



Hardware-accelerated neural network model for early prediction of sudden cardiac arrest based on heart rate variability metrics

Sheng-Yueh Pan¹ · Duc Huy Nguyen¹ · Paul C.-P. Chao¹

Received: 2 February 2024 / Accepted: 20 September 2024 / Published online: 16 October 2024
© The Author(s) 2024

Abstract

Sudden Cardiac Arrest (SCA) constitutes a dire medical condition, marked by the abrupt cessation of effective blood circulation due to the heart's failure to contract properly. This leads to acute circulatory collapse, often culminating in loss of consciousness within an hour and potentially resulting in fatality within minutes if left unattended. Heart rate variability (HRV) serves as a critical biometric, derived from electrocardiogram (ECG) signals with ventricular depolarization waves to further calculate the R-R Intervals (RRIs). These intervals provide the basis for extracting various characteristics of cardiac rhythm, encompassing time-domain, frequency-domain, and nonlinear features. **This study presents a neural network-based classification algorithm that leverages HRV metrics to categorize patients into SCA and Normal Sinus Rhythm (NSR) cohorts. The proposed neural network (NN) model showcased impressive results, achieving an accuracy of 96.23%, a sensitivity of 94.35%, and a specificity of 98.12% in detecting SCA, as evaluated through leave-one-subject-out analysis.** In order to harness the benefits of hardware acceleration, the algorithm is implemented on a Field-Programmable Gate Array (FPGA). Its computational efficiency is subsequently benchmarked against traditional software-based methodologies. **The hardware-level implementation is made possible in Verilog hardware description language (HDL) and was verified successfully with expected performance by register-transfer level (RTL) simulation via Vivado 2020.2.**

1 Introduction

Cardiovascular diseases continue to be a significant global healthcare burden, with sudden cardiac arrest (SCA) emerging (Ha et al. 2022) as one of the most critical and fatal cardiovascular events. Characterized by the sudden cessation of the heart's electrical activities, SCA often progresses to sudden cardiac death (SCD) within a matter of minutes if not promptly addressed. To empirically validate the significance of SCA and its indicators, electrocardiograms (ECG) datasets were obtained from the MIT-BIH Physionet database (Goldberger, et al. 2000). Specifically, the MIT-BIH normal sinus rhythm database for healthy controls and the sudden cardiac death Holter

database for SCA patients were employed as presented in Fig. 1. Despite advancements in emergency medical response systems (Shah 2006), recent data reveal an alarmingly low post-SCA survival rate, underscoring the pressing need for improvements in early diagnosis and rapid intervention.

Traditionally, ECG has been at the forefront of cardiac diagnostics, providing invaluable insights into the heart's electrical patterns and rhythms. Their non-invasive nature coupled with their ability to yield continuous recordings has cemented their role in cardiac monitoring. As the field of cardiology evolved, researchers have pivoted their attention to more intricate measurements derived from ECGs. One such metric that has garnered considerable attention is heart rate variability (HRV). Recent studies have underscored the potential of HRV, revealing its capacity to serve as a predictive metric for sudden arrhythmic episodes. Especially noteworthy are its implications in anticipating ventricular tachyarrhythmias (VT) and ventricular fibrillation (VF) (Daubert et al. 2006), which are significant culprits in the onset of SCA and its fatal progression to SCD. Moreover, numerous studies (Murugappan et al. 2021; Abubakar et al. 2018; Alonso-

✉ Paul C.-P. Chao
pchao@nycu.edu.tw

Sheng-Yueh Pan
jason1019.ece09g@nctu.edu.tw

Duc Huy Nguyen
ndhuyvn199@gmail.com

¹ Department of Electrical and Control Engineering, National Yang Ming Chiao Tung University, Hsinchu 300, Taiwan

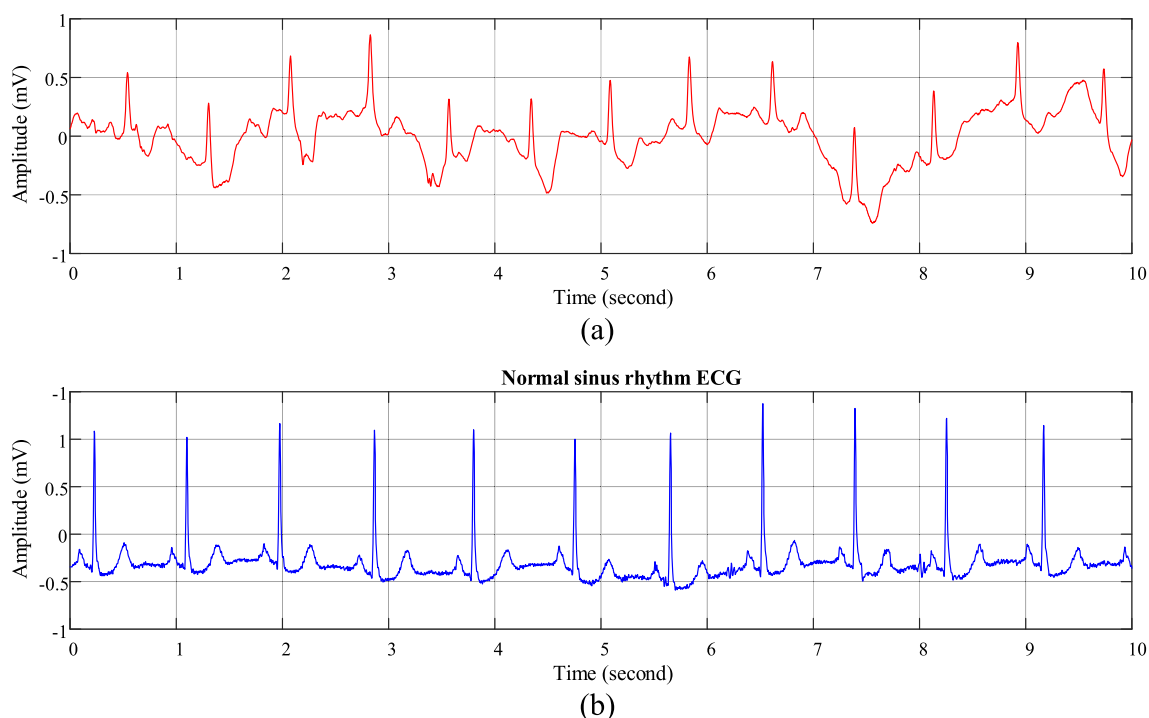


Fig. 1 **a** Sudden cardiac arrest due to VT/VF ECG waveform; **b** Normal sinus rhythm ECG waveform

Atienza et al. 2013; Shen et al. 2007; Lai et al. 2019) have leveraged ECG specifically for the early detection and analysis of SCA, underscoring its pivotal role in this domain. While the promising nature of HRV as a diagnostic tool is evident, it is crucial to consider the methodologies employed in its analysis. Historically, linear analysis techniques have been the mainstay. However, they are not without their drawbacks. One of the primary challenges with such methods is the requisite for extended data lengths to achieve accurate frequency-domain HRV computations. This demand can potentially lead to delayed interventions, especially in situations where every second counts. Furthermore, a significant concern arises when relying solely on linear techniques; they tend to overlook the inherent complexities of HRV and ECG signals. The heart's electrophysiological signals are intrinsically non-linear and exhibit non-stationary behavior (Berwal et al. 2019). By confining the analysis to linear methods, there's a potential loss in capturing the full spectrum of information these signals offer, potentially compromising the richness and depth of insights that can be gleaned.

While there is an acknowledgment in existing literature about the potential of nonlinear signal processing techniques in forecasting SCA and SCD events (Suboh et al. 2019), discernible gaps in research remain. A significant number of these studies have been observed to provide short prediction windows, often reducing their practical application in real-world emergency settings. Moreover,

there's a prevailing focus on software-based methodologies, which, although effective in certain contexts, may not harness the full computational potential available with modern hardware. In addressing these gaps, an innovative approach was undertaken. A Neural Network-centric classification model is meticulously developed and refined, tailoring it specifically for the early and accurate prediction of SCA events induced by VT/VF. Distinct from previous studies, this model is architected to incorporate an expansive feature set, designed to capture nuanced cardiac signal variations that could be indicative of impending SCA. Beyond just offering a broader spectrum of features, the model also boasts enhanced predictive capabilities, aiming to provide longer lead times for medical interventions.

