Congestive heart failure waveform.pdf

by

Submission date: 16-Sep-2022 04:54PM (UTC+0700)
Submission ID: 1901201082
File name: Congestive heart failure waveform.pdf (4.44M)
Word count: 6786
Character count: 34493

Informatics in Medicine Unlocked 21 (2020) 100441

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: http://www.elsevier.com/locate/imu

Congestive heart failure waveform classification based on short time-step analysis with recurrent network

Annisa Darmawahyuni, Siti Nurmaini *, Meiryka Yuwandini, Muhammad Naufal Rachmatullah, Firdaus Firdaus, Bambang Tutuko

Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Palembang, 30139, Indonesia

ARTICLE INFO

Keywords:
Congestive heart failure
Normal sinus rhythm
Recurrent neural networks
Long short-term memory

ABSTRACT

Congestive heart failure (CHF) is characterized by the heart's inability to pump blood adequately throughout the body without increased intracardiac pressure. Diverse approaches are used to treat CHF. These approaches, which include physical examination, echocardiography, and laboratory testing, require a high degree of competence to interpret findings and make diagnoses. Moreover, existing methods do not account for the relationships between variables and thus provide limited performance. Electrocardiogram (ECG), as a non-invasive test, may be used for CHF early diagnosis, which would require further examination to be referred. A previous study revealed a significant correlation between heart failure (HF) and ECG features. However, the method was only performed on small, balanced data; then, the features must be derived from trial and error. The current paper proposes deep-learning techniques—recurrent neural networks (RNNs) with long short-term memory (LSTM) architectures—to create a diagnostic algorithm that achieves high accuracy with limited information and automated feature extraction. The ECG signals used in this study were obtained from the public PhysioNet databases. We fine-tuned the hyperparameters of 24 LSTM models to obtain the best model. Moreover, ECG signal segmentation was compared among the first five and fifteen minutes as features. Out of the 24 LSTM models, the model with the first fifteen minutes of ECG signals (model 1) obtained the highest accuracy, sensitivity, specificity, precision, and F1-score (99.86%, 99.85%, 99.85%, 99.87%, and 99.86%, respectively). The first fifteen minutes of ECG signals performed well because the LSTM model learned an increasing number of features. In conclusion, the proposed LSTM model could give a clinician a preliminary CHF diagnosis for further medical attention. Deep learning can be a useful predictive method for increasing the number of identified CHF patients.

1. Introduction

Congestive heart failure (CHF) is a chronic disease in which the heart fails to maintain the blood circulation adequately. CHF is a global pandemic affecting about 26 million people [1]. CHF is also a significant health concern in Asia, where its prevalence (6.7%) is much higher than that in Western countries [2]. The diagnosis of CHF is a challenging task because it is affected by reduced ejection fraction (HFrEF), preserved ejection fraction (HFpEF), and mid-range ejection fraction (HFmrEF). In the normal heart, there is good stroke volume (blood flow volume ejected per heartbeat), and oxygen-rich blood from the left ventricle is pumped into the body. In CHF, the volume of the stroke decreases, and the heart cannot effectively pump oxygen-rich blood to the rest of the body [3]. The New York Heart Association (NYHA) classifies CHF into four categories. There are significant symptoms only in patients in classes 3 and 4, namely, patients with insufficient physical activity experience signs of heart failure or while resting, respectively [4].

Several methods are used to diagnose CHF patients, such as chest X-ray, nuclear imaging, magnetic resonance imaging (MRI), invasive angiography, and echocardiography [5]. Echocardiography is the most commonly used test; it uses ultrasound to measure the volume of the stroke, the end of the diastolic volume, and the ratio of these quantities (ejection fraction) [6]. However, it can be time-consuming and costly [5, 7], and it is highly operator dependent [5]. Another method for detecting CHF is the use of an electrocardiogram (ECG) [5]. ECG is inexpensive, widely accessible, and can be the most useful instrument in

* Corresponding author. Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Palembang, 30139, Indonesia.
E-mail addresses: riset.annisadarmawahyuni@gmail.com (A. Darmawahyuni), siti_nurmaini@unsri.ac.id, sitinurmaini@gmail.com (S. Nurmaini), meiryka17@gmail.com (M. Yuwandini), naufalrachmatullah@gmail.com (Muhammad Naufal Rachmatullah), virdauz@gmail.com (F. Firdaus), bambangtutuko60@gmail.com (B. Tutuko).

https://doi.org/10.1016/j.imu.2020.100441
Received 27 July 2020; Received in revised form 29 September 2020; Accepted 30 September 2020
Available online 7 October 2020

the diagnosis and prognosis of CHF patients. However, the manual examination of ECG signals by experts is complicated and subject to intra- and interobserver variabilities [8]. Deciphering minute shifts in ECG signals is difficult because their amplitude is measured in millivolts [5]. A well-designed computer-aided detection (CAD) system for CHF based on ECG signals is needed to overcome these problems.

