



Automated ECG multi-class classification system based on combining deep learning features with HRV and ECG measures

Ahmed S. Eltrass¹ · Mazhar B. Tayel¹ · Abeer I. Ammar¹

Received: 29 November 2020 / Accepted: 4 January 2022 / Published online: 29 January 2022

© The Author(s) 2022

Abstract

Electrocardiogram (ECG) serves as the gold standard for noninvasive diagnosis of several types of heart disorders. In this study, a novel hybrid approach of deep neural network combined with linear and nonlinear features extracted from ECG and heart rate variability (HRV) is proposed for ECG multi-class classification. The proposed system enhances the ECG diagnosis performance by combining optimized deep learning features with an effective aggregation of ECG features and HRV measures using chaos theory and fragmentation analysis. The constant-Q non-stationary Gabor transform technique is employed to convert the 1-D ECG signal into 2-D image which is sent to a pre-trained convolutional neural network structure, called AlexNet. The pair-wise feature proximity algorithm is employed to select the optimal features from the AlexNet output feature vector to be concatenated with the ECG and HRV measures. The concatenated features are sent to different types of classifiers to distinguish three distinct subjects, namely congestive heart failure, arrhythmia, and normal sinus rhythm (NSR). The results reveal that the linear discriminant analysis classifier has the highest accuracy compared to the other classifiers. The proposed system is investigated with real ECG data taken from well-known databases, and the experimental results show that the proposed diagnosis system outperforms other recent state-of-the-art systems in terms of accuracy 98.75%, specificity 99.00%, sensitivity of 98.18%, and computational time 0.15 s. This demonstrates that the proposed system can be used to assist cardiologists in enhancing the accuracy of ECG diagnosis in real-time clinical setting.

Keywords Convolutional Neural Network (CNN) · Deep Learning (DL) · Electrocardiogram (ECG) · And Heart Rate Variability (HRV)

1 Introduction

As stated by the World Health Organization (WHO), heart diseases are responsible of about 31% of deaths worldwide [1]. The electrocardiogram (ECG) is a noninvasive test for monitoring the heart function by detecting the electrical activity of heart muscles. It provides cardiologists with all needed information about heart conditions, and therefore, ECG represents an efficient tool for identifying various cardiac disorders. Congestive heart failure (CHF) is a serious cardiac disorder and a major contributor to global mortality rates. In CHF, the heart becomes unable to pump

sufficiently to maintain blood flow and supply the body tissue's needs for oxygen and metabolism. More than 26 million adults suffer from CHF all over the world, with an increase rate of 3.6 million every year [2]. However, early detection of CHF makes a significant contribution in improving treatment choices and impeding the CHF progression. Another important heart disorder responsible for several cases of sudden death is the arrhythmia (ARR). ARR refers to abnormal heart rhythm resulted from irregular heart rate.

To provide an effective and accurate identification of ARR and CHF, careful and uniform assessment via cardiologists is necessary which is hard and time-consuming. Therefore, a fully automated diagnosis system is urgently needed for accurate identification of heart diseases. Development of diagnosis systems can assist cardiologists in making accurate and expeditious diagnosis of ECG

✉ Ahmed S. Eltrass
ahmed.eltrass@alexu.edu.eg

¹ Electrical Engineering Department, Faculty of Engineering, Alexandria University, Alexandria, Egypt

recordings as well as reducing the consumed time and cost associated with the clinical interpretation. Over the last decades, several diagnosis systems based on machine learning (ML) methods were proposed for distinguishing distinct heart disorders [3, 4]. Recently, deep learning (DL) has become one of the most essential subfields of ML in signal processing studies especially in ECG interpretation. DL architecture is structured to learn and pick-up distinct characteristics automatically and then optimize the model weights and gradients by back propagation algorithm [5, 6]. There are various DL structures, including convolutional neural networks (CNNs) and deep belief networks (DBNs) [7, 8].

Heart rate variability (HRV) is one of the most important tools to evaluate the overall cardiac health. It reflects the cardiac system ability to adapt with any changes in internal and external stimuli. HRV time series is the variation in period time between consecutive heartbeats (RR intervals) [9]. Chaos theory allows better comprehension of heart rate dynamics, taking that the electrical activity of human heart is not a perfect oscillator, but it is slightly irregular and can be considered as a chaotic system. According to the chaos theory, chaos exists when there is a substantial dependence on the initial conditions [10]. Several studies reported that certain cardiac ARRs are instances of chaos [11]. In [12], several tests were applied on ECG signal, and the results showed the existence of deterministic chaos and nonlinear dynamics in the ECG signal. In [13], a nonlinear prediction algorithm was proposed to investigate the predictability and sensitivity to initial conditions of ECG records during atrial fibrillation, and the results confirmed the chaotic ventricular response in atrial fibrillation. Two important chaotic system parameters, namely the Lyapunov exponents (LEs) [14] and correlation dimension (CD) [15], are utilized to model the chaotic nature of ECG signals.

Several ML studies revealed that statistical, geometrical, spectral, and nonlinear measures of HRV are very effective for CHF diagnosis [16–26] and ARR discrimination [27–29]. In [18], the standard time features are combined with Renyi entropy exponents to classify CHF with better accuracy than using only time domain features. The classification results with this approach are 80% sensitivity, 94.4% specificity, and 87.9% accuracy using KNN classifier. A multistage risk evaluation approach for CHF detection based on short-term HRV dynamic measures [19] was proposed, and the experimental results showed an accuracy of 96.61% using decision-tree-based support vector machine (SVM) classifier. In [21], HRV fuzzy and permutation entropies are extracted and sent to the least squares SVM classifier for CHF diagnosis with an accuracy of 98.21%. A new technique based on combining morphological and statistical features of individual heartbeats

was proposed in [28] for classification of cardiac ARRs. In [30], the fragmentation indices are combined with traditional linear and nonlinear analysis of HRV to enhance the classification performance of cardiac diseases.

Despite the high ECG diagnosis results of HRV analysis used in most of the feature-based ML techniques, their robustness is not guaranteed. This is because the most important HRV features are inevitably affected by spontaneous fluctuations, respiration, drug interferences, age, and gender. As a result, HRV analysis should not form the main backbone for cardiac diseases assessment. These drawbacks can be avoided by developing new DL techniques for cardiac diseases diagnosis [31–38]. Note that training a deep CNN from scratch is hard because it requires a huge amount of labeled training ECG data and it may suffer from the problem of overfitting when using small dataset [39]. Transfer learning represents an adaptable solution to avoid this issue as it allows utilizing existing neural networks (NNs) that were trained with a huge amount of data and move this knowledge to the targeted classification system [40].

Different pre-trained CNNs [41–43] were trained on a huge dataset, namely the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). Various ML techniques and pre-trained CNN architectures were investigated for medical imaging classification [44–46]. In contrast, few DL models have reported robust performance for ECG classification of cardiac diseases [47–51]. A pre-trained CNN model with a distance distribution matrix in entropy calculation was proposed in [48] for CHF diagnosis with an accuracy of 81.9% and sensitivity of 80.99%. In [49], a DL model based on simple convolutional units and time–frequency characteristics was investigated to classify CHF and ARR with an accuracy of 93.75%. In [51], an ensemble approach based on short-term HRV data and deep NN was developed for CHF detection.

The majority of exiting ECG classification systems report the reasonable classification results of distinguishing CHF and NSR cases. However, it is still very difficult to build up an efficient automated framework that can distinguish accurately CHF, ARR, and NSR cases, while running in real time with low hardware complexity. The aim of this paper is to develop a novel automatic ECG multi-class classification framework for differentiating CHF, NSR, and ARR cases with high classification performance and low hardware complexity. The proposed framework is based on combining features from three parallel channels: (1) HRV features representing the fundamental variations between CHF, ARR, and NSR, (2) optimal selected DL features from a pre-trained CNN model, AlexNet, allowing to capture more subtle differences between the three classes, and (3) dynamic features of ECG signals using chaos theory. The concatenated

features are sent to different ML classifiers to distinguish CHF, ARR, and NSR cases.

The main contributions of this work can be summarized as follows: (1) A new hybrid real-time automated system is proposed for distinguishing CHF, ARR, and NSR cases with high classification performance and low computational burden. (2) The proposed framework is based on a new signal transformation technique, namely constant-Q non-stationary Gabor transform (CQ-NSGT), for extracting efficiently the changes in the ECG signal and transforming it into 2-D image to be passed to the AlexNet structure. (3) The proposed system investigates the pair-wise feature proximity (PWFP) feature reduction technique for selecting the optimal subtle DL features and combining them with both HRV measures and ECG features representing the fundamental differences between CHF, ARR, and NSR classes. (4) The computational efficiency of the proposed structure is enhanced over the end-to-end AlexNet CNN architecture by replacing the MLP classification layer with ML classifier to differentiate the three classes. (5) Different pre-trained CNNs, including VGGNet16, VGGNet19, ResNet50, ResNet101, Inceptionv3, and DenseNet are examined by replacing the AlexNet model and comparing the classification results using the same techniques for all other stages of the proposed framework. The performance of the proposed system is examined using several evaluation metrics, including accuracy, sensitivity, precision, specificity, and computational time using fivefold cross-validation. Extensive experiments have been performed to examine the performance of the proposed framework, and the experimental results reveal that the proposed approach outperforms other exiting state-of-the-art diagnosis systems. This demonstrates that the proposed framework represents an efficient tool to assist cardiologists in real-time ECG multi-class classification.

The remaining parts of this paper are organized as follows. Section 1 explains the structure of the proposed ECG diagnosis system, including the extracted features from the AlexNet CNN architecture, HRV, and ECG. Then, the feature selection algorithm and the classification technique are presented. In Sect. 3, the experimental results are investigated in terms of several evaluation metrics for different ML classifiers. In Sect. 4, the results are discussed and compared with other recent state-of-the-art systems.

2 Methodology

The proposed ECG multi-class classification system consists of three stages. The first stage includes the ECG data acquisition, preprocessing, segmentation, peak detection, and signal-to-image transformation. The second stage is composed of feature extraction, feature selection, and

feature concatenation. Finally, the classification is employed in the third stage. The structure of the proposed approach is illustrated in Fig. 1. The work concerning each part is explained in detail in the following subsections.

2.1 ECG data acquisition and preprocessing

To examine the effectiveness of the proposed approach, 162 ECG records of three different cases (ARR, CHF, and NSR) were retrieved from the following public databases: Massachusetts Institute of Technology–Beth Israel Deaconess Medical Center (MIT-BIH) ARR [52, 53], MIT-BIH NSR [52], and Beth Israel Deaconess Medical Centre (BIDMC) CHF [52, 54]. All the 162 ECG records were measured from lead II and VI, and they are divided into 96 ARR, 30 CHF, and 36 NSR records. All retrieved records were analyzed and labeled by several cardiologists. All ECG signals were resampled to a fixed sampling frequency of 128 Hz and normalized to eliminate the offset effect.

Note that ECG signals are frequently contaminated with several artifacts and noise sources which affect the diagnosis efficiency [55, 56]. In this work, an automated approach based on adaptive filtering [57] is employed to suppress all unwanted artifact components from the input raw ECG data, while keeping all essential characteristics of the ECG signal. After eliminating all unwanted noises and artifacts, each filtered ECG record is partitioned into 6 segments each with a particular length of 10,000 samples (78 s), yielding a total number of 972 ECG segments. The whole 972 segments are utilized in the training, validation, and testing processes [58]. Figure 2 illustrates three ECG records for ARR, CHF, and NSR subjects. Comparing Fig. 2a and c, the difference between irregularity in RR intervals for ARR subject and regular rhythm in the NSR case can be clearly observed. Figure 2b shows the small amplitudes of the R-peaks demonstrating the low pumping power resulted from CHF in comparison with the normal amplitudes of R-peaks for the NSR subject shown in Fig. 2c.