The contribution of this research extends beyond the mere development of a sophisticated predictive model. Recognizing the potential bottlenecks in software-based solutions, particularly in situations demanding real-time predictions, this study ventured into the domain of hardware optimization. The feasibility of implementing the proposed model on Field-Programmable Gate Array (FPGA) was extensively investigated. FPGA, with their parallel processing capabilities and reconfigurable architectures, presents a promising avenue for enhancing the speed and efficiency of complex signal processing tasks. Such hardware-accelerated approaches (Cunha et al. 2022) are projected to not only achieve faster results but also ensure more consistent and reliable predictions, a critical

factor when dealing with life-threatening conditions like SCA. Through this multi-faceted approach that encompasses advanced machine learning algorithms, expansive feature engineering, and hardware optimization, this study posits a comprehensive solution designed to redefine the standards of early-stage SCA and SCD detection. These contributions catalyze further advancements in the field, ultimately elevating the standards of cardiac care and improving patient survival outcomes.

2 Classification algorithm of sudden cardiac arrest based on heart rate variability analysis

In cardiac healthcare, the accurate prediction and classification of SCA is crucial. This section elucidates the role of HRV analysis, focusing on its intricate assessment of consecutive heartbeats. The HRV analysis is segmented into time-domain and frequency-domain methods, with an additional exploration of its dynamical properties via the Poincaré plot (Nguyen et al. 2022; Rodriguez et al. 2017). The study's core objective is to differentiate ECG data related to SCA from normal sinus rhythm (NSR), utilizing datasets from the MIT-BIH PhysioNet database. Emphasis is placed on the R-R interval, representing the time between successive R-wave peaks in a standard ECG. A neural network model is then introduced to classify five-minute ECG segments for SCA or NSR identification. HRV encompasses variations in the temporal intervals between successive heartbeats, also termed inter-beat intervals (IBIs). Contrary to a metronomic heart, a healthy cardiac system exhibits complex, non-linear oscillations. HRV can be classified into three categories based on

duration; these are 24-h, short-term which spans 5 min, and ultra-short-term lasting less than 5 min. The parameters of HRV are segmented into time-domain, frequency-domain, and non-linear metrics, as elaborated in Table 1.

2.1 Time domain analysis

In the HRV analysis, time-domain indices are utilized to quantify the variability observed in interbeat intervals (IBI). This IBI is often referred to as the R-R interval within an ECG waveform, denoting the temporal gap between successive peaks of the R wave. Key time-domain metrics encompass including standard deviation of RR intervals (SDRR); root mean square of successive differences between R-R intervals (RMSSD); percentage of successive R-R intervals with differences exceeding 20 ms (pNN20); and percentage where such differences surpass 50 ms (pNN50) as

$$\text{SDRR} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (RR_i - \overline{RR})^2}, \quad (1)$$

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

$$\text{pNN20} = \frac{\text{NN20}}{\text{Total NN}} \times 100\%, \quad (3)$$

$$\text{pNN50} = \frac{\text{NN50}}{\text{Total NN}} \times 100\%, \quad (4)$$

where, N represents the total number of R-R intervals, RR_i denotes the individual R-R interval, and \overline{RR} is the mean of all R-R intervals. NN20 and NN50 refer to the count of

Table 1 HRV measures

Parameter	Unit	Description
Time-domain		
SDRR	ms	Standard deviation of RR intervals
RMSSD	ms	Root mean square of successive RR interval differences
pNN20	%	Percentage of successive RR intervals that differ by more than 20 ms
pNN50	%	Percentage of successive RR intervals that differ by more than 50 ms
Frequency-domain		
LF power	ms ²	Absolute power of the low-frequency band (0.04–0.15 Hz)
HF power	ms ²	Absolute power of the high-frequency band (0.15–0.4 Hz)
LF/HF	%	Ratio of LF-to-HF power
Non-linear measures		
SD1	ms	Poincaré plot standard deviation perpendicular the line of identity
SD2	ms	Poincaré plot standard deviation along the line of identity
SD1/SD2	%	Ratio of SD1-to-SD2

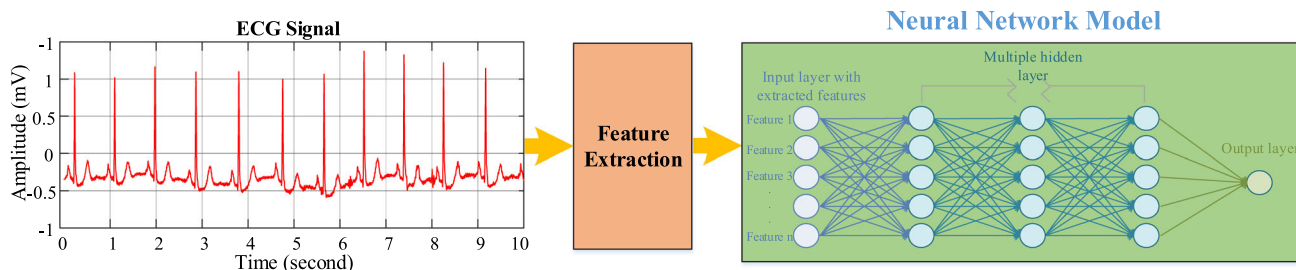


Fig. 2 Computational flow overview

successive R-R intervals with differences exceeding 20 ms and 50 ms, respectively.

2.2 Frequency domain analysis

Frequency-domain measurements are employed to estimate the distribution of absolute power across distinct frequency bands. The categorization of heart rate oscillations into two frequency bands has been delineated by the Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology (Camm et al. 1996), specifically into low-frequency bands ranging from 0.04 to 0.15 Hz and high-frequency bands ranging from 0.15 to 0.4 Hz. The term frequency power is used to describe the signal energy confined within a particular frequency band and can be denoted either in terms of absolute or relative power. Absolute power is computed as the square of milliseconds divided by cycles per second, expressed as ms^2/Hz . Relative power, expressed in normal units, is obtained by dividing the absolute power of a specific frequency band by the total absolute power aggregated across both low-frequency and high-frequency bands. The low-frequency band contains rhythms that exhibit periods between 7 and 25 s, while the high-frequency band comprises rhythms with periods between 2.5 and 7 s. Estimation of the power spectrum density is achieved through the Fast Fourier Transform-based Welch's periodogram method. This approach segments the time-series data into overlapping intervals, calculates the modified periodogram for each segment, and averages these power spectrum density estimates. Welch's method proves superior to traditional periodogram-based power spectrum density estimation techniques as it employs averaging across modified periodograms of segmented data rather than relying on a singular data record. In the application of Welch's method, data segments of length L are overlapped and periodograms are subsequently computed from these windowed segments. To account for the loss of signal energy due to the windowing procedure, periodograms are normalized by a factor U as

$$U = \frac{1}{L} \sum_{n=0}^{L-1} w^2(n), \quad (5)$$

where $w(n)$ represents the function applied to the data segments. Meanwhile, the power density spectral estimate generated by Welch's method is presented as

$$P_{WE} = \frac{1}{S} \sum_{i=0}^{L-1} P_i(f), \quad (6)$$

where $P_i(f)$ represents the individual periodogram of the i -th segment, and S denotes the normalization factor to ensure proper scaling of the power density estimate.

2.3 Nonlinear analysis

The Poincaré plot serves as a two-dimensional geometric approach for examining the dynamical properties of HRV. In this diagram, each successive R-R interval is graphed in relation to its preceding R-R interval. Each pair of adjacent R-R intervals forms a unique point within the plot. As a part of the non-linear HRV metrics, the Poincaré plot is quantitatively assessed through three key parameters including standard deviation 1 (SD1), standard deviation 2 (SD2), and the ratio of SD1 to SD2 (SD1/SD2). SD1 represents the standard deviation of points in the plot that is perpendicular to the line of identity and is mathematically defined as

$$SD1 = \sqrt{\frac{1}{2} SDSD^2}, \quad (7)$$

where $SDSD$ is the standard deviation of the points in the Poincaré plot that is perpendicular to the line of identity. SD2 is the standard deviation of points parallel to the line of identity as

$$SD2 = \sqrt{\frac{1}{2} SDRR^2 - \frac{1}{2} SDSD^2}, \quad (8)$$

where $SDRR$ is the standard deviation of the points in the Poincaré plot parallel to the line of identity. SD1 is a statistical metric used in time-series analysis quantifying the dispersion of data points along a specific axis, often

Table 2 Feature ranking

	SDRR	RMSSD	pNN20	pNN50	HF	LF	LF/HF	SD1	SD2	SD2/SD1
Mutual Info	0.208	0.407	0.213	0.223	0.361	0.361	0.362	0.408	0.202	0.490
Feature Ranking	8	3	7	6	5	5	4	2	9	1

