [CrossRef]
10. Hannun, A.Y.; Rajpurkar, P.; Haghpanahi, M.; Tison, G.H.; Bourn, C.; Turakhia, M.P.; Ng, A.Y. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat. Med.* **2019**, *25*, 65–69. [CrossRef]
11. Wang, K.; Zhang, K.; Liu, B.; Chen, W.; Han, M. Early prediction of sudden cardiac death risk with Nested LSTM based on electrocardiogram sequential features. *BMC Med. Inform. Decis. Mak.* **2024**, *24*, 94. [CrossRef]
12. Bizopoulos, P.; Koutsouris, D. Deep learning in cardiology. *IEEE Rev. Biomed. Eng.* **2018**, *12*, 168–193. [CrossRef] [PubMed]
13. Khazaei, M.; Raeisi, K.; Goshvarpour, A.; Ahmadzadeh, M. Early detection of sudden cardiac death using nonlinear analysis of heart rate variability. *Biocybern. Biomed. Eng.* **2018**, *38*, 931–940. [CrossRef]
14. Ebrahimzadeh, E.; Foroutan, A.; Shams, M.; Baradaran, R.; Rajabion, L.; Joulani, M.; Fayaz, F. An optimal strategy for prediction of sudden cardiac death through a pioneering feature-selection approach from HRV signal. *Comput. Methods Programs Biomed.* **2019**, *169*, 19–36. [CrossRef] [PubMed]
15. Shi, M.; He, H.; Geng, W.; Wu, R.; Zhan, C.; Jin, Y.; Zhu, F.; Ren, S.; Shen, B. Early detection of sudden cardiac death by using ensemble empirical mode decomposition-based entropy and classical linear features from heart rate variability signals. *Front. Physiol.* **2020**, *11*, 118. [CrossRef]
16. Shi, M.; Yu, H.; Wang, H. Automated detection of sudden cardiac death by discrete wavelet transform of electrocardiogram signal. *Symmetry* **2022**, *14*, 571. [CrossRef]
17. Reddy, K.V.; Kumar, N. Automated prediction of sudden cardiac death using statistically extracted features from electrocardiogram signals. *Int. J. Electr. Comput. Eng.* **2022**, *12*, 4960–4969. [CrossRef]
18. Centeno-Bautista, M.A.; Rangel-Rodriguez, A.H.; Perez-Sanchez, A.V.; Amezcua-Sanchez, J.P.; Granados-Lieberman, D.; Valtierra-Rodriguez, M. Electrocardiogram analysis by means of empirical mode decomposition-based methods and convolutional neural networks for sudden cardiac death detection. *Appl. Sci.* **2023**, *13*, 3569. [CrossRef]

19. Yang, J.; Sun, Z.; Zhu, W.; Xiong, P.; Du, H.; Liu, X. Intelligent prediction of sudden cardiac death based on multi-domain feature fusion of heart rate variability signals. *EURASIP J. Adv. Signal Process.* **2023**, 2023, 32. [\[CrossRef\]](#)
20. Telangore, H.; Azad, V.; Sharma, M.; Bhurane, A.; Tan, R.S.; Acharya, U.R. Early prediction of sudden cardiac death using multimodal fusion of ECG Features extracted from Hilbert–Huang and wavelet transforms with explainable vision transformer and CNN models. *Comput. Methods Programs Biomed.* **2024**, 257, 108455. [\[CrossRef\]](#)
21. Gao, W.; Liao, J. Sudden Cardiac Death Risk Prediction Based on Noise Interfered Single-Lead ECG Signals. *Electronics* **2024**, 13, 4274. [\[CrossRef\]](#)
22. Greenwald, S.D. The development and analysis of a ventricular fibrillation detector. Master’s Thesis, MIT Dept. of Electrical Engineering and Computer Science, Cambridge, MA, USA, 1986.