2.2 QRS detection and ECG transformation

Three copies of the denoised ECG segments are sent to the second stage which is composed of three parallel paths. The first path delivers a copy of ECG segments directly to the feature extraction stage. In the second path, the ECG segments are processed by a Pan–Tompkins QRS detector [59] to extract the HRV measures. The Pan–Tompkins algorithm is employed to detect the QRS complexes through three main steps: highlighting the frequency content of the QRS complexes with a band-pass filter, passing the filtered ECG signal through a differentiator to search the highest slopes that differentiate the QRS complexes

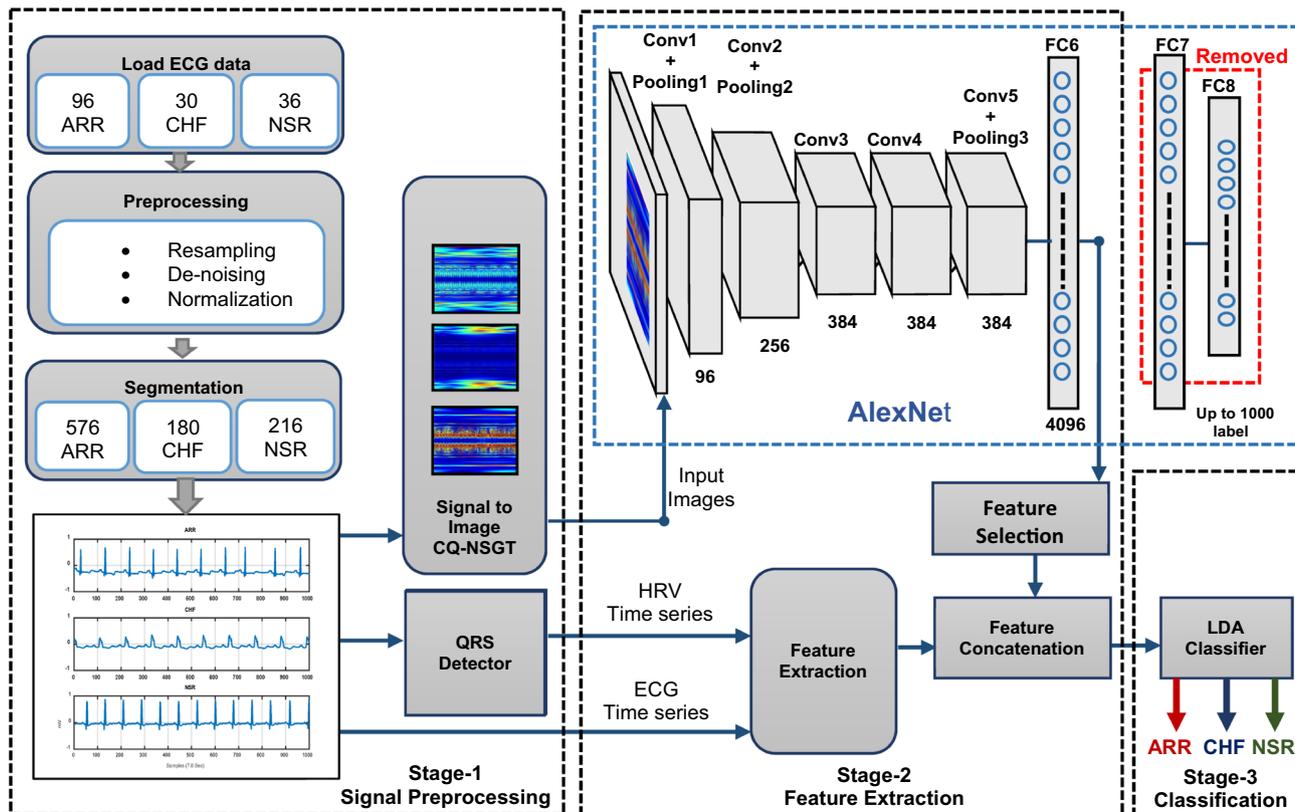


Fig. 1 The architecture of the proposed ECG multi-class diagnosis system

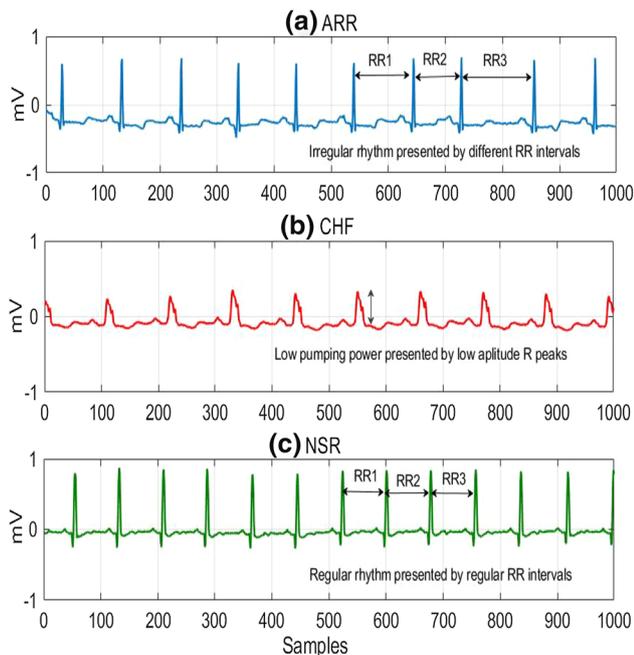


Fig. 2 ECG signals each of length 1000 samples for three subjects: a ARR; b CHF; and (NSR)

from other ECG components, and finally amplifying the QRS contribution and smoothing the ECG peaks. In the

third path, a copy of the denoised ECG segments should be converted into images to be fed as input to the proposed CNN architecture.

In the current study, the CQ-NSGT technique [60–62] is proposed to extract the changes in the ECG signal and transform it into time–frequency representation. Conventional algorithms, including short-time Fourier transform (STFT) and continuous wavelet transform (CWT) [63–66] suffer from the fixed frequency resolution over the entire operating frequency and the high computational time, respectively. In contrast, the CQ-NSGT algorithm provides a variable frequency resolution which is well suited for capturing the ECG dynamic variations and frequency changes over time. Also, the CQ-NSGT algorithm allows tracking the non-stationarity of ECG signal containing low- and high-frequency components, while requiring low computational complexity. Moreover, the CQ-NSGT algorithm overcomes the main drawbacks of traditional constant-Q transforms [62], specifically invertibility issues and computational burden problems. All mathematical analyses of the CQ-NSGT algorithm can be found in our work of [60].

Figure 3 illustrates the images of ARR, CHF, and NSR subjects. The transformed 2-D time–frequency representation illustrates the ECG signal strength (normalized

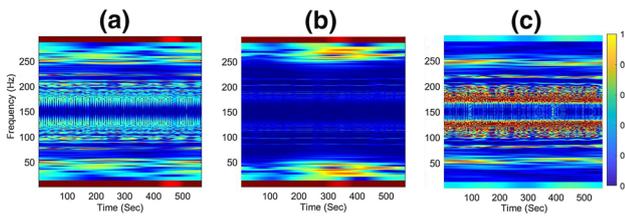


Fig. 3 Time–frequency representations for **a** ARR, **b** CHF, and **c** NSR using the CQ-NSGT algorithm

amplitude), represented by the color bar, over time at distinct frequencies. The difference between the three cases is determined by comparing the normalized amplitude values over time at distinct frequencies. Figure 3a shows the irregular heart rate of the ARR subject, while Fig. 3c illustrates the normal variation in both time and frequency of the NSR case. Figure 3b demonstrates the low heart’s pumping power of the CHF case, represented by the low amplitude over most of the frequency range.

2.3 Feature extraction

Feature extraction is the most important phase in the learning process. If the feature quality is low, it may lead to low performance even with using a powerful classification algorithm. Therefore, a primary goal of this work is to achieve an effective hybrid combination of different important features. The proposed methodology concatenates three groups of features. The first group of features, including standard linear and nonlinear HRV measures and fragmentation indices, is extracted to capture the fundamental differences between ARR, CHF, and NSR cases. The subtle variations of ECG signals are captured in the second group of features by selecting the optimal DL features extracted from the pre-trained AlexNet CNN model. Finally, the third group of features is extracted directly from the ECG segments to track the ECG dynamic changes using chaos theory. Other features requiring high computational cost are avoided in the current work in order not to affect the computational efficiency of the proposed diagnosis system.

2.3.1 Linear and nonlinear HRV measures

Linear HRV analysis includes several measurements of successive RR interval variations in both time and frequency domains. Time domain measures are the simplest indices for HRV analysis by representing the statistical variability in the RR time series. In this work, the mean, variance, and standard deviation of RR intervals are calculated to measure the amount of dispersion in the RR interval relative to its mean. Also, the higher-order statistics such as kurtosis (Ku) and skewness (Sk) are obtained to

assess the probability distribution of each HRV series. Ku indicates whether the RR intervals data is flat or peaked distribution in comparison with the normal distribution, while Sk measures the tails asymmetry of the RR data distribution. Table 1 shows the HRV time domain measures for M successive RR intervals with interval length d_j and index $j = 1, 2, 3, \dots, M$.

Frequency domain measures of HRV, extracted from the HRV power spectral density (PSD), are very important for evaluating the risk of different cardiac diseases [67]. Frequency domain analysis provides the power distribution of RR intervals over the frequency. The PSD of HRV shows two different peaks: one in the low-frequency (LF) band (0.04–0.15 Hz) and another one in the high-frequency (HF) band (0.15–0.4 Hz). Table 1 shows the total power in all bands, the power in LF and HF bands, and all other frequency domain measures.

Due to the non-stationary nature of ECG signals, conventional linear methods of signal analysis are not suitable for robust feature extraction and accurate ECG diagnosis. Linear HRV analysis cannot represent accurately the complex dynamics of cardiovascular system. Combining both linear and nonlinear HRV measures allows enhancing the classification performance. The most important nonlinear HRV measures are fractal scaling exponent α measured by detrended fluctuation analysis (DFA), approximate entropy ($ApEn$), Shannon entropy ($ShanEn$), and Poincaré descriptors [68–71].

α is calculated using DFA, in which the input RR interval time series, with N samples, is first integrated, and then the integrated time series is segmented into windows of equal length n . In each window, a least squares line $y_n(k)$ is fit to the data representing the trend in that window. The integrated time series $y(k)$ is then detrended by subtracting it from the local trend $y_n(k)$ in each window. The root mean square fluctuation of the detrended time series, $F(n)$, is calculated as follows: [68]

$$F(n) = \sqrt{\frac{1}{N} \sum_{k=1}^N [y(k) - y_n(k)]^2} \tag{1}$$

The calculation is repeated over all windows to investigate the relationship between $F(n)$ and the window size n . $F(n)$ always increases with the window size increase. This relationship is mapped onto a log–log plot to indicate the presence of power law (fractal) scaling. α is the slope of the line relating $\log F(n)$ to $\log(n)$ which characterizes the fluctuations.

$ApEn$ is a statistical metric for quantifying the regularity and predictability in data series without any a priori knowledge about the system generating them [69]. In the current study, $ApEn$ represents a model-independent measure for evaluating both dominant and subordinant patterns

Table 1 Linear HRV time-domain and frequency domain measures

Measure	Description	Equation
<i>Statistical time domain measures</i>		
μ	Mean of RR intervals time series	$\frac{1}{M} \sum_{j=1}^M d_j$
σ^2	Variance of the length of RR intervals time series	$\frac{1}{M} \sum_{j=1}^M (d_j - \mu)^2$
σ	Standard deviation of the length of RR intervals time series	$\sqrt{\sigma^2}$
Sk	Skewness of the probability distribution of RR intervals time series	$\frac{1}{M} \sum_{j=1}^M \left[\frac{d_j - \mu}{\sigma} \right]^3$
Ku	Kurtosis of the probability distribution of RR intervals time series	$\frac{1}{M} \sum_{j=1}^M \left[\frac{d_j - \mu}{\sigma} \right]^4$
<i>Frequency domain measures</i>		
TP	Total power in all bands	
$VLF P$	Power in the Very Low-Frequency (VLF) band (0.003–0.04 Hz): influence of thermoregulation on HRV	
$LF P$	Power in the LF band (0.04–0.15 Hz): marker for sympathetic nerve stimulation	
$HF P$	Power in the HF band (0.15–0.4 Hz): marker for vagal modulation of heart rate	
$VLFNorm$	Ratio between $VLFP$ and TP	
$LFNorm$	Ratio between LFP and TP	
$HFNorm$	Ratio between HFP and TP	
LFP/HFP	Ratio between LFP and HFP	
$HFPeak$	Frequency of highest peak in the HF band	
$LFPeak$	Frequency of highest peak in the LF band	

in the HRV data, and hence quantifying their regularity and predictability. $ApEn$ sets a nonnegative number to the HRV time series, with larger values corresponding to more irregularity in the HRV data. The more regular and predictable the HRV series, the lower will be the $ApEn$ measure. The value of $ApEn$ relies on three main parameters: the embedding dimension m , the tolerance value r , and the data length N . $ApEn$ measures the logarithmic probability that runs of patterns that are close (within r) for m adjoining observations stay close (within the same tolerance width r) on the next incremental comparisons.