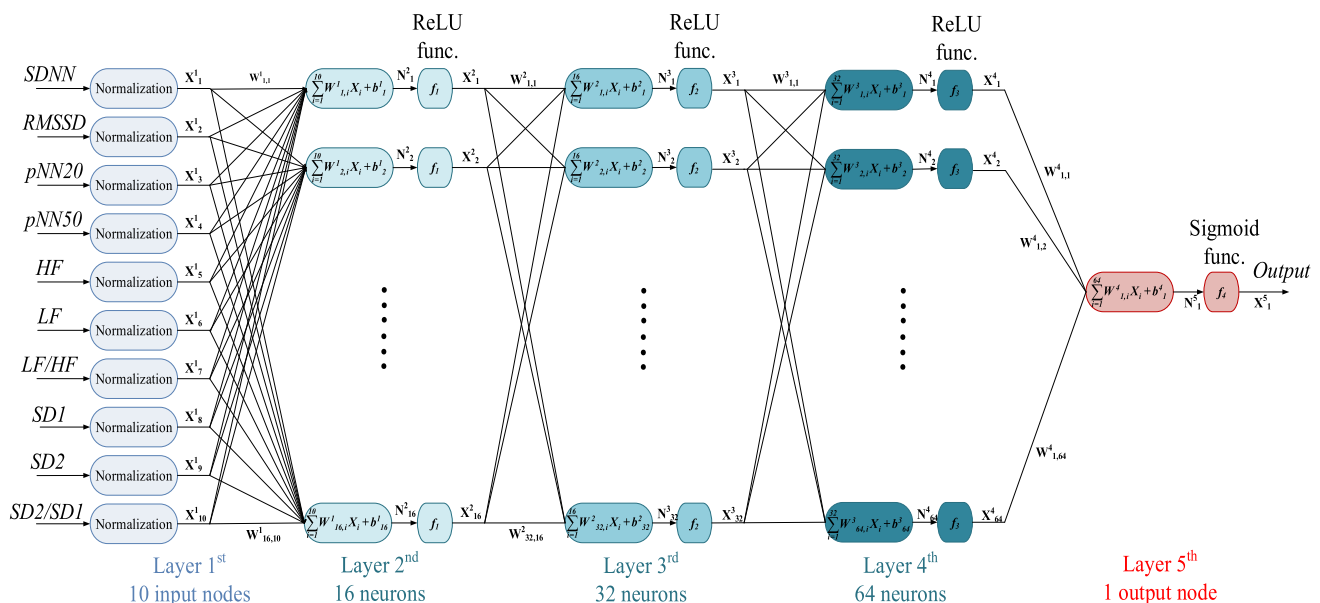
associated with short-term variability. In applications such as heart rate analysis, SD1 reveals rapid changes in the data, providing valuable insights into dynamic processes. Similarly, SD2 is another statistical measure used in time-series analysis. SD2 measures the dispersion of data points along a different axis, typically linked to long-term variability. In dynamic system modeling, SD2 characterizes the overall stability and complexity of the system, offering a perspective on its behavior over extended periods. Furthermore, the SD1/SD2 ratio is a derived parameter that provides a unique perspective on the interplay between short-term and long-term variability within a dataset. Researchers often use this ratio to gain insights into autonomic nervous system dynamics and employ it as a diagnostic tool across scientific disciplines. It offers a quantitative understanding of the balance between short-term and long-term fluctuations in data.

3 Data acquisition and methodology

This section provides an overview of the data acquisition and methodology employed in our study. As presented in Fig. 2, it summarizes the computational flow of the entire process, illustrating the sequential progression from ECG data acquisition through feature extraction to the structure of the proposed neural network model. This computational flow presents methodologies discussed in this section including data acquisition, feature extraction, and the architecture of the proposed neural network model.

3.1 Data acquisition

Data for ECG analysis were sourced from the MIT-BIH PhysioNet database. Specifically, two databases were utilized including the MIT-BIH Normal Sinus Rhythm Database for healthy control subjects (Goldberger, et al. 2000) and the Sudden Cardiac Death Holter Database for patients who have experienced sudden cardiac arrest (Greenwald 1986). The MIT-BIH Sudden Cardiac Death Holter Database consists of 23 comprehensive Holter

**Fig. 3** The structure of the proposed neural network model

recordings. Each recording represents a two-lead ECG with a minimum duration of four hours. The database encompasses 18 patients with an underlying sinus rhythm, one patient under continuous pacing, and four patients diagnosed with atrial fibrillation. Every patient in this set demonstrated sustained ventricular tachyarrhythmia, and 20 of them experienced cardiac arrest. However, it's important to highlight that only 18 recordings in the database contain complete ECG annotation. Due to this constraint, the analysis was restricted to these 18 patients, and they were classified as sudden cardiac death cases. From a demographic perspective, the database comprises nine males ranging from 34 to 80 years old, eight females aged between 30 and 89, and one individual with an undisclosed gender, aged 62. Figure 1(a) displays the ECG waveform associated with SCA resulting from ventricular tachycardia/fibrillation (VT/VF). The MIT-BIH Normal Sinus Rhythm Database encompasses 18 extended ECG recordings. These recordings cover subjects that include five males aged between 26 and 45, and 13 females aged between 20 and 50. Notably, none of these individuals showed any significant arrhythmias. For consistency in the analysis, 18 subjects were selected from each database, leading to a combined cohort of 36 subjects. Each database provides recordings that exceed one hour in duration for every individual. In this study, the focus was on ECG signals immediately preceding any cardiac arrest episodes, with Fig. 1(b) illustrating the typical ECG waveform of a normal sinus rhythm.

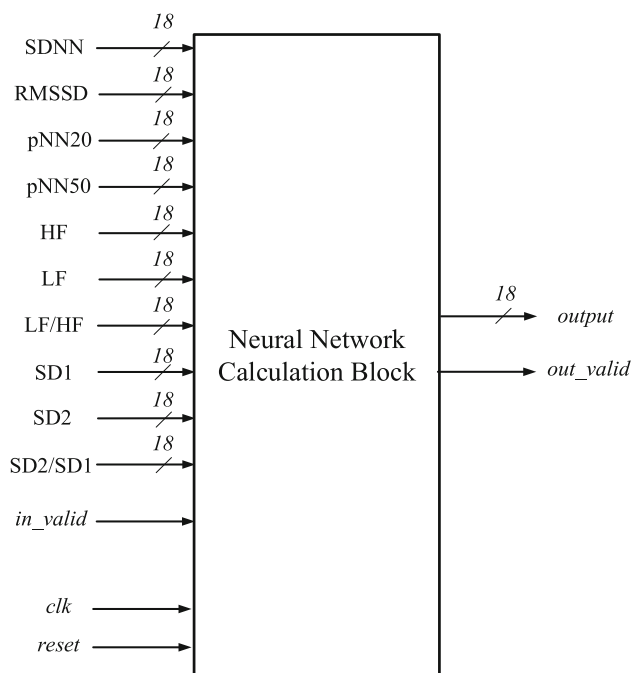


Fig. 4 The pin assignment of neural network hardware

3.2 Feature extraction

In line with standards on HRV previously published (Natali et al. 2017), ECG signals are required to have durations of at least two minutes for the extraction of high-frequency (HF) power and four minutes for low-frequency (LF) power in the frequency domain of HRV. As a result, 60-min ECG signals were partitioned into five-minute segments, with each segment having a one-minute overlap. This partitioning resulted in 56 five-minute ECG segments for each subject, leading to a cumulative total of 2016 segments for extraction. HRV represents the fluctuation of R-R intervals, with R-peaks annotated in both selected MIT-BIH databases. All HRV metrics were computed using Python 3.8. A set of nine features was extracted from each five-minute ECG segment, including standard deviation of R-R intervals (SDRR), root mean square of successive differences (RMSSD), proportion of adjacent intervals differing by more than 20 ms (pNN20), and proportion of adjacent intervals differing by more than 50 ms (pNN50) for time-domain measures. For frequency-domain measures, LF, HF, and LF/HF ratio are considered. Additionally, SD1, SD2, and the ratio SD1/SD2 are used for calculation of nonlinear measurements. Mutual information serves as a quantitative indicator of the informational correspondence between two variables. In this study, mutual information between HRV features and target variables was computed to ascertain correlations. Table 2 presents the calculated mutual information along with the feature ranking. Higher values denote greater interdependence between the HRV features and the target variables. A value of zero signifies independence between the HRV measure and the target variables.