23. Goldberger, A.; Amaral, L.; Glass, L.; Hausdorff, J.; Ivanov, P.C.; Mark, R.; Mietus, J.E.; Moody, G.B.; Peng, C.-K.; Stanley, H.E. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation* **2000**, 101, e215–e220. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Xie, C.X.; Wang, L.H.; Yu, Y.T.; Ding, L.J.; Yang, T.; Kuo, I.C.; Wang, X.-K.; Gao, J.; Abu, P.A.R. Clinical sudden cardiac death risk prediction: A grid search support vector machine multimodel base on ventricular fibrillation visualization features. *Comput. Electr. Eng.* **2025**, 123, 110022. [\[CrossRef\]](#)
25. Mallat, S.G. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **1989**, 11, 674–693. [\[CrossRef\]](#)
26. Balta, D.; Akyemiş, E.M. Arrhythmia detection using Pan-Tompkins algorithm and Hilbert transform with real-time ECG signals. *Acad. Perspect. Procedia* **2021**, 4, 307–315. [\[CrossRef\]](#)
27. Zhang, Q.; Manriquez, A.I.; Medigue, C.; Papelier, Y.; Sorine, M. An algorithm for robust and efficient location of T-wave ends in electrocardiograms. *IEEE Trans. Biomed. Eng.* **2006**, 53, 2544–2552. [\[CrossRef\]](#)
28. Pukkila, T.; Molkari, M.; Kim, J.; Räsänen, E. Reduced RR interval correlations of long QT syndrome patients. In Proceedings of the 2022 Computing in Cardiology (CinC), Tampere, Finland, 4–7 September 2022; IEEE: Boston, MA, USA; Volume 498, pp. 1–4.
29. Hu, K.; Ivanov, P.C.; Chen, Z.; Carpena, P.; Stanley, H.E. Effect of trends on detrended fluctuation analysis. *Phys. Rev. E* **2001**, 64, 011114. [\[CrossRef\]](#)
30. Hernesniemi, J.A.; Pukkila, T.; Molkari, M.; Nikus, K.; Lyytikäinen, L.-P.; Lehtimäki, T.; Viik, J.; Kähönen, M.; Räsänen, E. Prediction of sudden cardiac death with ultra-short-term heart rate fluctuations. *Clin. Electrophysiol.* **2024**, 10, 2010–2020. [\[CrossRef\]](#)
31. Tseng, L.M.; Tseng, V.S. Predicting ventricular fibrillation through deep learning. *IEEE Access* **2020**, 8, 221886–221896. [\[CrossRef\]](#)
32. Chen, H.; Das, S.; Morgan, J.M.; Maharatna, K. Prediction and classification of ventricular arrhythmia based on phase-space reconstruction and fuzzy c-means clustering. *Comput. Biol. Med.* **2022**, 142, 105180. [\[CrossRef\]](#)
33. Abrishami, H.; Han, C.; Zhou, X.; Campbell, M.; Czosek, R. Supervised ECG interval segmentation using LSTM neural network. In Proceedings of the International Conference on Bioinformatics & Computational Biology (BIOCOMP), Las Vegas, NV, USA, 30 July–2 August 2018; pp. 71–77.
34. Saragih, Y.V.; Isa, S.M. CNN performance improvement using wavelet packet transform for SCA prediction. *J. Theor. Appl. Inf. Technol.* **2022**, 100, 5458–5468.
35. Centeno-Bautista, M.A.; Perez-Sanchez, A.V.; Amezcua-Sanchez, J.P.; Valtierra-Rodriguez, M. Sudden cardiac death prediction based on the complete ensemble empirical mode decomposition method and a machine learning strategy by using ECG signals. *Measurement* **2024**, 236, 115052. [\[CrossRef\]](#)
36. Jabło, F.D.; Hesar, H.D. A novel method for early prediction of sudden cardiac death through nonlinear feature extraction from ECG signals. *Phys. Eng. Sci. Med.* **2025**, 48, 343–358. [\[CrossRef\]](#) [\[PubMed\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.