Consider N data points from RR time series $\{x(n)\} = x(1), x(2), \dots, x(N)$. $ApEn$ can be calculated using the following steps [69]:

- A set of m -vectors $X(1), X(2), \dots, X(N - m + 1)$ in R^m is formed, where $X(i) = [x(i), x(i + 1), \dots, x(i - m + 1)]$, $i = 1, \dots, N - m + 1$. These vectors represent m consecutive x values, starting with the i th point.
- The distance between $X(i)$ and $X(j)$, $d[X(i), X(j)]$, is defined as the maximum absolute difference between their respective scalar components:

$$d[X(i), X(j)] = \max_{u=1,2,3,\dots,m} |x(i + u - 1) - x(j + u - 1)| \tag{2}$$

- Define $N^m(i)$ as the number of $j(j = 1, \dots, N - m + 1, j \neq i)$ that satisfies the condition $d|X(i) - X(j)| < r$ for a certain given $X(i)$. Then, the frequency of patterns like a given one of window length m within window width $r, C_r^m(i)$, can be expressed as:

$$C_r^m(i) = \frac{N^m(i)}{(N - m + 1)} \tag{3}$$

- Calculate the natural logarithm of each $C_r^m(i)$ and average it over i to obtain the mean of fraction patterns, $\emptyset^m(r)$, as follows:

$$\emptyset^m(r) = \frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} \ln C_r^m(i) \tag{4}$$

- After increasing the dimension to $m + 1$, repeat the previous steps to get $C_r^{m+1}(i)$ and $\emptyset^{m+1}(r)$.
- $ApEn$ is defined as the difference between $\emptyset^m(r)$ and $\emptyset^{m+1}(r)$:

$$ApEn(m, r, N) = \emptyset^m(r) - \emptyset^{m+1}(r) \tag{5}$$

The two input parameters r and m should be chosen to achieve good statistical reproducibility for $ApEn$. The value of r can be selected as a fraction of the standard deviation (SD) of RR time series $x(n)$ [69]. In the current study, the best reproducibility for $ApEn$ is achieved using the optimized values of $m = 2$ and $r = 0.2SD$.

ShanEn is another important measure for the complexity of the RRI series distribution based on information theory [70]. *ShanEn* becomes large when the distribution is flat (the patterns are uniformly distributed) and the series contains the maximum amount of information. In contrast, *ShanEn* becomes small when some patterns are more likely, while others are missing (the patterns are normally distributed).

Consider HRV time series $RR = \{RR_i, i = 1, 2, \dots, N\}$. Each sample of the series RR is normalized to obtain the series $rr = \{rr_i, i = 1, 2, \dots, N\}$. The dynamic range ($rr_{max} - rr_{min}$) is divided into a fixed number of quantization levels ξ by performing a coarse graining of the dynamics with a resolution of $(rr_{max} - rr_{min})/\xi$. The quantization levels are labeled with numbers ranging from 0 to $\xi - 1$ to form a quantized series $rr^\xi = \{rr_i^\xi, i = 1, 2, \dots, N\}$. Using the quantized series rr^ξ , patterns of L delayed samples are formed as $rr_{L,i}^\xi = (rr_i^\xi, rr_{i-1}^\xi, \dots, rr_{i-L+1}^\xi)$, where $rr_L^\xi = \{rr_{L,i}^\xi, i = 1, 2, \dots, N - L + 1\}$. The *ShanEn* can be defined as:

$$ShanEn(L, \xi) = - \sum_{i=1}^{N-L+1} p(rr_{L,i}^\xi) \log(p(rr_{L,i}^\xi)) \quad (6)$$

where $p(rr_{L,i}^\xi)$ is the probability of the pattern $rr_{L,i}^\xi$ approximated by its sample frequency and the sum is taken over all different patterns. In this work, ξ and L are set to 6 and 3, respectively, to achieve the best results.

Poincaré plot is another commonly nonlinear method that illustrates graphically the correlation between successive RR intervals, i.e., plot of RR_n versus RR_{n+1} . Each pair of consecutive RR intervals is represented as one point in the Poincaré plot. The plot shape represents an important feature, and it can be parametrized by fitting an ellipse to

the plot (see Fig. 4). The ellipse orientation is identified based on the line-of-identity ($RR_n = RR_{n+1}$). The points above the line-of-identity represent all prolongations ($RR_n < RR_{n+1}$), and the points below this line correspond to all shortenings of the interval between 2 consecutive RR intervals ($RR_n > RR_{n+1}$). The standard deviation measuring the dispersion of points across the identity line of the Poincaré plot, denoted by $SD1$, identifies the width of the Poincaré cloud, and therefore determines the level of short-term HRV [71]. In contrast, the standard deviation measuring the dispersion of points along the line-of-identity, denoted by $SD2$, indicates the length of the Poincaré cloud and therefore describes the long-term HRV. Both $SD1$ and $SD2$ represent the axes of the fitted ellipse as shown in Fig. 4a.

Figure 4 shows the Poincaré plot for three cases: NSR, ARR, and CHF. It can be noted from Fig. 4a that the dominant Poincaré pattern of NSR subject contains cluster of RR intervals lying between 0.6 and 1.1 s with accepted deviation along the line-of-identity axis. This demonstrates the ability of heart to adapt with different stimulations. Figure 4b shows the scattered RR intervals of ARR case with higher deviation than the NSR case, showing the irregular heart rate of ARR. Figure 4c illustrates the limited RR intervals of the CHF case lying between 0.2 and 0.8 s.

Short-term HRV analysis can be enhanced using a group of indices that assess the fragmentation of HRV, defined by frequent changes in heart rate acceleration sign [30]. From the ECG signal, two time series are constructed: (1) HRV time series with length N , $\{RR_i\} = \{t_{Ri} - t_{Ri-1}\}$, $1 \leq i \leq N$, where t_{Ri} is the time of occurrence of the i^{th} beat, and (2) increment time series $\{\Delta RR_i\} = \{RR_i - RR_{i-1}\}$ (differences between consecutive RR intervals). Four fragmentation indices can be calculated from those time series as follows [30]:

- The percentage of inflection points in the time series $\{RR_i\}$, where a time instant t_{Ri} is said to be an inflection point if $\Delta RR_i \Delta RR_{i+1} \leq 0$.
- The inverse of average segment length (IALS) for both acceleration and deceleration segments. The acceleration segment length is calculated by counting the number of intervals between consecutive inflection points when the difference between two RR intervals is less than 0. The deceleration segment length is calculated using the same way but when the difference between two RR intervals is greater than 0.
- The percentage of short segments (PSS). It is the complement of the percentage of RR intervals in acceleration and deceleration segments with three or more RR intervals.

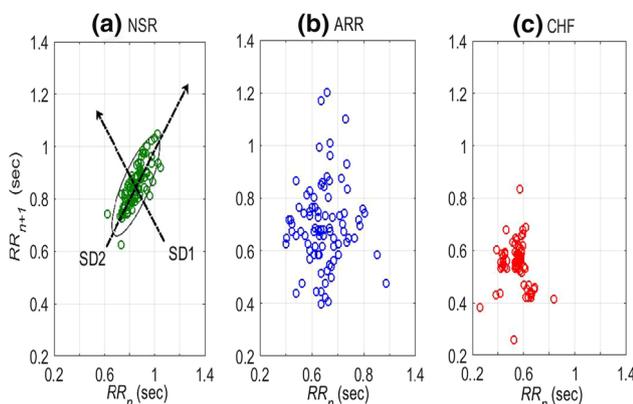


Fig. 4 Poincaré plot for three subjects: **a** NSR, **b** ARR, and **c** CHF. Length and width are shown graphically on the plot

- The percentage of RR intervals in alternation segments (PAS), where an alternation segment is defined as a sequence of at least four RR intervals on condition that the sign of heart rate acceleration is changed every beat.

Detailed analysis of HRV fragmentation indices can be found in [30].

2.3.2 ECG features from chaotic analysis

The electrical activity of human heart can be considered as a chaotic system [11, 12]. Note that the available information about the heart as a dynamical system is a group of ECG measurements from skin-mounted sensors. There is no mathematical formulation of the heart dynamics and only some observables whose total number of state variables is not known can be investigated. To study the chaotic behavior of such system, two important parameters, namely the Lyapunov exponents (LEs) and correlation dimension (CD), are investigated. The importance of the two parameters in modeling the chaotic nature of ECG signals can be summarized as follows:

- Lyapunov exponents (LEs): they are important measures for the sensitive dependence of nonlinear system variables on the initial conditions [14]. Note that for m -dimensional system, there are always m values of LEs. It is common to refer to the largest one as the Largest Lyapunov Exponent (LLE). The chaotic behavior of ECG signal can be identified based on the value of LLE which is estimated from the ECG time series. If the value of LLE of an ECG signal is positive, then the ECG signal is said to have chaotic characteristics [10]. For strict periodic activity, the LLE value is zero, and its value is infinite for pure noise. Note that LLE is a nonlinear metric that calculates the exponential rate of divergence of contiguous trajectories of the state space built by the ECG data, and it shows the overall characteristics of the instantaneous heart rhythm. The two common methods of calculating LEs are Rosenstein [72] and Wolf [73]. Both techniques have been reported to be appropriate for short time series [74]. Although neither of the two methods is robust enough to the noise [74], several studies reported that Rosenstein's method achieved better performance than Wolf's method under different noise conditions [75]. Also, Rosenstein's method is more appropriate for estimating the LLE than the Wolf's algorithm, especially for small datasets, because it avoids approximations by disregarding the procedure of neighbor replacement [72]. Moreover, estimates from the Rosenstein's method are less influenced by changing the embedding dimension and the reconstruction delay than the Wolf's method. This reveals that Rosenstein's

method is quite robust to changes in these quantities in comparison with Wolf's algorithm. Rosenstein's method is more appropriate for estimating the LLE, yielding stable results for several applications [76]. Note that the instability of calculating LEs arises mainly from the noise sensitivity of the used algorithm. In the current study, all noise sources are eliminated automatically before extracting any features, allowing for stable estimates of LLE. In this work, the LLE of each ECG segment is estimated using the efficient method of Rosenstein [72]. Rosenstein's method is simple in realization and it is characterized by high computational efficiency.

- Correlation dimension (CD): it is one the most important parameters in the chaos theory which can be utilized to determine the number of primary controls of the variable and thus estimates the degree of freedom of the underlying process. CD is considered as a measure of the attractor size, and it is derived from the correlation integral, describing the spatial arrangement of points on the chaotic attractor. Although it is not possible to find the direct correlation between the pattern of phase space plots of ECG data and the cardiac status, there is a unique CD range for each cardiac condition. Therefore, the CD of each ECG signal represents a characteristic parameter for diagnosing different heart disorders. Several techniques were developed for calculating the CD, among which the Grassberger and Procaccia algorithm [15] has been employed most and is also implemented in this work with some adaptation. The Grassberger and Procaccia method utilizes the basic idea of phase space reconstruction and the embedding theorem to calculate the CD, which requires choosing an appropriate scaling region. The original Grassberger and Procaccia method is not adaptive and relies heavily on visual inspections for selecting the scaling regions, which is prone to human errors and thus not objective. In the current study, a simple automated method is utilized for identifying the linear scaling region without the need for visual inspection. To explain this method, assume that the dimension D of an object is the exponent that scales the bulk b of an object with linear distance r (i.e., $b \propto r^D$). In the Grassberger–Procaccia method, the correlation integral $C(r)$ is utilized to represent the bulk [77]. The CD is obtained by calculating the slope of the linear region of the plot of $\log(C(r))$ against $\log(r)$ for low values of r . In this work, a simple automated method for identifying the linear region is employed by computing the second derivative of the $\log(C(r))$ against $\log(r)$ plot and looking for the longest plateau with values lower than a predefined threshold (set here

as 0.1). If several linear regions are found to have the same length, the one that provides the maximum CD value is chosen.

Several studies reported the importance of both CD and LEs in different ECG classification problems [78, 79]. In [78], the CD and LLE are utilized to distinguish five different arrhythmia classes of ECG signals, and the results revealed the usefulness of both features in arrhythmia classification. In this work, the CD and LLE are calculated to model the chaotic behavior of ECG signals and hence provide additional features for enhancing the ECG classification performance. Note that the two chaotic parameters are not sufficient alone to make an accurate ECG diagnosis so that they are used in conjunction with other ECG features, HRV measures, and DL features.