3.3 Structure of proposed neural network model

A neural network with a greater number of hidden layers exhibits enhanced function-fitting capabilities, potentially leading to superior performance. However, excessive layer depth may introduce challenges such as overfitting and convergence issues. As delineated by Goodfellow et al. (2016), neural networks without hidden layers are restricted to linearly separable functions. Conversely, networks with more than two hidden layers can approximate any smooth mapping to arbitrary accuracy and represent complex decision boundaries. To address potential overfitting, our methodology incorporates leave-one-subject-out testing to ensure the model's generalizability to unseen data. As presented in Fig. 3, the proposed neural network model comprises one input layer, one output layer, and multiple hidden layers. Initially, ten heart rate variability (HRV) metrics, including SDNN, RMSSD, pNN20, and pNN50 in

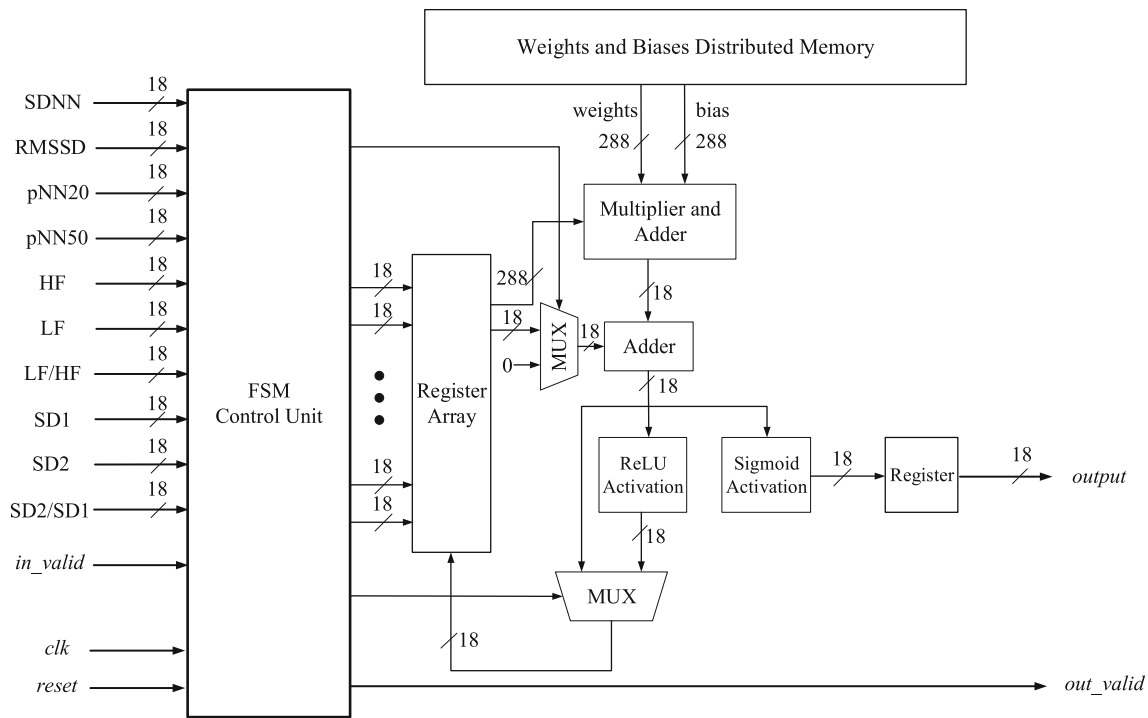


Fig. 5 The hardware architecture of proposed neural network model

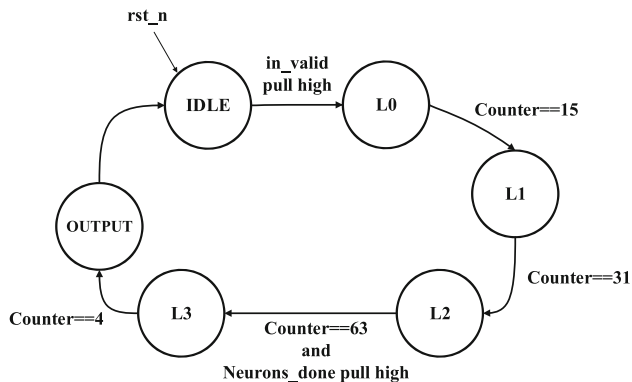


Fig. 6 The finite state machine of the hardware

the time domain, along with HF power and LF power in the frequency domain, as well as SD1, SD2, and the ratio SD2/SD1, are normalized before feeding into the network. The hidden layers consist of 16, 32, 64, and 128 neurons, respectively. Each neuron in the hidden layers employs the Rectified Linear Unit (ReLU) (Hara et al. 2015) as the activation function while binary cross-entropy serves as the loss function for this binary classification. Moreover, regularization technique is applied to impose constraints on the network parameters during training, preventing them from becoming reliant on the training data. Furthermore, dropout layers were incorporated into the network architecture, randomly deactivating a certain percentage of neurons during each training iteration, thus reducing the

likelihood of co-adaptation among neurons and promoting generalization to unseen data. The output node, equipped with the sigmoid activation function, generates a probability score indicative of SCA. Any score exceeding 0.5 is classified as SCA, while scores below this threshold are categorized as NSR. By designing the neural network architecture and incorporating leave-one-subject-out testing, the aim is to avoid overfitting while maximizing the model's predictive performance on unseen data.

3.4 Hardware implementation based on the proposed neural network model

3.4.1 Overall hardware design

To expedite computational efficiency of the classification algorithm a hardware-based feed-forward neural network is employed. Given that the Artificial Intelligence (AI) algorithm is intended for wearable devices, circuit area is prioritized over computational speed. Figure 4 delineates the FPGA pin assignment for hardware implementation. The input interface comprises control pins namely clock, reset, and in_valid signals as well as a range of Heart Rate Variability (HRV) metrics such as SDNN, RMSSD, pNN20, pNN50, HF, LF, LF/HF, SD1, SD2, and SD2/SD1 ratio. The output interface consists of neural network output and an out_valid signal. All HRV input metrics are represented as 18-bit fixed point numbers and data validity

Fig. 7 The logic implementation of single neuron multiplication and addition

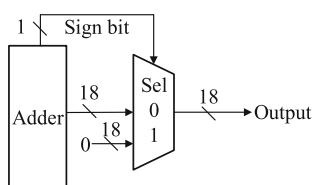
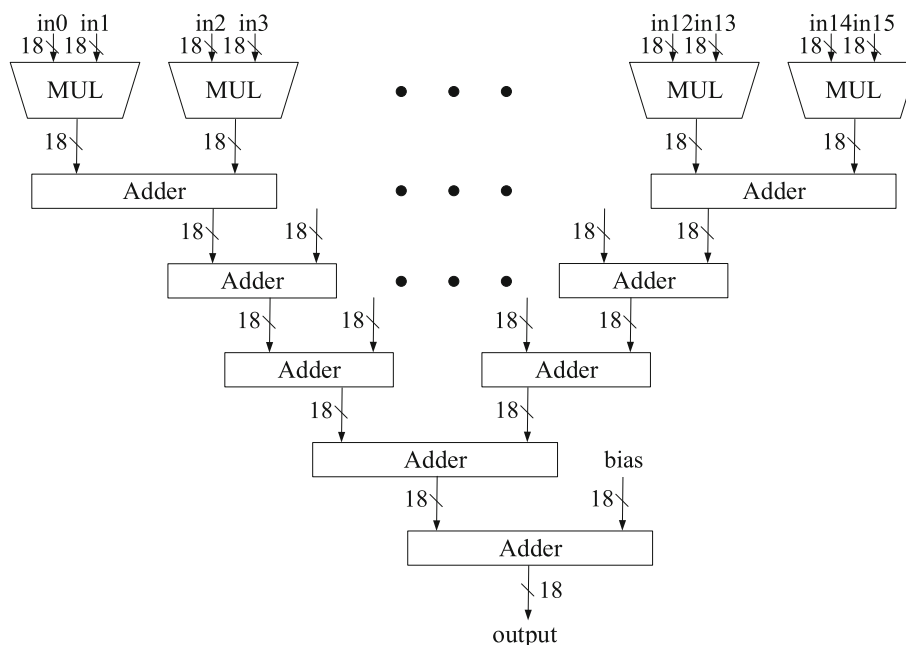


Fig. 8 The logic implementation of ReLU activation function

Table 3 Approximation of sigmoid activation function

x	$f(x)$
$0 \leq x < 1$	$0.25 x + 0.5$
$1 \leq x < 2.375$	$0.125 x + 0.625$
$2.375 \leq x < 5$	$0.03125 x + 0.84375$
$ x \geq 5$	1

is signaled when the *in_valid* pin is high. Pretrained weights and biases for the neural network are also 18-bit fixed point numbers stored within FPGA's distributed memory. The pretrained model used in this context is the sixth model obtained from cross validation, boasting an accuracy of 96.73 percent. Upon completion of the neural network's computation, the *out_valid* signal transitions to a high state, indicating that the output is ready for further processing. Figure 5 presents the hardware architecture of the proposed neural network model. The computational block for neural network calculations consists of an 18-bit multiplier and adder unit for the manipulation of 16 neurons, an 18-bit adder for aggregating the results of current

and preceding sets of 16 neurons, along with dedicated units for both ReLU and sigmoid activation functions. Specifically, the ReLU activation unit serves the hidden layers while the sigmoid activation unit is utilized in the output layer, constraining the output values to fall within the range of 0 to 1.

3.4.2 Finite state machine

The architecture of the finite state machine is illustrated in Fig. 6. Upon activation of the reset signal, the system transitions to the idle state (Tiwari and Tomko 2004; Alagar et al. 2011), awaiting the onset of the *in_valid* signal. Once *in_valid* is activated, the state machine progresses sequentially through L0, L1, L2, and L3 states, corresponding to computations in the first, second, third, and fourth dense layers, respectively. A counter variable is utilized to keep track of the number of neurons that have completed computations within each dense layer. In the first dense layer, which comprises 16 neurons, the counter increments from 0 to 15. Upon reaching a counter value of 15, the system proceeds to state L1. In the second dense layer, containing 32 neurons, the counter extends to 31 before transitioning to state L2. The third dense layer involves 64 neurons, prompting the counter to extend up to 63. Due to hardware limitations that restrict the parallel computation to 16 neurons, the third dense layer employs a two-cycle calculation for each neuron. Specifically, the hardware accumulates the outputs of the first 16 neurons in the preceding layer and subsequently adds these to the accumulated results of the remaining 16 neurons in the following cycle. Upon completion of calculations for each

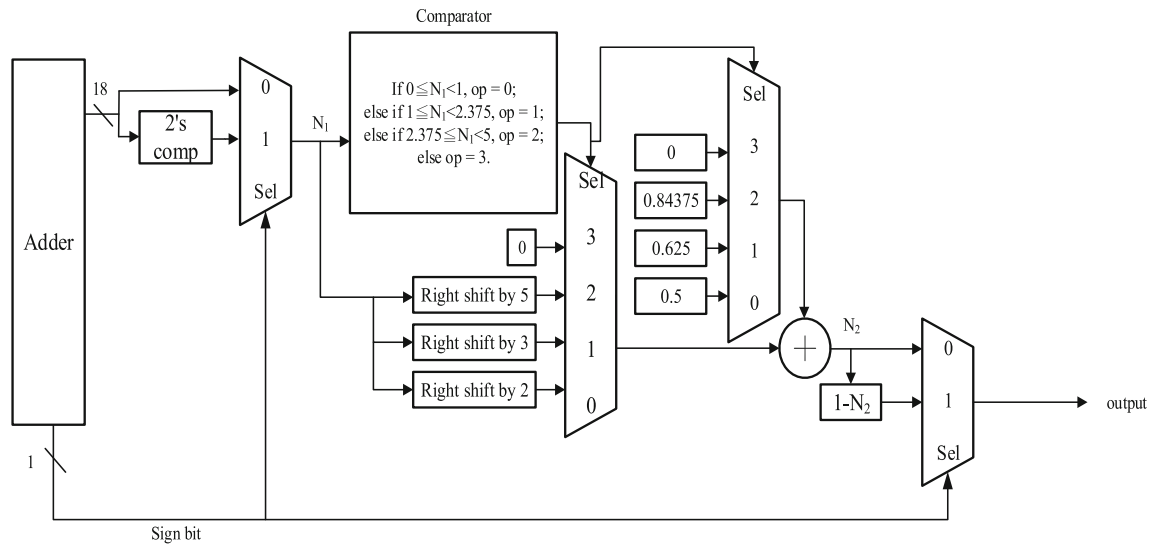


Fig. 9 The logic implementation of the sigmoid activation function approximation

Table 4 Performance of different layer numbers

MLP architecture	10–16–32–1%	10–16–32–64–1%	10–16–32–64–128–1%
Training accuracy	97.10	97.13	97.25
Training sensitivity	95.85	95.75	95.69
Training specificity	98.35	98.51	98.81
Testing accuracy	96.28	96.33	95.63
Testing sensitivity	95.04	94.84	92.96
Testing specificity	97.52	97.82	98.31

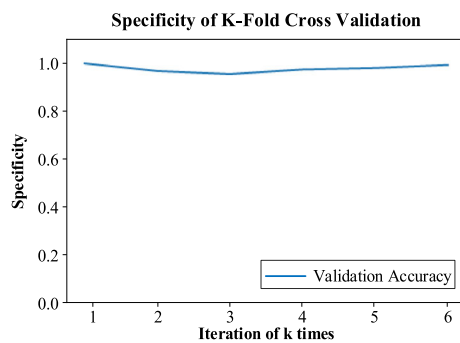


Fig. 10 6-fold cross validation performance of proposed neural network model.

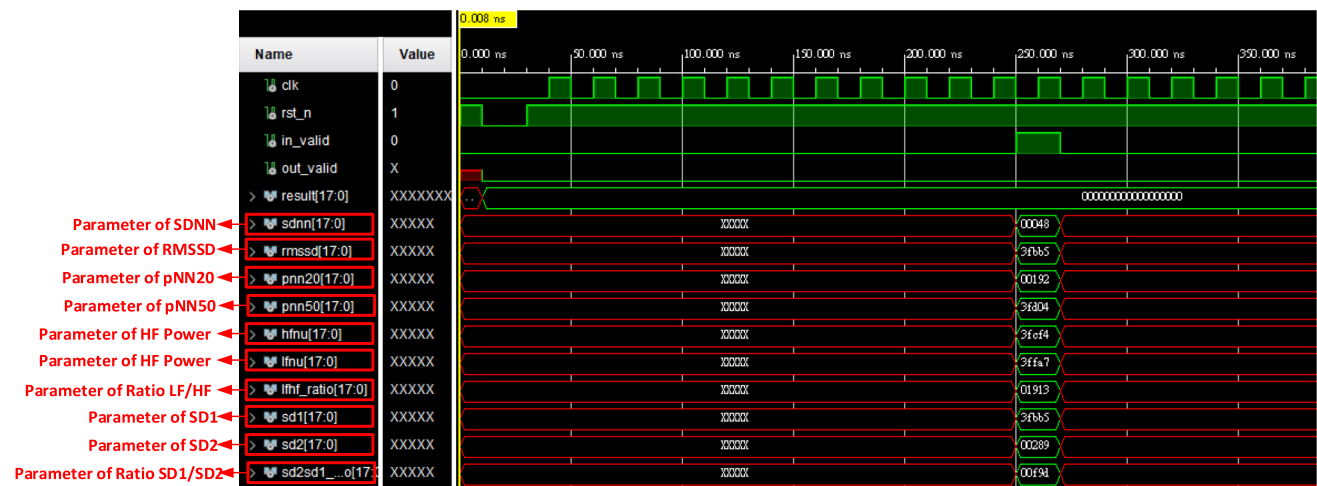
neuron, the *neurons_done* signal is activated. The state transitions to L3 when the counter reaches 63 and *neurons_done* is activated. The output layer contains a single neuron. Therefore, the 64 neurons in the third dense layer undergo computation in four cycles. Concurrently, the counter reaches a value of 4, prompting the system to transition to the OUTPUT state. This state facilitates the release of the neural network's computed result. Finally, the state machine reverts to the IDLE state after spending one cycle in the OUTPUT state.