2.3.3 ECG Transferred DL features

The AlexNet model is investigated to automatically capture the ECG characteristics, instead of training a CNN from the beginning, which would require a huge amount of labeled data and high computational time along with the problem of overfitting. Although there are many improvements on top of the AlexNet model [41] emerged from VGGNet [42] and GoogleNet [43] such as VGGNet16, VGGNet19, ResNet50, ResNet101, Inceptionv3, and DenseNet, the experimental results showed that the AlexNet architecture achieved superior performance over all other models. The AlexNet model comprises of two major parts: (1) feature extraction for the input image, and (2) fully connected multi-layer perceptron (MLP) which executes the classification operation. In the current work, the fully connected MLP is not used for classification due to its high computational cost, and instead a feature selection technique is used first to choose the optimal features from the AlexNet output feature vector to be concatenated with ECG and HRV features. Then, the concatenated features are sent to a ML classifier. Figure 1 shows the structure of the proposed framework. All details of the AlexNet architecture can be found in [41].

The AlexNet architecture is employed to extract automatically the important subtle features by constructing hierarchical representations of the transformed 2-D time–frequency representation of the ECG signal. After transforming the input ECG signal into image using the CQ-NSGT technique, the resultant 2-D image is sent to the AlexNet network. The AlexNet network consists of eight layers, the first five are convolutional layers, some of them followed by max-pooling layers, and the last three are fully connected layers. Each layer of the convolution layers is followed by a Rectified Linear Unit (ReLU) except the output layer. In comparison with the used ReLU function,

new activation functions like the scaled polynomial constant unit activation function (SPOCU) [80] may increase the gradient flow of sparse training and avoid the vanishing gradient problem. The SPOCU function has a finite range, allowing more stability for gradient-based training methods. Also, the continuous differentiability of SPOCU enables identity approximation near the origin and allows the use of gradient-based optimization methods which is not the case for ReLU function. Despite the sophisticated structure of the SPOCU function, some recent studies have reported similar classification performance for both SPOCU and ReLU activation functions [81]. In comparison with SPOCU, the computation of the ReLU function is much simpler and easier which reduces the computational burden. Also, the ReLU function speeds up the training process when using different parameter initialization methods and accelerates the forward propagation process with low computational cost. SPOCU may achieve better performance than ReLU in image classification problems with a large database of complex images. However, in the current ECG classification, it is more efficient to use the ReLU function of the AlexNet architecture to avoid any computational complexity, while achieving real-time ECG diagnosis with high performance.

Compared with other common pre-trained CNNs, AlexNet has the following privileges: (i) The overfitting issue is alleviated with the use of a dropout layer that discards randomly some neurons in the training process; (ii) the richness of features is improved with the use of max-pooling layers; (iii) the nonlinear activation function ReLU is used to accelerate the forward propagation process. AlexNet is chosen for this study as it is a proper trade-off between speed and accuracy. The first layers of the AlexNet network extract the basic image features, which are then processed by deeper network layers. The deeper layers capture the higher-level features, which are created using the simple features of the first layers. These higher-level features transform all basic features into a richer image representation that is appropriate for classification. This means that early layers determine the common features, while subsequent layers identify the more subtle features. Weights of the AlexNet network were trained on approximately 1.2 million images from ImageNet dataset in order to fit the training ECG images and fine-tune the model and hence alleviate the overfitting problems, while improving the accuracy of ECG classification. Note that fine-tuning makes the pre-trained CNN model learn new task faster and easier than train network from scratch using randomly initialized weight.

In total, 4096 features were extracted from the pre-trained AlexNet architecture, and a feature reduction technique is employed to reduce the number of features by selecting the optimal subset of features. The main

disadvantage of AlexNet architecture is the discriminative ability of the learned features, especially in the latter layers of the network, which can significantly influence the performance. This problem can be solved by replacing the MLP classification layer of AlexNet architecture with other ML classification algorithms. The training parameters are tuned to get the best performance. This yields a batch size of 15, learning rate of 10^{-4} , and maximum epochs of 20. To reduce overfitting problems, the AlexNet architecture is modified by changing the default 0.5 dropout probability of the FC6 layer to be 0.6. Also, the MLP classification layers (FC7 and FC8) are removed and instead the output features are sent to a feature selection technique (see Fig. 1).

2.3.4 Feature selection

Since the number of output features extracted from the AlexNet is 4096, the pair-wise feature proximity (PWF) feature reduction technique [82] is investigated to reduce the number of features by selecting the optimal subset of features. The resultant selected optimal DL features are 500, discarding 3596 features. In the PWF method, the feature dimension closeness among a pair of samples from the same (or different) class is used to choose the important features for class differentiation. Unlike other techniques which use all feature vector for distance measurement, the PWF technique utilizes only a subset of the feature vector. Therefore, optimal features, which minimize the within-class distance and maximize the between-class distances for each pair of samples, should bring the points of the same class closer and maintain a far distance from the other class instances. The PWF technique allows selecting the optimal features based on the pair-wise proximity in feature values, while reducing the computational burden using a heuristic search approach for feature ranking. In the current study, the output features from the AlexNet are sent to the PWF algorithm to reduce the feature vector dimension by selecting the optimal features and hence reducing the computational load, avoiding the risk of overfitting, and improving the feature vector quality.

Assume the data contains n number of d -dimensional points $(x_i = [x_i^1, x_i^2, \dots, x_i^d] \in \mathbb{R}^d)$, given by $\{(x_i, y_i)\}_{i=1}^n$, where $y_i \in \{1, 2, \dots, c\}$ is the class label of the corresponding data. Assume $X = [x_1, x_2, \dots, x_n] \in \mathbb{R}^{d \times n}$ is the total data matrix and $f^i = [f^{i1}, f^{i2}, \dots, f^{im}]$ is the i^{th} row of the matrix X . The feature selection process can be formulated as obtaining m features from d dimensions, which will achieve the optimum classification performance. Heuristic strategies are usually employed to find the best features by examining each feature independently from the d features and choosing the top m features.

The PWF algorithm employs a naive way of selecting features based on pair-wise feature similarity with Fisher Score (FS) computation. In FS, the evaluation criterion maximizes the between-class variance while minimizing the within-class variance. The FS formulation can be expressed as [82]:

$$F(f^i) = \frac{\sum_{k=1}^c n_k (\mu_k^i - \mu^i)^2}{\sum_{k=1}^c n_k (\sigma_k^i)^2} \quad (7)$$

where μ^i denotes the overall mean of i^{th} feature and n_k denotes the number of samples of k^{th} class. μ_k^i and σ_k^i are the i^{th} feature mean and variance of k^{th} class, respectively. In PWF, the pair-wise feature similarity is employed to select the appropriate features, in which a feature is said to be an important feature if it conserves the samples of the same class very close, while retaining the points from distinct classes very far [82]. This means that for a pair of points, the feature selection process is performed based on the point pairs which are very close for the same class and very far for distinct classes.

Define $p_{jk} = [b_1, b_2, \dots, b_d]^T$, $b_i \in \{0, 1\}$, with $b_i = 1$ as the features along which the points pair $(x_i, x_k); y_j = y_k$ are close to each other (the pair-wise within-variance is minimum). These features can be selected by sorting the distance between individual features in ascending order and choosing the first few features satisfying the following optimization problem

$$\max_{p_{jk}} p_{jk}^T p_{jk}; \text{ s.t. } |x_j - x_k|^T p_{jk} < \tau \quad (8)$$

where τ is a threshold. In the PWF algorithm, we use the Manhattan distance between x_i and x_k , defined as $\sum_{i=1}^d |x_j^i - x_k^i| = |x_j - x_k|^T \mathbf{1}$, where $\mathbf{1}$ is a vector of ones. Assume that we need to keep β number of features, which are close for the pair (x_i, x_k) , out of d features. This allows representing the selection process as follows:

$$\min_{p_{jk}} |x_j - x_k|^T p_{jk}; \text{ s.t. } p_{jk}^T p_{jk} = \beta \quad (9)$$

where the term $|x_j - x_k|^T p_{jk}$ can be interpreted as a distance measure with a subset of features for which $b_i = 1$.

Similarly, let $q_{jk} = [b_1, b_2, \dots, b_d]^T$, $b_i \in \{0, 1\}$ be the features along which the pair $(x_i, x_k); y_j \neq y_k$ are farthest if $b_i = 1$. A similar way of finding the features that differentiate the points from distinct classes can be formulated as follows:

$$\max_{q_{jk}} |x_j - x_k|^T q_{jk}; \text{ s.t. } q_{jk}^T q_{jk} = \beta \quad (10)$$

The PWF algorithm gathers the information from all possible pairs which are represented by $P = [p^1, p^2, \dots, p^d]$

and $Q = [q^1, q^2, \dots, q^d]$. P and Q denote, respectively, the normalized histogram of features based on their contribution toward closeness/discriminating power between different classes. The reduced feature dimensions are identified based on the existence in both P and Q . An important feature should be capable of differentiating points from distinct classes while bringing the points of same class closer. Therefore, an important feature is the one which has higher probability of occurrence in both P and Q .

2.4 ECG classification

Since the computational time using an end-to-end CNN is relatively high, the MLP classification part of the AlexNet architecture is replaced in the current study with another classification methodology (see Fig. 1). After using the PWF feature selection algorithm, the optimized DL features are concatenated with both HRV measures and ECG features.

Note that the total number of linear features is 18 classified as follows: 15 linear HRV features, including 10 frequency domain features and 5 statistical time domain features (see Table 1), along with 3 linear ECG features (mean, variance, and standard deviation). The total number of nonlinear features is 11 classified as follows: 9 nonlinear HRV features, including α , $ApEn$, $ShanEn$, Poincaré descriptors ($SD1$ and $SD2$), and the four fragmentation indices, beside 2 nonlinear ECG features (LLE and CD). The optimized selected 500 DL features are concatenated with both HRV measures and ECG features (18 linear features and 11 nonlinear features). Then, the resulted concatenated features are fed to a ML classifier to distinguish CHF, ARR, and NSR subjects. In this work, the linear discriminant analysis (LDA) classifier [83] is investigated for ECG multi-class classification, and its performance is compared with different ML classifiers such as kernel SVM, K-nearest neighbor (KNN), and decision tree [84, 85].

The selection of ML classifier depends on many factors, including the data size, number of features, accuracy, and computational time. Some ML classifiers perform better with larger datasets than others. For small datasets, classifiers with high bias/ low variance such as LDA and linear SVM work better than low bias/ high variance classifiers such as KNN and decision trees. Algorithms such as SVM and decision tree, which involve tough tuning of learning parameters, require higher computational time to process and train the data, while algorithms such as logistic regression (LR) and LDA are easy to understand and implement, and hence they have fast execution. To achieve the best classification performance, different ML classifiers (LDA, kernel SVM, KNN, and decision tree)

spanning all classes were investigated. All learning parameters of the selected classifiers are optimized using the well-known grid search technique. Grid search is the most basic hyperparameter tuning method, which builds a model for each possible combination of all learning parameters of the classifier, evaluating each model, and choosing the architecture which achieves the best classification results.

LDA is a generalization of the statistical model of Fisher's linear discriminant [83]. It is used to find a linear combination of features that separates two or more classes in a dataset. The classes of the labelled patterns are used to calculate the discriminant functions $\{f_1, f_2, \dots, f_c\}$, where c is the number of classes. Then, these discriminant functions are utilized to specify the decision boundaries between classes and the region of each class. A comparison between the c different discriminant functions is made to assign the class label with the maximum score to the unknown pattern.

Note that LR and LDA are multivariate statistical techniques that can be utilized to investigate the same research studies. The two methods do not differ in their functional forms, a combination of the independent variables and a base for classification. However, they differ in the way of the coefficient estimation. The sample size has a significant influence on the difference between the two approaches. The main difference between the two methods is very clear for small sample size, as the difference between the distribution of the training sample and that of the test sample can be essential for the LR to be able to achieve high classification performance [86]. In contrast, LDA assumes normality, and the errors occurred in prediction are caused by the errors in estimation of the mean and the variance on the sample. The differences between the two methods are negligible when the sample size is adequately large. LDA outperforms LR for small sample size, while the results of the two methods are getting closer for large sample size. Since the data size under investigation is not very large, the LDA is chosen for the current multi-class classification problem.