Table 5 Performance of leave-one-subject-out

MLP architecture	Training accuracy %	Training sensitivity %	Training specificity %	Testing accuracy %	Testing sensitivity %	Testing specificity %
10–16–32–64–1	97.15	95.79	98.50	96.23	94.35	98.12

Table 6 Performance of proposed model comparison with published journals

	Loganathan et al. (2014)	Ebrahimzadeh et al. (2014)	Murugappan et al. (2015)	Fujita et al. (2016)	Khazaei et al. (2018)	Rabin Kaspal et al. (2021)	Jianli Yang et al. (2023)	Proposed
Features	5 min HRV (2 min before SCA)	1 min HRV (4 min before SCA)	1 min HRV (6 min before SCA)	1 min HRV (4 min before SCA)	1 min HRV (6 min before SCA)	5 min HRV (10 min before SCA)	70 min before SCA	55 min before SCA
Method	Support Vector Machine	Multilayer perceptron %	K-Nearest Neighbor	Support Vector Machine %	Decision Tree Classifier %	Hybrid Deep Learning (Convolutional Neural Network and Recurrence Complex Network)	Support Vector Machine	Multilayer perceptron %
Sensitivity	–	83.75	–	95.00	95.00	–	96.15%	94.35
Specificity	–	84.04	–	94.40	95.00	–	89.59%	98.12
Accuracy	96.36%	83.96	92.77%	94.70	95.00	93.24%	91.22%	96.23

**Fig. 11** The input waveform of RTL simulation

3.4.3 Hardware architecture of the multilayer perceptron unit

The multi-layer perceptron (MLP) unit (Khalil Alsmadi et al. 2009) within the dense layers consists of an 18-bit multiplier-adder unit configured for 16-neuron computation, as well as ReLU activation function. Following that, the 18-bit fixed-point multiplier constrains data width to 18 bits throughout the computational process. During multiplication, the higher-order 6 bits of the integer part are truncated, as are the lower 11 bits of the fractional part. This architectural design of the single-neuron multiplication and addition unit is depicted in Fig. 7. The adder

module processes two 18-bit fixed-point data inputs to produce a singular 18-bit fixed-point data output. The adder employs parallel computation to mitigate critical path delay, as opposed to a sequential approach. Figure 8 outlines the logic implementation of the ReLU activation function, which is designated for the hidden layers. The multiplexer (MUX) selects the adder's output when its most significant bit (MSB), also known as the sign bit, is zero. In contrast, a zero output is produced when the sign bit is one. Given that the sigmoid activation function in the output layer is based on an exponential function, it represents the most computationally intensive task within the hardware architecture. Accordingly, an approximation

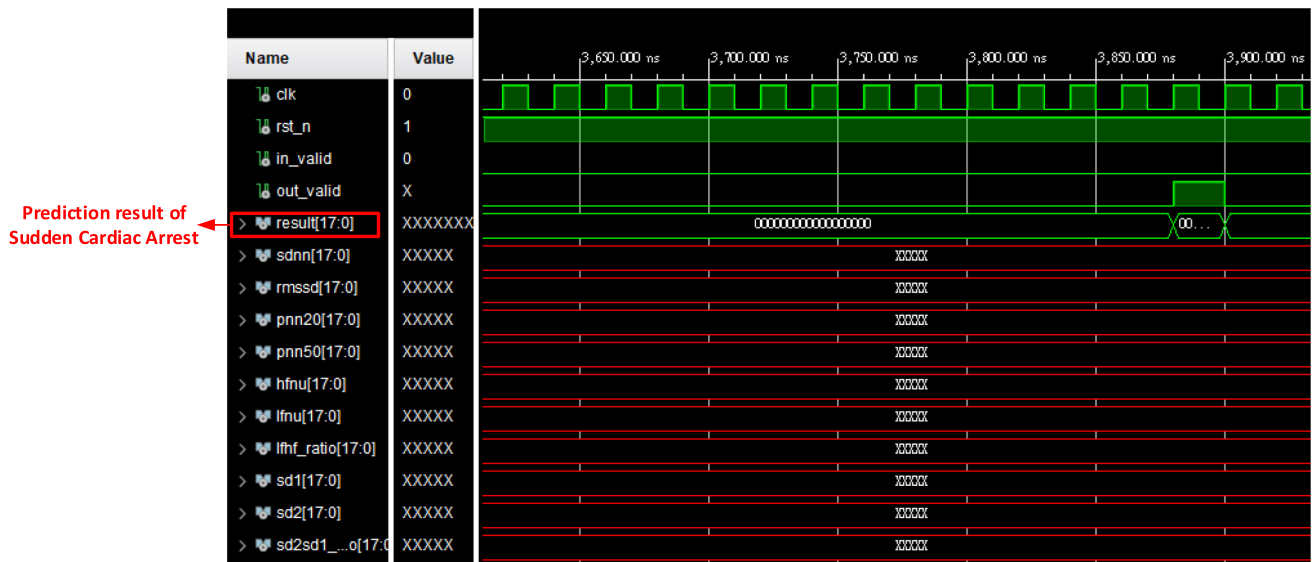


Fig. 12 The output waveform of RTL simulation

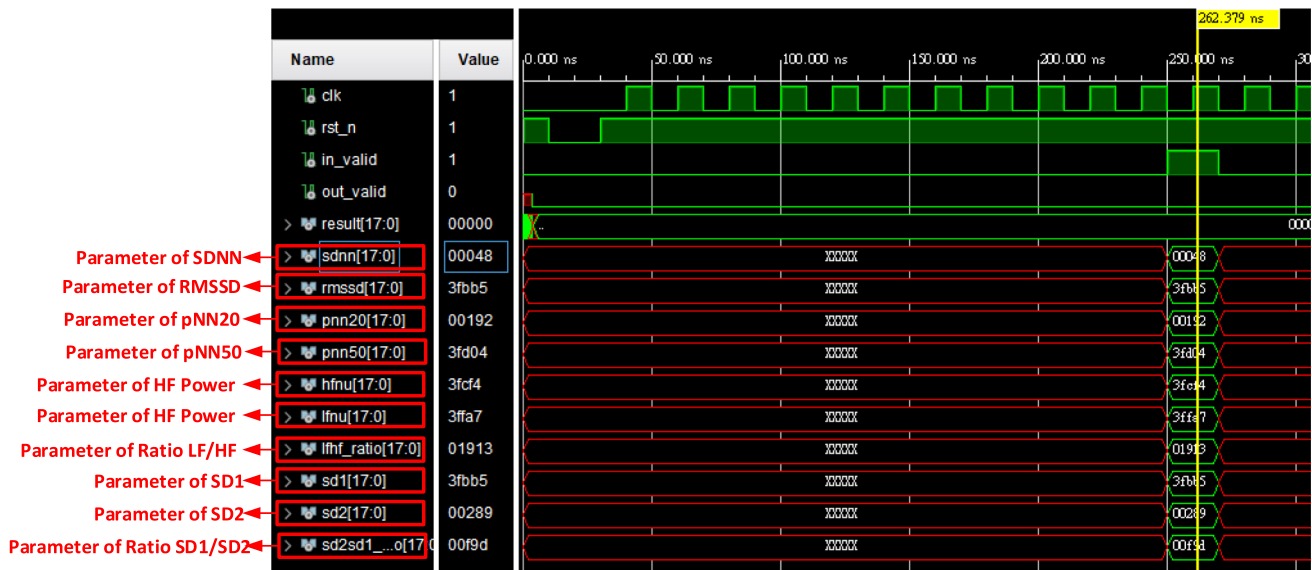


Fig. 13 The input waveform of FPGA implementation

method is employed to facilitate sigmoid function calculation, requiring merely a multiplication, a shift operation, and an addition. The mathematical description of this approximation method is detailed in Table 3, and the logic implementation of the sigmoid activation function approximation is illustrated in Fig. 9. For neurons in the upper layer, if the sign bit is one, the calculated result undergoes a two's complement operation prior to estimating the sigmoid output.

4 Experimental results

4.1 Performance evaluation

The K-fold cross-validation (Wong and Yeh 2019) serves as a robust technique to evaluate the performance of the model across multiple test datasets. Concerning data partitioning, the 36 subjects are segmented into six distinct groups, each comprising six subjects. Each group is balanced, containing three patients with sudden cardiac death and three subjects exhibiting normal sinus rhythm. Consequently, six training iterations are performed, and the average accuracy across these iterations serves as a