In comparison with LDA, LR has some limitations. When classes are well-separated, the estimation process for the parameters of LR is surprisingly unstable, which is not the case for LDA. LR uses maximum likelihood estimation (MLE) technique to derive all parameters [87]. Note that MLE depends on the large-sample asymptotic normality which may cause a degradation in the estimation performance when there are few cases for each observed collection of independent variables. LR requires a large number of cases to achieve reliable estimates for the logistic coefficients [87]. The more unequal groups are constructed from the dependent variable, the more cases are required. This reveals that for a small number of cases,

the LDA is more stable than the LR. Also, several studies showed that LDA achieves better performance than LR in multi-class classification problems, while both methods have nearly the same performance for binary classification problems [86]. The LDA has high bias/low variance, and hence, it outperforms other ML classifiers for relatively small datasets. Unlike other algorithms which require difficult tuning of learning parameters, the LDA classifier is easy to understand and implement, and hence it has high computational efficiency. It is very critical to reduce the computational cost and hence achieve real-time diagnosis. Therefore, the LDA is employed in the current ECG multi-class classification problem as it represents a perfect compromise by achieving high classification performance with low computational requirements. The experimental results showed that the LDA algorithm achieves maximum separation among the classes in the

feature set and consequently has superior performance over all other classifiers.

3 Results

To examine the effectiveness of the proposed framework in distinguishing CHF, ARR, and NSR cases, all the available ECG records are investigated. Table 2 shows the distribution of ECG signals retrieved from all databases under investigation. It can be noted that the ECG records utilized in the current work include distinct subjects of patient age and sex. To assess the usefulness of the current ECG multi-classification framework, various evaluation metrics, including accuracy (*Acc*), sensitivity (*Se*), precision (*Pr*), and specificity (*Sp*) are calculated based on four main parameters: the number of true positives (*TP*), true negatives (*TN*), false negatives (*FN*), and false positives (*FP*). *Acc* is a measure of true predictions, while *Pr* is the positive predictive value or the fraction of *TP* among the positive predictions. *Se* is the proportion of positives that are correctly identified, while the *Sp* is the proportion of negatives that are correctly identified. Those evaluation metrics are given by [88]:

$$Se = TP / (TP + FN) \tag{11}$$

$$Sp = TN / (TN + FP) \tag{12}$$

Table 2 Distribution of ECG signals retrieved from the three databases under investigation

Class	Segments	Cases	Cases		Age (years)	
			Male	Female	Male	Female
ARR	576	47	25	22	32–89	23–89
CHF	180	15	11	4	22–71	54–63
NSR	216	18	5	13	26–45	20–50

Table 3 Average classification results of the proposed ECG multi-class classification system using fivefold cross-validation

Fold	Type	Evaluation metric			
		<i>Se</i> %	<i>Sp</i> %	<i>Pr</i> %	<i>Acc</i> %
Fold-1	ARR	100.00	97.22	98.00	98.82
	CHF	93.75	100.00	100.00	98.82
	NSR	100.00	100.00	100.00	100.00
Fold-2	ARR	98.02	98.55	99.00	98.24
	CHF	96.55	98.58	93.33	98.24
	NSR	100.00	100.00	100.00	100.00
Fold-3	ARR	97.03	97.10	98.00	97.06
	CHF	93.10	97.87	90.00	97.06
	NSR	100.00	100.00	100.00	100.00
Fold-4	ARR	98.04	100.00	100.00	98.82
	CHF	100.00	98.59	93.33	98.82
	NSR	100.00	100.00	100.00	100.00
Fold-5	ARR	96.15	100.00	100.00	97.65
	CHF	100.00	97.90	90.00	98.24
	NSR	100.00	99.24	97.50	99.41
Average	ARR	97.85 ± 1.44	98.57 ± 1.42	99.00 ± 1.00	98.12 ± 0.77
	CHF	96.68 ± 3.30	98.59 ± 0.86	93.33 ± 4.08	98.24 ± 0.72
	NSR	100.00 ± 0.0	99.85 ± 0.34	99.00 ± 1.12	99.88 ± 0.26
Overall average performance		98.18	99.00	97.11	98.75

$$Pr = TP / (TP + FP) \tag{13}$$

$$Acc = (TP + TN) / (TP + FP + TN + FN) \tag{14}$$

Separate ECG signals are used in the training and testing phases as follows: 802 ECG segments taken from 79 ARR, 24 CHF, and 30 NSR records are utilized for training and validation, while 170 ECG segments taken from the remaining 17 ARR, 6 CHF, and 6 NSR records are kept for the test phase. To authenticate the diagnosis results and tackle deviation issues, the k-fold cross-validation technique is utilized [58]. In this method, the training and validation data are randomly partitioned into k equal sized groups, each group is utilized once for validation, and (k-1) times for training. In this work, the fivefold cross-validation is utilized, where the 802 ECG segments, kept for training and validation, are randomly partitioned into five equal sized groups, each time one group is retained as a validation set and the remaining four groups are retained as a training set. The 170 ECG segments, reserved for testing, are examined to obtain a fair assessment of the final model.

Table 3 summarizes the fivefold evaluation metrics and standard deviation of the five rounds for the CHF, ARR, and NSR classes. For each fold, the evaluation metrics (*Acc*, *Se*, *Sp*, and *Pr*) are calculated for the test data. The average *Acc*, *Se*, *Sp*, and *Pr* are calculated by averaging the results of the 5-rounds (see Table 3). The standard deviation is calculated for each metric relevant to the three categories. It can be noted that the proposed approach provides very accurate diagnosis results in terms of average *Se* (97.85%, 96.68%, and 100%), *Acc* (98.12%, 98.24%, and 99.88%), *Pr* (99.00%, 93.33%, and 99.00%) and *Sp*

(98.57%, 98.59%, and 99.85%) for ARR, CHF, and NSR, respectively.

As mentioned before, the concatenated features (optimized DL features, ECG features, and HRV measures) are sent to a ML classifier to distinguish CHF, ARR, and NSR subjects. Different ML classifiers including LDA, kernel SVM, KNN, and decision tree are investigated, and the classification results are summarized in Table 4 and Fig. 5. A detailed comparison between the proposed LDA classifier and other classification methods is performed by investigating the average of all evaluation metrics (*Acc*, *Se*, *Sp*, and *Pr*) for each fold using all classifiers. Figure 5 illustrates the average *Acc*, *Se*, *Sp*, and *Pr* of the proposed system for each fold using LDA, kernel SVM, KNN, and decision tree classifiers.

The average classification results and standard deviation for each evaluation metric are calculated for all classifiers as shown in Table 4. Note that the comparison between all the classifiers is made using the same techniques in all other steps of the proposed system shown in Fig. 1. Table 4 and Fig. 5 show that the kernel SVM outperforms both KNN and decision tree classifiers. However, the LDA achieves better performance than kernel SVM in terms of the average *Acc* (98.75% versus 97.41%), *Se* (98.18% versus 95.19%), *Sp* (99.00% versus 97.45%), and *Pr* (97.11% versus 95.96%). The LDA algorithm achieves maximum separation among the classes in the feature set and consequently has superior performance over all other classifiers. Figure 6 shows the overall classification results of the proposed system using LDA, kernel SVM, KNN, and decision tree classifiers. It can be noted that the LDA

Table 4 The classification results of the proposed ECG multi-class classification system using different ML classifiers

Classifier	Type	Evaluation metric			
		<i>Se</i> %	<i>Sp</i> %	<i>Pr</i> %	<i>Acc</i> %
KNN	ARR	83.56 ± 0.79	80.56 ± 3.29	87.40 ± 2.79	82.47 ± 1.49
	CHF	63.02 ± 5.63	89.40 ± 0.87	48.00 ± 5.06	85.76 ± 1.28
	NSR	89.74 ± 4.42	98.43 ± 0.56	87.40 ± 1.77	96.24 ± 1.48
Overall performance		78.77	89.56	74.27	88.16
Decision tree	ARR	91.93 ± 0.36	90.42 ± 2.64	93.40 ± 2.07	91.29 ± 1.13
	CHF	76.65 ± 4.76	94.34 ± 0.46	73.33 ± 2.36	91.29 ± 1.13
	NSR	100.00 ± 0.0	99.70 ± 0.68	93.4 ± 2.24	99.76 ± 0.53
Overall performance		89.53	94.82	86.71	94.12
Kernel SVM	ARR	97.79 ± 1.73	93.97 ± 2.04	95.60 ± 1.67	96.12 ± 0.79
	CHF	91.21 ± 6.20	99.30 ± 1.20	96.67 ± 5.77	97.65 ± 0.72
	NSR	95.56 ± 2.07	99.08 ± 0.34	95.60 ± 1.12	98.47 ± 0.53
Overall performance		95.19	97.45	95.96	97.41
Proposed LDA	ARR	97.85 ± 1.44	98.57 ± 1.42	99.00 ± 1.00	98.12 ± 0.77
	CHF	96.68 ± 3.30	98.59 ± 0.86	93.33 ± 4.08	98.24 ± 0.72
	NSR	100.00 ± 0.0	99.85 ± 0.34	99.00 ± 1.12	99.88 ± 0.26
Overall performance		98.18	99.00	97.11	98.75

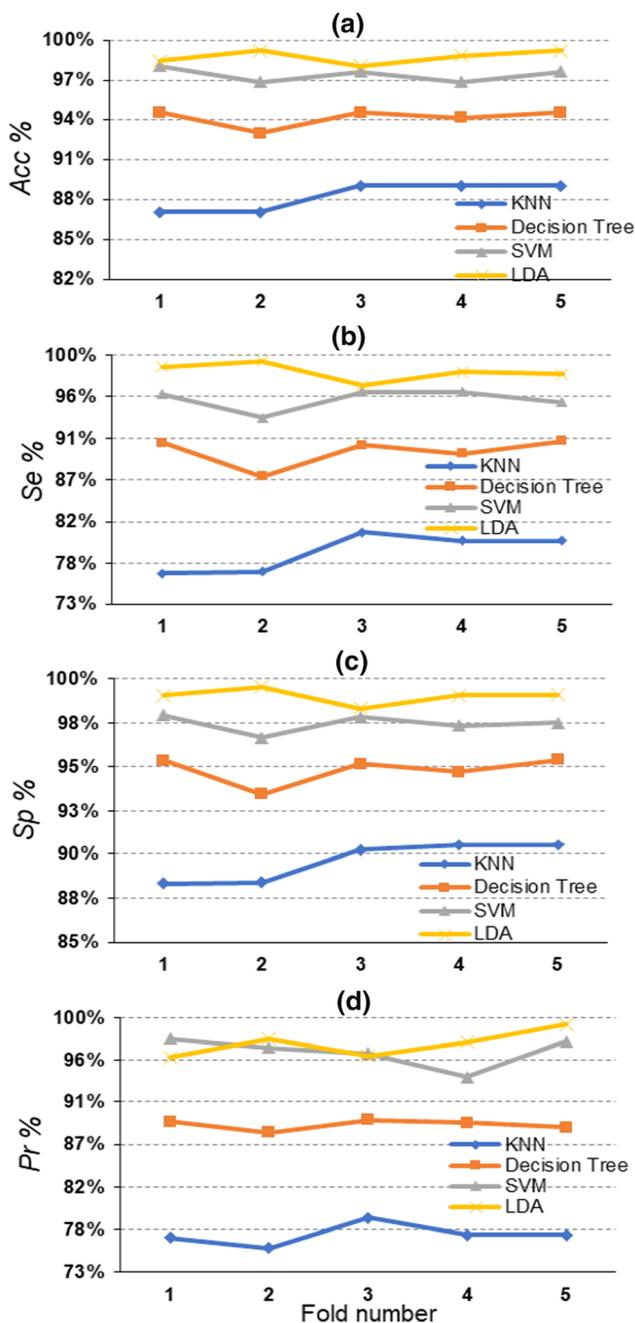


Fig. 5 The average a Acc, b Se, c Sp, and d Pr of the proposed diagnosis system for each fold using different ML classifiers

method is more efficient than other ML classifiers with nearly the same consumed time. This elucidates the usefulness of the LDA classifier in boosting the ECG multi-class classification performance over other ML classifiers.

The overall average ECG multi-class classification performance is investigated for different combinations of DL, linear, and nonlinear features (before and after feature selection, with and without linear/nonlinear features). The results are summarized in Table 5. It can be noted that

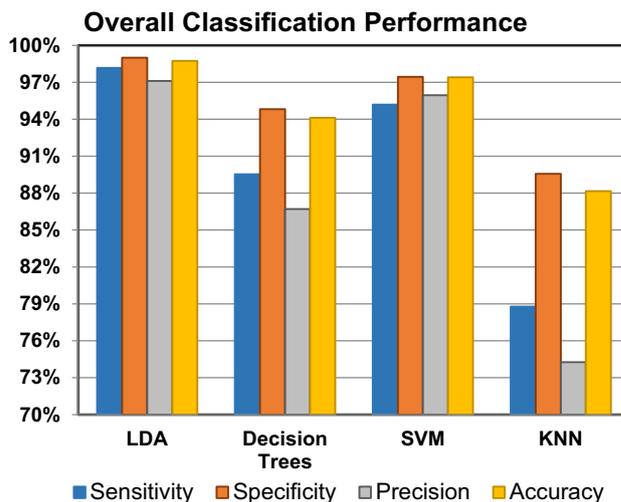


Fig. 6 The average evaluation metrics of the proposed framework using different classification algorithms

combining the optimized selected DL features with both linear and nonlinear HRV/ECG features achieves superior performance over other combinations. This means that applying the PWF algorithm on the extracted features from the AlexNet model allows reducing the feature vector dimension by selecting the optimal DL features and hence significantly reducing the computational complexity. Moreover, combining the optimized DL features with both HRV measures and ECG features succeeds to improve the feature vector quality and enhance the classification performance without high computational cost. This reveals the usefulness of the proposed ECG multi-class classification approach in real-time clinical setting.

4 Discussion and conclusions

This work proposes a novel hybrid automatic diagnosis system of pre-trained CNN model combined with HRV measures and ECG features for distinguishing CHF, ARR, and NSR cases with high classification performance and low hardware complexity. The proposed system investigates the PWF technique for selecting the optimal subtle DL features from the output features of the AlexNet model. The optimized DL features are then combined with both HRV measures and ECG features representing the fundamental differences between CHF, ARR, and NSR classes. The resulted concatenated features are sent to the LDA classifier to differentiate the three classes. An important contribution of the proposed technique is to overcome the high computational cost of the end-to-end CNN by replacing the MLP classification layer with the DLA classifier. Different ML classifiers including kernel SVM,

Table 5 Average fivefold classification results of the proposed ECG multi-class classification system for different combinations of DL, linear, and nonlinear features using LDA classifier

Features	<i>Se</i> %	<i>Sp</i> %	<i>Acc</i> %
DL features without feature selection, linear features, and nonlinear features	95.43	97.57	96.35
Selected DL features and linear features	97.29	96.76	96.36
Selected DL features and nonlinear features	97.71	98.21	96.87
Selected DL features, linear features, and nonlinear features	98.17	99.00	98.74

KNN, and decision tree are compared with the LDA algorithm, and the classification results show the superior performance of the LDA method. To the best of the authors' knowledge, it is the first time to merge optimized subtle DL features with both HRV measures and ECG features in one multi-class classification system that results in very accurate classification performance using reasonable hardware and software requirements. The proposed approach is applied to real ECG records, and the performance is evaluated in terms of *Acc*, *Se*, *Sp*, *Pr*, and computational time. The experimental results demonstrate that the proposed method provides very accurate multi-class classification results with overall average *Acc* of 98.74%, *Se* of 98.17%, *Sp* of 99.00%, and consumed time of 0.154 s.

The proposed ECG multi-class classification system is compared with other recent state-of-the-art systems [18, 19, 21–24, 48–51, 89–93], and the results are summarized in Table 6. To validate this comparison, all results of the systems under comparison were obtained using the same databases investigated in the current study (MIT-BIH and BIDMC databases). Also, we examined all the available 162 ECG records from the databases under investigation to avoid any selectivity of the subjects and to authenticate the classification results. Moreover, the five-fold cross-validation technique is employed to get all classification results so that overfitting and bias problems are avoided, and the model generalization on different dataset is achieved. Table 6 shows that the proposed framework provides the highest *Acc* of 98.74% utilizing shorter ECG segments than all state-of-the-art systems under comparison. Also, Table 6 demonstrates that the proposed method outperforms most of the systems under comparison [18, 19, 21–23, 48–51, 89, 90, 93] in reducing the *FN*, and consequently providing the highest *Se* of 98.17%. It can be observed that the *Acc* of the proposed approach is better than the methods of [18, 19, 21–23, 48–51, 89–93] and is nearly the same as the technique of [24]. Also, the *Sp* of the proposed approach is better than the systems of [18, 21–24, 48–50], and comparable with the methods of [19, 51].

Unlike other systems which distinguish only patients diagnosed with CHF from NSR [18, 19, 21–24, 48, 50, 51], the proposed system provides multi-class classification for CHF, ARR, and NSR with average *Acc* of 98.74%, *Se* of

98.17%, and *Sp* of 99.00%. This is more challenging than classifying only CHF cases from NSR subjects but very useful in several clinical applications. Few studies [49, 89–93] focused on the three-class ECG classification problem and reported classification results for the three classes (CHF, ARR, and NSR). The proposed approach outperforms all these techniques. The two-class ECG classification problem of distinguishing only patients diagnosed with CHF from NSR is investigated by comparing the classification results of the proposed approach with the techniques in [18, 21, 50] using 18 NSR and 15 CHF records. The experimental results reveal that the proposed approach provides accurate diagnosis results in terms of average *Acc* (98.83% and 99.82%), *Se* (97.78% and 99.98%), and *Sp* (98.92%, and 99.76%) for CHF and NSR, respectively. This yields an average *Acc* of 99.33%, *Se* of 98.88%, and *Sp* of 99.34% for the two-class classification of CHF and NSR. It can be noted that the proposed approach achieves superior performance over the studies reported in [18, 21, 50]. This reveals the effectiveness of the proposed approach in both two-class and three-class ECG classification problems.

The proposed method requires shorter ECG segments (1.3 min) than other exiting systems [18, 19, 21–24, 48–51, 89–93], revealing its robustness in real-time clinical setting. Although the proposed approach could not achieve the highest diagnosis results in comparison with very few studies in Table 6, it achieved automatic real-time diagnosis with accurate classification performance and without any dependency on the cardiologists using shorter ECG segments than all other systems under comparison. Note that some ECG diagnosis systems are sensitive to artifacts, and their classification results were produced after removing all artifacts manually. For automated ECG classification framework, either the artifacts are suppressed with an automatic technique or the system should be evaluated on unprocessed data. In the current study, an automated adaptive filter method [57] was utilized to suppress all unwanted artifacts from the raw ECG signal. Also, some studies reported very high diagnosis results without any cross-validation techniques (classification results without standard deviation values). Moreover, some studies are not suitable for real-time applications due to time-consuming preprocessing steps. This reveals that the

Table 6 Comparison between the proposed ECG multi-class classification system and other recent state-of-the-art systems

Reference, Year	Subjects from MIT-BIH NSR, MIT-BIH ARR, and BIDMC CHF Databases	Time Scale	Approach	Performance		
				<i>Se</i>	<i>Sp</i>	<i>Acc</i>
[18], 2016	18 NSR 15 CHF	1000 RR intervals (13 min)	Renyi entropy combined with conventional HRV features using K-NN classifier	80.0	94.4	87.9
[19], 2016	54 NSR RR 18 NSR 29 CHF RR 15 CHF	5 min	Multistage CHF detection system using short-term HRV dynamic measures and decision-tree-based SVM classifier	95.39	100	96.91
[21], 2017	18 NSR 15 CHF	500 samples	HRV fuzzy and permutation entropies at different frequency scales with the least squares SVM classifier	98.07	98.33	98.21
[22], 2018	52 NSR RR 18 CHF RR	5 min	ECG handcrafted features with SVM classifier	91.31	90.04	90.95
[23], 2019	54 NSR RR 29 CHF RR	From 5 min to 10 h	Multiple time scales HRV analysis with SVM classifier	93.33	98.33	94.44
[24], 2019	54 NSR RR 29 CHF RR	5 min	Multi-stage classification based on short-term HRV measures and nonlinear (<i>ShanE</i> , <i>ApEn</i> , DFA, and Poincaré plot) features	100	98.1	98.80
[48], 2018	54 NSR RR 29 CHF RR	5 min	Pre-trained CNN combined with distance distribution matrix in entropy calculation	80.99	–	81.34
[49], 2019	18 NSR 15 CHF 47 ARR	8 min	CNN with CWT algorithm	–	–	93.75
[50], 2019	18 NSR 15 CHF	2 s	11-layer deep CNN for CHF diagnosis	96.52	95.75	95.98
[51], 2019	54 NSR RR 29 CHF RR	500 sample	Short-term HRV analysis and deep CNN	76.71	99.22	87.54
[89], 2021	18 NSR 15 CHF 47 ARR	771 samples	Hybrid AlexNet-SVM deep neural networks and STFT	95.00	–	96.77
[90], 2021	18 NSR 15 CHF 47 ARR	8 min	DWT features with SVM classifier	92.59	–	95.92
[91], 2020	15 NSR 15 CHF 30 ARR	8 min	Feature fusion with wavelet and classification using SVM	–	–	93.33
[92], 2020	18 NSR 15 CHF 47 ARR	8 min	CNN architecture with CWT algorithm	–	–	96.00
[93], 2020	18 NSR 15 CHF 47 ARR	8 min	Feature extraction using CWT algorithm and classification with SVM	92.59	–	95.92
Proposed System	18 NSR 15 CHF 47 ARR	Very short segments 10,000 samples (1.3 min)	Hybrid deep CNN approach combined with HRV and ECG features using PTWF feature selection algorithm and LDA classifier	98.17	99.00	98.74

proposed approach can be used to assist cardiologists in real-time ECG classification with high efficiency.

In [24], a diagnosis system was proposed for discriminating CHF subjects from NSR cases using short-term HRV analysis. The HRV time-domain, frequency-domain features are calculated and combined with nonlinear HRV measures, including *ShanE*, *ApEn*, DFA, and Poincaré plot. A three-stage classifier comprising of simple perceptron algorithms in the initial two stages, and a more complex ML classifier algorithm in the last stage is employed for classification with *Acc* of 98.8%, *Sp* of 98.1%, and *Se* of 100%. In comparison with the proposed approach, the work of [24] reported high classification results of discriminating only CHF patients from NSR cases without any validation techniques. This means that the system of [24] was

proposed for a binary classification problem which is much easier than the ECG multi-class classification under investigation. Unlike the proposed approach which employs fivefold cross-validation, the classification results of [24] were obtained without any cross-validation analysis which may lead to bias problems and the lack of model generalization on different datasets. This may explain the very high *Se* of 100% which may not be obtained for any other datasets. Also, the study of [24] did not report any computational time analysis of the complicated three-stage classifier system which may prevent its application in real-time setting. In contrast, the proposed approach represents an efficient automated ECG multi-class classification system that can distinguish accurately CHF, ARR, and NSR

Table 7 Average evaluation metrics of the AlexNet model and other CNN structures using the same methods for all other stages of the proposed framework

Performance	AlexNet	VGG-16	VGG-19	ResNet-50	ResNet-101	Inception-v3	DenseNet
<i>Se</i> %	98.17	97.13	97.67	94.88	96.34	94.39	95.41
<i>Sp</i> %	99.00	98.26	96.12	96.36	96.24	95.16	96.20
<i>Acc</i> %	98.74	97.42	95.23	95.63	96.17	95.88	95.97

cases, while running in real-time with low hardware and software requirements.

Different pre-trained CNN architectures are investigated by replacing the AlexNet model and comparing the results using the same techniques for all other stages of the proposed framework shown in Fig. 1. Table 7 shows the overall average multi-class classification results of the AlexNet architecture and other pre-trained CNN models. For all these CNN architectures, the classification results are examined, and the results reveal that AlexNet model achieves the best classification results with the lowest consumed time. The superior performance of AlexNet model results from the simplicity of the input ECG images which do not require a depth layer in the CNN structure. In ECG diagnosis systems, the data variability is significantly lower than image classification applications, and hence, sophisticated deep CNN architectures with a huge number of free parameters may cause overfitting problems and low performance for ECG classification, while requiring more computational burden. The AlexNet architecture succeeds not only to provide better classification results than other CNN models, but also to achieve the highest computational efficiency.