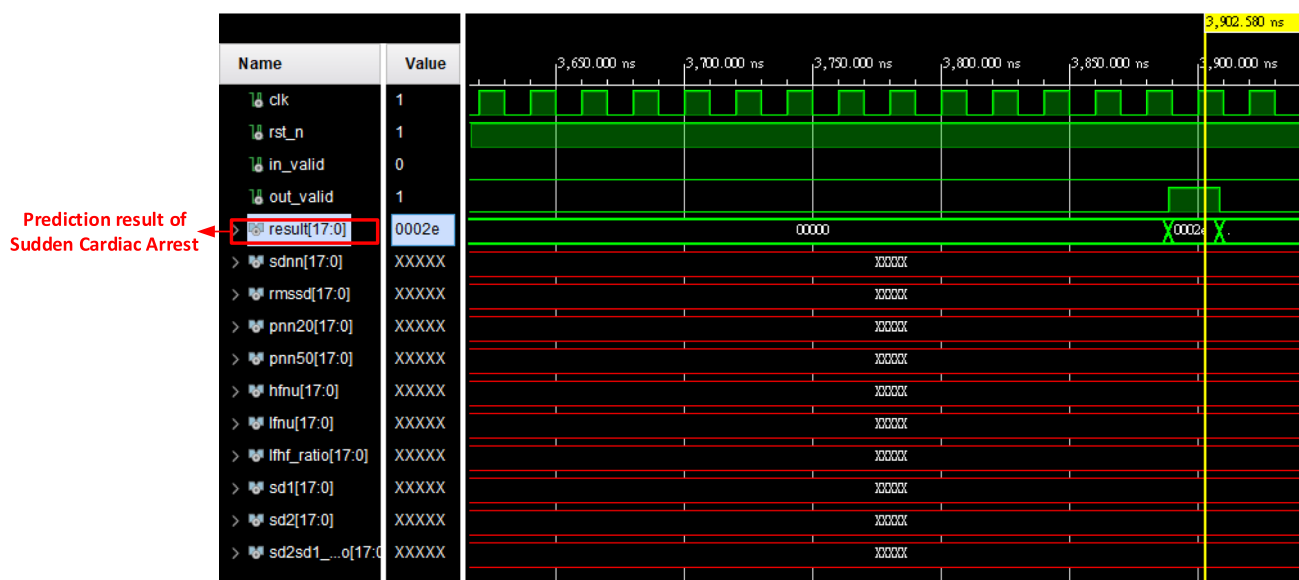


Fig. 14 The output waveform of FPGA implementation

Table 7 Performance of different precision bits and software

	Software %	16-bit %	18-bit %	20-bit %
Accuracy	96.73	95.54	96.43	96.73
Sensitivity	93.45	91.07	92.86	93.45
Specificity	100.00	100.00	100.00	100.00

Table 8 Synthesis results of different precision bits

	Clock Speed	LUT	Register	DSP
16-bit	61.728 MHz	3159	1575	16
18-bit	60.606 MHz	3593	1811	16
20-bit	59.880 MHz	4575	2000	16

Table 9 Implementation result

	Clock Speed	LUT	Register	DSP
Utilization	50 MHz	3561	1810	16
Utilization %	–	2.65	0.67	2.16

performance indicator for the model's quality by calculating the average performance. In addition to k-fold cross-validation, leave-one-subject-out cross-validation is employed to further ensure that subject-specific biases do not affect performance evaluation metrics.

4.2 Performance of neural network

The comparative performance across neural networks with varying numbers of layers is presented in Table 4. The training accuracy for the 4-layer neural network at 97.10%, lower than its 5-layer and 6-layer counterparts. Following that, the test accuracy for the 4-layer model is inferior to that of the 5-layer model. Although the 5-layer and 6-layer networks demonstrate near-identical training accuracies, the 5-layer network outperforms the 6-layer model in test accuracy. It is important to note that the 4-layer model exhibits underfitting patterns, while the 6-layer model tends to overfit. Consequently, the 5-layer model emerges as the most effective neural network configuration. Its architecture comprises one input layer with 10 nodes, three hidden layers with node counts of 16, 32, and 64, respectively, and a single-node output layer. The sensitivity, specificity, and accuracy averages for the training sets in the 5-layer model are 95.75%, 98.51%, and 97.13%, respectively. The corresponding averages for the test sets are 94.84%, 97.82%, and 96.33%. These metrics under sixfold cross-validation are depicted in Fig. 10. Performance under leave-one-subject-out cross-validation is summarized in Table 5, with marginal degradation in test accuracy compared to sixfold cross-validation without any overfitting.

Single-layer neural networks typically classify linearly separable data, while multi-layer architectures can model more intricate relationships. However, a surge in layers and neurons can lead to overfitting. This issue was addressed through empirical experimentation. The optimal neuron counts in hidden layers are observed to follow powers of 2, with neuron distributions in the layers such as 16–32–64,

Table 10 Performance of hardware comparison with other published journals

		Basterretxea et al. (2014)	Bahoura (2016)	Zhai et al. (2016)	Gaikwad et al. (2019)	Proposed
MLP Architecture		14–19–19–7	12–7–3	12–3–1	7–6–5	10–16–32–64–1
FPGA Platform		Spartan-6	Artix-7	Zynq-7000	Artix-7	Artix-7
Classification Time		800 ns	19,968 ns	540 ns	270 ns	3620 ns
Hardware Resources Utilization	LUT	1032	21,648	4032	3466	3547
	FF	774	13,330	2863	569	1810
	DSP	7	219	28	81	16
	BRAM	39	2	2	0	0

facilitating the reduction of resource allocation overhead on the hardware. The performance of this model compares favorably when contrasted with other studies utilizing the same database, as illustrated in Table 6. The adopted approach, leveraging 5-min HRV metrics as inputs, demonstrates superiority both in accuracy and early classification of SCA onset.

4.3 Hardware implementation based on pre-trained neural network model

Figure 11 illustrates the input waveform presented in Table 1 such as SDNN, RMSSD, pNN20, pNN50, HF, LF, LF/HF, SD1, SD2, and SD2/SD1 ratio, which associated with register-transfer level (RTL) (Pundir et al. 2022) simulation. All ten inputs are simultaneously provided when the *in_valid* signal transitions to a high state. After a span of 181 cycles, the output from the neural network (NN) is computed, prompting the *out_valid* signal to rise. The corresponding output waveform is depicted in Fig. 12. Upon completion of RTL simulation, the synthesis and implementation stages are executed using Xilinx Vivado 2020, targeting the Artix-7 FPGA. Figures 13 and 14 display these 10 inputs and output waveforms for the hardware implementation, which are consistent with the RTL simulation results. Experiments assessing model accuracy and synthesis outcomes across different bit configurations are summarized in Tables 7 and 8. During the synthesis phase, timing constraints are set such that both input and output delays constitute half of a clock cycle. Comparative evaluation reveals that the 16-bit hardware variant exhibits an accuracy approximately 1.2% inferior to its software-based counterpart. Conversely, the 20-bit hardware version matches the software in accuracy but incurs a hardware cost approximately 1.45 times greater than the 16-bit variant. Weighing these trade-offs, the 18-bit hardware emerges as the most balanced solution in terms of both cost and accuracy. The characteristics of this 18-bit

implementation on Artix-7 (xc7a200tfbg676-2) are presented in Table 9. Notably, the software latency is measured at 1.82756 ms, while the hardware implementation reduces this latency to 0.00362 ms for input-to-output transactions achieving effectively a speedup factor of approximately 500.

Table 10 presents a comparative analysis of FPGA performance alongside other contributions documented in the literature. A trade-off between computational speed and hardware area is observed. Given the emphasis on wearable device implementation, a reduced hardware footprint was prioritized over computational latency. The specifications of the implementation encompass a maximum operating frequency of 50 MHz, a classification latency of 3620 ns, usage of 3547 look-up tables (LUTs), 1810 registers, 16 digital signal processors (DSPs), and no block random access memory (BRAMs). Although the topology has more neurons than those in other published works, the model does not significantly exceed the hardware resources, especially in terms of DSP slices. Early identification of patients susceptible to SCD allows for timely preventive measures, thereby notably enhancing survival rates.

5 Conclusion

The approach taken in this study analyzes ECG waveforms 60 min before the onset of SCD, representing the earliest time-window examined to date. Through k-fold cross-validation, the proposed model exhibited superior performance, achieving 96.23% accuracy, 94.35% sensitivity, and 98.12% specificity via leave-one-subject-out evaluation. Additionally, the proposed neural network model was implemented in FPGA, accelerating the classification result calculations approximately 500 times faster than software-based methods. In subsequent studies, exploration based on additional databases may serve to refine the SCA classification model. Greater predictive insights might be obtained

by examining other ECG attributes associated with SCA, such as bradyarrhythmia. In terms of hardware advancements, real-time extraction of HRV in both time and frequency domains, as well as nonlinear measures, might be incorporated. Additionally, the hardware might be integrated into wearable devices for continuous heart rate monitoring, facilitating timely alerts for potential cardiac events.

Acknowledgements The authors appreciate the support from the Taiwan Semiconductor Research Institute (TSRI) for fabricating the IC. This work was also supported by National Science and Technology Council via 113-2640-E-A49-008, 113-2923-E-A49-006, 113-2223-E-A49-001, 112-2811-E-A49-508-MY2 and 112-2223-E-A49-006., and also by Hsinchu and Southern Taiwan Science Park Bureaus, Ministry of Science and Technology, Taiwan via 110CE-2-02 and 112AO28B.