Further insight requires investigating the proposed system in different ECG multi-class diagnosis systems. An important ECG classification system for distinguishing multiple heart rhythm classes, including 10 ARR abnormalities, NSR, and noise, can be a goal for future investigation using the proposed approach. In [94], a novel end-to-end DL algorithm with residual block was proposed to take raw ECG as input without signal processing and classify 12 rhythm categories (10 ARR abnormalities, NSR, and noise) using 30-s single-lead ECG rhythms. This model was trained on a huge dataset, comprising of 91,232 single-lead ECG records from 53,549 patients in an ambulatory setting. When validated against a test dataset of 328 records from 328 unique patients annotated by expert cardiologists, the *Se* of this model outperforms the average cardiologist *Se* for all rhythm categories when fixing the *Sp* at the average *Sp* obtained by cardiologists. Despite the high diagnostic performance of this DL approach [94], it may require high computational time due to the sophisticated structure of the end-to-end CNN architecture. Also,

this approach focuses only on heartbeat features extracted by the end-to-end CNN model, while ignoring the HRV features and ECG measures. Several studies reported improved diagnosis performance for multiple heart rhythm classes when integrating DL features with RR interval features [95, 96]. This lends further support for examining the proposed hybrid approach of combining DL features with both HRV features and ECG measures for classifying the 12 rhythm classes. This will be investigated in the near future and the results will be reported elsewhere.

The proposed hybrid approach can be also investigated for EEG multi-class classification problems by converting the input EEG signals into 2-D images with the CQ-NSGT technique and feeding the resulted images to the pre-trained AlexNet CNN model. Then, the optimal DL features extracted from the AlexNet model can be combined with the essential features of EEG signals, including time-domain, frequency-domain, and nonlinear dynamics measures which distinguish distinct brain diseases like epileptic seizure [97] or identify specific characteristics in the EEG signals such as motor movement/imagery tasks. Recently, few end-to-end CNN models were proposed for EEG multi-class classification of epileptic seizure type [98] and for distinguishing the five classes of EEG motor movement (4 motor imagery tasks and one rest) [99]. On the other hand, combining the optimal DL features from pre-trained CNN models with the important EEG handcrafted features is still largely unexplored research area for EEG multi-class classification. A goal of future investigation is to compare the performance of proposed hybrid approach with other recent state-of-the-art systems in such EEG multi-class applications.

The overall consumed time of the ECG multi-class classification system is around 364.52 s for training and 0.154 s for testing on a laptop with the following characteristics: Intel core i7-4600 M CPU, 2.9 GHz, 8 GB RAM, and windows 10 operating system. The computational efficiency of the proposed hybrid approach is demonstrated by comparing it with the end-to-end AlexNet architecture. In the end-to-end structure, the ECG image is fed to the AlexNet model, and the output features are the only relevant features to be classified by the MLP. The results reveal that the diagnosis time of the proposed hybrid

approach outperforms the end-to-end AlexNet model in terms of training time (364.52 s versus 10,440 s) and testing time (0.154 s versus 61 s). This means that applying the PWF feature selection algorithm on the extracted features from the AlexNet model allows reducing the feature vector dimension by selecting the optimal DL features and hence significantly reducing the computational complexity. Moreover, combining the optimized DL features with both HRV measures and ECG features succeeds to improve the feature vector quality and enhance the classification performance without high computational cost. This reveals the usefulness of the proposed ECG multi-class classification approach in real-time clinical setting.

Medical diagnosis systems based on ML, including the current ECG multi-class classification problem may need some sort of explainability and interpretability to trust the model's predictions. Explainability techniques can be generally categorized into two classes, intrinsic and post hoc. Intrinsic techniques are obtained by developing interpretable prediction models like all white-box models. In the current study, the whole framework can be only intrinsic explainable when discarding all features from the AlexNet CNN model as they are considered as black box features. Future work may involve investigating the possible model's performance degradation when the AlexNet features are missing. Unlike intrinsic explainability frameworks, post hoc techniques use secondary models to interpret the predictions of a black box model without any access to the model's inner architecture like its weights. In the future, the proposed approach can be modified by incorporating post hoc explainability methods on the CNN's final feature's map output and producing a visual map highlighting the sub regions of the input images which can contribute significantly to the model's predictions.

Although the proposed approach has been proven to be a promising tool for ECG multi-class classification, it should be widely examined by investigating larger ECG datasets with many more patients for identifying different classes of heart diseases such as rhythm classes and myocardial infarction classes. Despite the high performance of the proposed approach, there are few limitations that can be investigated in the future. First, the proposed system requires an automated approach for eliminating all noise and artifact sources from the ECG signals before extracting the ECG features and HRV measures or passing the ECG signals to the CNN model. Second, relationship among different ECG components in adjacent heartbeats may be useful indicators for different heart diseases. Future work on this point may involve combining the inter-beat features with ECG features, HRV measures, and optimal DL features to investigate their effect on the classification performance. Third, more recent and advanced pre-trained CNN architectures can be investigated to enhance the

classification accuracy. However, the implementation of such recent CNNs is very challenging on low-power embedded devices or wearable devices for long-term mobile monitoring due to the huge computational requirements. The hardware design and implementation of the proposed ECG multi-class classification system can be a goal for future investigation. These points are challenges that will be explored and reported in the near future.

Funding Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB).

Declarations

Conflict of interest Ahmed Eltrass, Mazhar Tayel, and Abeer Ammar declare that they have no conflict of interest.

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

1. The WHO CVD Risk Chart Working Group (2019) World Health Organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions. *Lancet Glob Health* 7(10):e1332–e1345
2. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JG, Coats AJ, Falk V, González-Juanatey JR, Harjola VP, Jankowska EA, Jessup M (2016) ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: the task force for the diagnosis and treatment of acute and chronic heart failure of the European society of cardiology (ESC). Developed with the special contribution of the heart failure association (HFA) of the ESC. *Eur J Heart Fail* 18(8):891–975
3. Faezipour M, Saeed A, Bulusu SC, Nourani M, Minn L (2010) A patient adaptive profiling scheme for ECG beat classification. *IEEE Trans Inf Technol* 14(5):1153–1165
4. Kishi T (2012) Heart failure as an autonomic nervous system dysfunction. *J Cardiol* 59(2):117–122
5. I. Lee, D. Kim, S. Kang, and S. Lee (2017), "Ensemble deep learning for skeleton based action recognition using temporal sliding LSTM networks," *Proceedings of the IEEE international conference on computer vision, Venice, Italy*, pp 1012–1020.
6. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521(7553):436–444
7. Hinton GE, Osindero S, Teh YW (2006) A fast learning algorithm for deep belief nets. *Neural Comput* 18(7):1527–1554

8. Schmidhuber J (2015) Deep learning in neural networks: an overview. *Neural Network* 61:85–117
9. ChuDuc H, NguyenPhan K, NguyenViet D (2013) A review of heart rate variability and its applications. *APCBEE Proc* 7:80–85
10. Gulick D (1992) *Encounters with Chaos*. McGraw-Hill, New York, USA
11. Witkowski FX, Kavanagh KM, Penkoske PA, Plonsey R, Spano ML, Ditto WL, Kaplan DT (1995) Evidence for determinism in ventricular fibrillation. *Phys Rev Lett* 75(6):1230
12. R. B. Govindan, K. Narayanan, and M. S. Gopinathan (1998), “On the evidence of deterministic chaos in ECG: Surrogate and predictability analysis,” *Chaos: an Interdisciplinary Journal of Nonlinear Science*, 8(2), 495–502.
13. Stein KM, Walden J, Lippman N, Lerman BB (1999) Ventricular response in atrial fibrillation: random or deterministic? *Am J Phys-Heart Circ Phys* 277(2):H452–H458
14. Abarbanel HDI, Brown R, Kennel MB (1991) Lyapunov exponents in chaotic systems: their importance and their evaluation using observed data. *Int J Mod Phys B* 5(09):1347–1375
15. Grassberger P and Procaccia I (1983), “Measuring the strangeness of strange attractors,” *Physica D9*, 189–208.
16. Pecchia L, Melillo P, Sansone M, Bracale M (2010) Discrimination power of short-term heart rate variability measures for CHF assessment. *IEEE Trans Inf Technol Biomed* 15(1):40–46
17. Liu G, Wang L, Wang Q, Zhou G, Wang Y, Jiang Q (2014) A new approach to detect congestive heart failure using short-term heart rate variability measures. *PLoS ONE* 9(4):e93399
18. Cornforth DJ and Jelinek HF (2016), “Detection of congestive heart failure using Renyi entropy,” *IEEE computing in cardiology conference (CinC)*, Vancouver, Canada, 43, 669–672, 2016.
19. Chen W, Zheng L, Li K, Wang Q, Liu G, Jiang Q (2016) A novel and effective method for congestive heart failure detection and quantification using dynamic heart rate variability measurement. *PLoS ONE* 11(11):e0165304
20. Masetic Z, Subasi A (2016) Congestive heart failure detection using random forest classifier. *Comput Methods Programs Biomed* 130:54–64
21. Kumar M, Pachori RB, Acharya UR (2017) Use of accumulated entropies for automated detection of congestive heart failure in flexible analytic wavelet transform framework based on short-term HRV signals. *Entropy* 19(3):92
22. Wang Y, Wei S, Zhang S, Zhang Y, Zhao L, Liu C, Murray A (2018) Comparison of time-domain, frequency-domain and non-linear analysis for distinguishing congestive heart failure patients from normal sinus rhythm subjects. *Biomed Signal Process Control* 42:30–36
23. Hu B, Wei S, Wei D, Zhao L, Zhu G, Liu C (2019) Multiple time scales analysis for identifying congestive heart failure based on heart rate variability. *IEEE Access* 7:17862–17871
24. Isler Y, Narin A, Ozer M, Perc M (2019) Multi-stage classification of congestive heart failure based on short-term heart rate variability. *Chaos, Solitons Fractals* 118(1):145–151
25. Melillo P, Luca ND, Bracale M, Pecchia L (2013) Classification tree for risk assessment in patients suffering from congestive heart failure via long-term heart rate variability. *J Biomed Health Inform* 17:727–733
26. Shahbazi F, Asl BM (2015) Generalized discriminant analysis for congestive heart failure risk assessment based on long-term heart rate variability. *Comput Methods Programs Biomed* 122(2):191–198
27. Sultan Qurraie S, Ghorbani Afkhami R (2017) ECG arrhythmia classification using time frequency distribution techniques. *Biomed Eng Lett* 7(4):325–332
28. Ghorbani Afkhami R, Azarnia G, Tinati MA (2016) Cardiac arrhythmia classification using statistical and mixture modeling features of ECG signals. *Pattern Recognit Lett* 70:45–51
29. Elhaj FA, Salim N, Harris AR, Swee TT, Ahmed T (2016) Arrhythmia recognition and classification using combined linear and nonlinear features of ECG signals. *Comput Methods Programs Biomed* 127:52–63
30. Costa MD, Davis RB, Goldberger AL (2017) Heart rate fragmentation: a new approach to the analysis of cardiac interbeat interval dynamics. *Front Physiol* 8:255
31. Wu Z, Ding X, and Zhang G (2016), “A novel method for classification of ECG arrhythmias using deep belief networks,” *Int J Comput Intell Appl*, 15(4).
32. Acharya UR, Oh SL, Y. Hagiwara Y, Tan JH, Adam M, Gertych A, and Tan RS (2017), “A deep convolutional neural network model to classify heartbeats,” *Computers in Biology and Medicine*, 89, 389–396.
33. Alrahhhal MM, Bazi Y, Alhichri H, Alajlan N, Melgani F, Yager RR (2016) Deep learning approach for active classification of electrocardiogram signals. *Inf Sci* 345:340–354
34. Chen W, Liu G, Su S, Jiang Q, Nguyen H (2017) “A CHF detection method based on deep learning with RR intervals”, The 39th annual international conference of the IEEE engineering in medicine and biology society (EMBC). Jeju Island, Korea, pp 3369–3372
35. Potes C, Parvaneh S, Rahman A, Conroy B (2016) “Ensemble of feature based and deep learning-based classifier for detection of abnormal heart sounds”, *IEEE computing in cardiology conference (CinC)*. Vancouver, Canada, pp 621–624
36. Hwang B, You J, Vaessen T, Myin-Germeys I, Park C, Zhang BT (2018) Deep ECGNet: an optimal deep learning framework for monitoring mental stress using ultra short-term ECG signals. *Telemedicine and e-Health* 24(10):753–772
37. Pourbabaee B, Roshtkhari MJ, Khorasani K (2018) Deep convolutional neural networks and learning ECG features for screening paroxysmal atrial fibrillation patients. *IEEE Trans Syst, Man, Cybern: Syst* 48(12):2095–2104
38. Altan G (2017) Diagnosis of coronary artery disease using deep belief networks. *European J Eng Nat Sci* 2(1):29–36
39. Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, Liang J (2017) Convolutional neural networks for medical image analysis: full training or fine tuning? *IEEE Trans Med Imaging* 35(5):1299–1312
40. H. Azizpour, A. S. Razavian, J. Sullivan, A. Maki, and S. Carlsson, “Factors of transferability for a generic ConvNet representation,” *IEEE Trans Pattern Anal Mach Intell*, 38(9), 1790–1802.
41. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst* 25(2):1097–1105
42. Simonyan K and Zisserman A (2014), “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint <http://arxiv.org/abs/1409.1556>*.
43. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, and Rabinovich A (2015), “Going deeper with convolutions,” The 28th IEEE conference on computer vision and pattern recognition (CVPR), Boston, Massachusetts, USA, pp. 1–9
44. Kumar A, Kim J, Lyndon D, Fulham M, Feng D (2016) An ensemble of fine-tuned convolutional neural networks for medical image classification. *IEEE J Biomed Health Inform* 21(1):31–40
45. Salama MS, Eltrass AS and Elkamchouchi HM (2018), “An improved approach for computer-aided diagnosis of breast cancer in digital mammography,” 13th Annual IEEE international symposium on medical measurements and applications, Rome, Italy, 1–5, 2018.
46. Eltrass A, Salama M (2020) Fully automated scheme for computer-aided detection and breast cancer diagnosis using digitised mammograms. *IET Image Proc* 14(3):495–505