Author contribution Sheng-Yueh Pan developed and implemented the neural network-based classification algorithm, leveraging Heart Rate Variability (HRV) metrics for categorizing patients into SCA and Normal Sinus Rhythm (NSR) cohorts. Duc Huy Nguyen contributed to the study by performing k-fold cross-validation to evaluate the algorithm's predictive accuracy, sensitivity, and specificity, as well as benchmarking the model's efficiency. Paul C.-P. Chao provided mentorship throughout, supervised the entire research process, and guided the hardware-level implementation and verification. All authors engaged in the review and editing of the manuscript.

Funding Not applicable.

Availability of data and materials No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare no conflict of interests.

Ethical approval Approved by the Institutional Review Board (IRB), National Taiwan University Hospital (NTUH), Taiwan numbered CS2-21096.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

Abubakar SM, Khan MR, Saadeh W, Altaf MAB (2018) "A wearable auto-patient adaptive ECG processor for shockable cardiac

- arrhythmia," In: 2018 IEEE Asian Solid-State Circuits Conference (A-SSCC): IEEE, pp. 267–268
- Alagar V, Periyasamy K, Alagar V, Periyasamy K (2011) "Extended finite state machine. Specification of software systems. Springer, London, pp 105–128
- Alonso-Atienza F, Morgado E, Fernandez-Martinez L, Garcia-Alberola A, Rojo-Alvarez JL (2013) Detection of life-threatening arrhythmias using feature selection and support vector machines. *IEEE Trans Biomed Eng* 61(3):832–840
- Bahoura M (2016) FPGA implementation of blue whale calls classifier using high-level programming tool. *Electronics* 5(1):8
- Basterretxea K, Echanobe J, Campo ID (2014) "A wearable human activity recognition system on a chip," In: Proceedings of the 2014 Conference on Design and Architectures for Signal and Image Processing, pp. 1–8, <https://doi.org/10.1109/DASIP.2014.7115600>
- Berwal D, Vandana C, Dewan S, Jiji C, Baghini MS (2019) Motion artifact removal in ambulatory ECG signal for heart rate variability analysis. *IEEE Sens J* 19(24):12432–12442
- Camm AJ et al (1996) Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. *Circulation* 93(5):1043–1065
- Cunha L, Roriz R, Pinto S, Gomes T (2022) Hardware-accelerated data decoding and reconstruction for automotive LiDAR sensors. *IEEE Trans Veh Technol* 72(4):4267–4276
- Daubert JP et al (2006) Predictive value of ventricular arrhythmia inducibility for subsequent ventricular tachycardia or ventricular fibrillation in Multicenter Automatic Defibrillator Implantation Trial (MADIT) II patients. *J Am Coll Cardiol* 47(1):98–107
- Ebrahimzadeh E, Pooyan M, Bijar A (2014) "A novel approach to predict sudden cardiac death (SCD) using nonlinear and time-frequency analyses from HRV signals," (in eng). *PLoS ONE* 9(2):e81896–e81896. <https://doi.org/10.1371/journal.pone.0081896>
- Fujita H et al (2016) Sudden cardiac death (SCD) prediction based on nonlinear heart rate variability features and SCD index. *Appl Soft Comput* 43:510–519. <https://doi.org/10.1016/j.asoc.2016.02.049>
- Gaikwad NB, Tiwari V, Keskar A, Shivaprakash NC (2019) Efficient FPGA implementation of multilayer perceptron for real-time human activity classification. *IEEE Access* 7:26696–26706. <https://doi.org/10.1109/ACCESS.2019.2900084>
- Goldberger AL et al (2000) PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* 101(23):e215–e220
- Goodfellow I, Bengio Y, Courville A (2016) Deep learning. MIT press
- Greenwald SD (1986) "The development and analysis of a ventricular fibrillation detector," Massachusetts Institute of Technology. Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology. [Online]. Available: <http://hdl.handle.net/1721.1/92988>
- Ha AC, Doumouras BS, Wang CN, Tranmer J, Lee DS (2022) Prediction of sudden cardiac arrest in the general population: review of traditional and emerging risk factors. *Can J Cardiol* 38(4):465–478
- Hara K, Saito D, Shouno H. (2015) "Analysis of function of rectified linear unit used in deep learning," In: 2015 international joint conference on neural networks (IJCNN),: IEEE, pp. 1–8
- Kaspar R, Alsadoon A, Prasad P, Al-Saiyd NA, Nguyen TQV, Pham DTH (2021) A novel approach for early prediction of sudden cardiac death (SCD) using hybrid deep learning. *Multim Tools Appl* 80:8063–8090

- Khalil Alsmadi M, Omar KB, Noah SA, Almarashdah I (2009) “Performance comparison of multi-layer perceptron (Back Propagation, Delta Rule and Perceptron) algorithms in neural networks,” in 2009 IEEE International Advance Computing Conference: IEEE, pp. 296–299.
- Khazaei M, Raeisi K, Goshvarpour A, Ahmadvadeh M (2018) Early detection of sudden cardiac death using nonlinear analysis of heart rate variability. *Biocybern Biomed Eng* 38(4):931–940. <https://doi.org/10.1016/j.bbe.2018.06.003>
- Lai D, Zhang Y, Zhang X, Su Y, Heyat MBB (2019) An automated strategy for early risk identification of sudden cardiac death by using machine learning approach on measurable arrhythmic risk markers. *IEEE Access* 7:94701–94716
- Loganathan LM, Murugappan M, Htut Y, Saravanan K (2014) Machine learning approach for sudden cardiac arrest prediction based on optimal heart rate variability features. *J Med Imaging Health Inform* 4:521. <https://doi.org/10.1166/jmihi.2014.1287>
- Murugappan M, Murugesan L, Omar I, Khatun S, Murugappan S (2015) Time domain features based sudden cardiac arrest prediction using machine learning algorithms. *J Med Imaging Health Inform* 5(6):1267–1271
- Murugappan M, Murugesan L, Jerritta S, Adeli H (2021) Sudden cardiac arrest (SCA) prediction using ECG morphological features. *Arab J Sci Eng* 46:947–961
- Natali JES, Starzynski PN, Chaui-Berlinck JG (2017) “Oscillatory patterns in heart rate variability and complexity: a meta-analysis,” (in English). *Biomed Signal Proces* 33:66–71. <https://doi.org/10.1016/j.bspc.2016.11.012>
- Nguyen DH, Chao PC-P, Chung C-C, Horng R-H, Choubey B (2022) Detecting atrial fibrillation in real time based on PPG via two CNNs for quality assessment and detection. *IEEE Sens J* 22(24):24102–24111
- Pundir N, Park J, Farahmandi F, Tehranipoor M (2022) “Power side-channel leakage assessment framework at register-transfer level. *IEEE Trans Very Large Scale Integr VLSI Syst* 30(9):1207–1218
- Rodriguez J, Voss A, Caminal P, Bayés-Genis A, Giraldo BF (2017) “Characterization and classification of patients with different levels of cardiac death risk by using Poincaré plot analysis,” In: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC): IEEE, pp. 1332–1335
- Shah MN (2006) The formation of the emergency medical services system. *Am J Public Health* 96(3):414–423
- Shen T-W, Shen H-P, Lin C-H, Ou Y-L (2007) “Detection and prediction of sudden cardiac death (SCD) for personal health-care,” In: 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, pp. 2575–2578
- Suboh M, Jaafar R, Nayan N, Harun N (2019) “ECG-based detection and prediction models of sudden cardiac death: current performances and new perspectives on signal processing techniques. *Int J Onl Eng*. <https://doi.org/10.3991/ijoe.v15i15.11688>
- Tiwari A, Tomko KA (2004) Saving power by mapping finite-state machines into embedded memory blocks in FPGAs. *Proceed Design, Autom Test Eur Conf Exhib* 2:916–921
- Wong T-T, Yeh P-Y (2019) Reliable accuracy estimates from k-fold cross validation. *IEEE Trans Knowl Data Eng* 32(8):1586–1594
- Yang J, Sun Z, Zhu W, Xiong P, Du H, Liu X (2023) Intelligent prediction of sudden cardiac death based on multi-domain feature fusion of heart rate variability signals. *EURASIP J Adv Signal Process* 2023(1):32
- Zhai X, Ali AAS, Amira A, Bensaali F (2016) MLP neural network based gas classification system on Zynq SoC. *IEEE Access* 4:8138–8146. <https://doi.org/10.1109/ACCESS.2016.2619181>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.