47. Ribas VJR, Wojdel A, Romero E, Ramos P, Brugada J (2016) ECG assessment based on neural networks with pre-training. *Appl Soft Comput* 49:399–406
48. Li Y, Zhang Y, Zhao L, Zhang Y, Liu C, Zhang L, Li Z, Wang B, Ng E, Li J (2018) Combining convolutional neural network and distance distribution matrix for identification of congestive heart failure. *IEEE Access* 6:39734–39744
49. Kaouter K, Mohamed T, Sofiene D, Abbas D, and Fouad M (2019) “Full training convolutional neural network for ECG signals classification,” In *AIP Conference Proceedings* (Vol. 2190, No. 1, p. 020055), AIP Publishing LLC, Athens, Greece, Dec. 2019.
50. Acharya UR, Fujita H, Oh SL, Hagiwara Y, Tan JH, Adam M, San Tan R (2019) Deep convolutional neural network for the automated diagnosis of congestive heart failure using ECG signals. *Appl Intell* 49(1):16–27
51. Wang L, Zhou W, Chang Q, Chen J, Zhou X (2019) Deep ensemble detection of congestive heart failure using short-term RR intervals. *IEEE Access* 7:69559–69574
52. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE (2000) Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation* 101(23):e215–e220
53. Moody GB, Mark RG (2001) The impact of the MIT-BIH Arrhythmia database. *IEEE Eng Med Biol Mag* 20(3):45–50
54. Baim DS, Colucci WS, Monrad ES, Smith HS, Wright RF, Lanoue A (1986) Survival of patients with severe congestive heart failure treated with oral milrinone. *J Am Coll Cardiol* 7(3):661–670
55. Ghanem N, Eltrass AS, Ismail NH (2018) “Investigation of EEG noise and artifact removal by patch-based and kernel adaptive filtering techniques, the 13th annual IEEE international symposium on medical measurements and applications (MeMeA). Italy, Rome
56. Eltrass AS, Ghanem N (2021) A new automated multi-stage system of Non-Local Means and multi-kernel adaptive filtering techniques for EEG noise and artifacts suppression. *J Neural Eng.* <https://doi.org/10.1088/1741-2552/abe397>
57. Tayel MB, Eltrass AS, Ammar AI (2018) A new multi-stage combined kernel filtering approach for ECG noise removal. *J Electrocardiol* 51:265–275
58. Stone M (1974) Cross-validatory choice and assessment of statistical predictions. *J R Stat Soc Ser B Methodol* 36(2):111–133
59. Pan J, Tompkins WJ (1985) A real-time QRS detection algorithm. *IEEE Trans Biomed Eng* 32(3):230–236
60. Eltrass AS, Tayel MB, Ammar AI (2012) A new automated CNN deep learning approach for identification of ECG congestive heart failure and arrhythmia using constant-Q non-stationary gabor transform. *Biomed Signal Process Control* 65:102326
61. Holighous N, Dorfler M, Velasco GA, and Grill T, “A framework for invertible, real-time constant-Q transforms,” *IEEE Trans Audio, Speech, and Lang Process*, 21(4).
62. Brown JC (1991) Calculation of a constant Q spectral transform. *J Acoustical Soc Am* 89(1):425–434
63. Yochum M, Renaud C, Jacquir S (2016) Automatic detection of P, QRS and T patterns in 12 leads ECG signal based on CWT. *Biomed Signal Process Control* 25:46–52
64. Remya RS, Indiradevi KP, Babu KA (2016) Classification of myocardial infarction using multi-resolution wavelet analysis of ECG. *Procedia Technol* 24(5):949–956
65. Nannaparaju V, Narasimman S (2015) Detection of T-wave alternations in ECGs by wavelet analysis. *Procedia Mater Sci* 10:307–313
66. Thomas M, Das MK, and Ari S (2015), “Automatic ECG arrhythmia classification using dual tree complex wavelet based features,” *AEU-Int J Electron Commun*, 69(4), 715–721
67. Lanfranchi PA, Colombo R, Cremona G, Baderna P, Spagnolati L, Mazzuero G, Wagner P, Perini L, Wagner H, Cavallaro C, Giannuzzi P (2006) Autonomic cardiovascular regulation in subjects with acute mountain sickness. *AJP Heart Circ Physiol* 289(6):H2364–H2372
68. Sallo ZG, Sallo MG (2016) Non-linear methods in HRV analysis. *Procedia Technol* 22:645–651
69. Hornero R, Abasolo D, Jimeno N, Sanchez C, Poza J, Aboy M (2006) Variability, regularity, and complexity of time series generated by schizophrenic patients and control subjects. *IEEE Trans Biomed Eng* 53(2):210–218
70. Porta A, Guzzetti S, Montano N, Furlan R, Pagani M, Malliani A (2001) Entropy, entropy rate, and pattern classification as tools to typify complexity in short heart period variability series. *IEEE Trans Biomed Eng* 48(11):1282–1291
71. Carrasco S, Caitán MJ, González R, Yáñez O (2001) Correlation among Poincaré plot indexes and time and frequency domain measures of heart rate variability. *J Med Eng Technol* 25(6):240–248
72. Rosenstein MT, Collins JJ, De Luca CJ (1993) A practical method for calculating largest Lyapunov exponents from small data sets. *Physica D* 65:117–134
73. Wolf A, Swift JB, Swinney HL, Vastano JA (1985) Determining Lyapunov exponents from a time series. *Physica D* 16(3):285–317
74. Cignetti F, Decker LM, Stergiou N (2012) Sensitivity of the Wolf’s and Rosenstein’s algorithms to evaluate local dynamic stability from small gait data sets. *Ann Biomed Eng* 40(5):1122–1130
75. Rispens SM, Pijnappels M, Van Dieën JH, Van Schooten KS, Beek PJ, Daffertshofer A (2014) A benchmark test of accuracy and precision in estimating dynamical systems characteristics from a time series. *J Biomech* 47(2):470–475
76. Awrejcewicz J, Krysko AV, Erofeev NP, Dobriyan V, Barulina MA, Krysko VA (2018) Quantifying chaos by various computational methods Part 1: simple systems. *Entropy* 203:175
77. Pritchard WS, Duke DW (1995) Measuring chaos in the brain: a tutorial review of EEG dimension estimation. *Brain Cogn* 27(3):353–397
78. Owis MI, Abou-Zied AH, Youssef AB, Kadah YM (2002) Study of features based on nonlinear dynamical modeling in ECG arrhythmia detection and classification. *IEEE Trans Biomed Eng* 49(7):733–736
79. Übeyli ED (2009) Detecting variabilities of ECG signals by Lyapunov exponent. *Neural Comput Appl* 18(7):653–662
80. Kiseľák J, Lu Y, Švihra J, Szépe P, Stehlik M (2020) SPOCU: scaled polynomial constant unit activation function. *Neural Comput Appl* 33(8):3385–3401
81. Van Molle P, Verbelen T, Vankeirsbilck B, De Vylder J, Diricx B, Kimpe T, Simoens P, Dhoedt B (2021) Leveraging the Bhattacharyya coefficient for uncertainty quantification in deep neural networks. *Neural Comput Appl.* <https://doi.org/10.1007/s00521-021-05789-y>.pp.1-17
82. Happy SL, Mohanty R, Routray A (2017) “An effective feature selection method based on pair-wise feature proximity for high dimensional low sample size data”, European signal processing conference (EUSIPCO). Kos, Greece, pp 1574–1578
83. Yu H, Yang J (2001) A direct LDA algorithm for high-dimensional data with application to face recognition. *Pattern Recogn* 34(10):2067–2070
84. Wu X, Kumar V, Quinlan JR, Ghosh J, Yang Q, Motoda H, McLachlan GJ, Ng A, Liu B, Philip SY et al (2008) Top 10 algorithms in data mining. *Knowl Inf Syst* 14(1):1–37

85. Rodriguez JJ, Kuncheva LI, Alonso CJ (2006) Rotation forest: a new classifier ensemble method. *IEEE Trans Pattern Anal Mach Intell* 28(10):1619–1630
86. Pohar M, Blas M, Turk S (2004) Comparison of logistic regression and linear discriminant analysis: a simulation study. *Metodoloski zvezki* 1(1):143
87. Antonogeorgos G, Panagiotakos DB, Priftis KN, and Tzonou A (2009), “Logistic regression and linear discriminant analyses in evaluating factors associated with asthma prevalence among 10-to 12-years-old children: divergence and similarity of the two statistical methods,” *Int J Pediatr*, vol. 2009, Article ID 952042.
88. Duda R, Hart E, Stork D (2000) *Pattern classification*. John Wiley & Sons, NY, USA
89. Çınar A, Tuncer SA (2021) Classification of normal sinus rhythm, abnormal arrhythmia and congestive heart failure ECG signals using LSTM and hybrid CNN-SVM deep neural networks. *Comput Methods Biomech Biomed Eng* 24(2):203–214
90. Kumari CU, Murthy ASD, Prasanna BL, Reddy MPP, and Panigrahy AK (2021), “An automated detection of heart arrhythmias using machine learning technique: SVM,” *Materials Today: Proceedings*, 24, part 2, pp.1393–1398.
91. Nahak S and Saha G (2020), “A fusion based classification of normal, arrhythmia and congestive heart failure in ECG,” In 2020 national conference on communications (NCC), IEEE, pp 1-6, Kharagpur, India
92. Krak I, Stelia O, Pashko A, Efremov M, and Khorozov O (2020), “Electrocardiogram classification using wavelet transformations,” In 2020 IEEE 15th International conference on advanced trends in radioelectronics, telecommunications and computer engineering (TCSET), IEEE, pp 930–933, Lviv-Slavske, Ukraine.
93. Kumari CU, Ankita R, Pavani T, Vignesh NA, Varma NT, Manzar MA, and Reethika A (2020), “Heart Rhythm Abnormality Detection and Classification using Machine Learning Technique,” In 2020 4th International conference on trends in electronics and informatics (ICOEI), 48184, IEEE, pp 580-584, Tirunelveli, India
94. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, Ng AY (2019) Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med* 25(01):65–69
95. Sannino G, De Pietro G (2018) A deep learning approach for ECG-based heartbeat classification for arrhythmia detection. *Futur Gener Comput Syst* 86:446–455
96. Mondéjar-Guerra V, Novo J, Rouco J, Penedo MG, Ortega M (2019) Heartbeat classification fusing temporal and morphological information of ECGs via ensemble of classifiers. *Biomed Signal Process Control* 47:41–48
97. Eltrass AS, Tayel MB, El-qady A (2021) Automatic epileptic seizure detection approach based on multi-stage quantized kernel least mean square filters. *Biomed Signal Process Control*. <https://doi.org/10.1016/j.bspc.2021.103031>
98. Raghu S, Sriraam N, Temel Y, Rao SV, Kubben PL (2020) EEG based multi-class seizure type classification using convolutional neural network and transfer learning. *Neural Netw* 124:202–212
99. Fadel W, Kollod C, Wahdow M, Ibrahim Y, and Ulbert I (2020), “Multi-class classification of motor imagery EEG signals using image-based deep recurrent convolutional neural network,” The 8th international winter conference on brain-computer interface (BCI), IEEE, pp 1–4, Gangwon, Korea (South).

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