Previously, CAD with machine-learning (ML) algorithms used ECG signals for CHF diagnosisand quantitative assessment for informed decision-making. Melilo et al. [9] first attempted to determine CHF disease severity using long-term heart rate variability (HRV) measurements. They developed a classification and regression tree (CART) as a classifier and achieved a 93.3% sensitivity and 63.6% specificity. In some cases, static HRV measurements might not fully quantify trend changes in CHF patients' autonomic activity during different daily activities [10]. Thus, Orhan et al. [11] presented the first application of the equal frequency in amplitude and equal width in time (EFiA-EWiT) approach to discriminate CHF and normal sinus rhythm (NSR) patterns in ECG signals. They extracted the best representative features using the primitive and fast linear regression prediction methods. The proposed method achieved accuracy, sensitivity, and specificity rates of 100%, 99.36%, and 99.30%, respectively. Masetic et al. [12] proposed the C4.5 decision tree method for creating a model that will detect and separate normal heart and CHF in a long-term ECG time series. The proposed method obtained 100% for both sensitivity and specificity and 99.20% for accuracy. Kamath et al. [13] performed the detrended fluctuation analysis (DFA) approach to calculate normal heart and CHF short-term (20 s) ECG segments. The approach obtained average sensitivity and specificity rates of 98.4% and 98%, respectively. Vidya et al. [14] proposed dual-tree complex transform wavelets (DTCWT) for the automated detection of CHF from standard ECG signals. They performed DTCWT on two seconds ECG segments up to six levels to obtain coefficients. Using 45 features of the ECG signal, the proposed method achieved accuracy, sensitivity, and specificity rates of 99.86%, 99.78%, and 99.94%, respectively. Isler et al. [15] classified CHF using short-term HRV analysis based on three-stage classifiers, namely, k-nearest neighbors (KNN), linear discriminant analysis (LDA), multilayer perceptron (MLP), SVM, and radial basis function artificial neuronal network (RBF ANN). They conducted their analysis using short-term (five minutes) and long-term (24 hours) HRV data. The five minutes HRV data obtained accuracy, sensitivity, and specificity rates of 98.8%, 100%, and 98.1%, respectively. Therefore, short waveform classification can be performed instead of examining long-term trends for CHF.

According to these previous studies, short waveforms of CHF can be classified instead of looking at long waves. The proposed methods, such as SVM, obtained promising CHF classification results. With its versatility, SVM performed well with small data pools (where there is no overfitting), and it captured nonlinearity well in features with its kernel function. However, only small and balanced data perform well in ML techniques. In reality, there is more to the existence of normal data that contribute to data imbalance. Furthermore, ML requires hand-crafted features; that is, these features must be derived from trial and error to achieve optimal classification accuracy. For useful and accurate classification, it is necessary to determine which visual features best represent the ECG signal for extraction. Therefore, ECG signal classification by automated feature extraction should be examined further.

Deep-learning (DL) algorithms have been explored to overcome the abovementioned issues. DL is an ML approach where a network automatically learns and picks up distinct features based on ECG input signals [16,17]. Recurrent neural networks (RNNs) are a form of DL that has been widely employed in speech recognition [18] and DNA sequencing [19], and it is receiving plenty of attention in the medical field [20]. Recently, researchers used RNNs models to develop CAD systems for the diagnosis of various medical conditions [21–24]. RNNs are a neural network where the output from the previous step is fed to the current step as the input. An RNN architecture consists of long short-term memory (LSTM) and a gated recurrent unit (GRU). However, in some cases, LSTM performs better than GRU [23]. A single LSTM unit consists of a cell, an input gate, an output gate, and a forget gate, enabling the cell to hold values for unspecified periods. These gates track information flow into and out of the LSTM cell [23,24]. Certain DL algorithms have been studied for CHF classification, such as convolutional neural networks (CNNs) [3] and sparse autoencoders (SAEs) [25]. However, when applied to ECG, a CNN sometimes cuts the window size of a fixed length, thereby reducing classification performance [26]. Still, with such an SAE, intervals have potential in CHF detection but cannot entirely reflect dynamic changes in 24 hours [25]. RNNs can address this shortcoming for sequential prediction to model the flow of time directly. Hence, the accurate automatic classification of short-term ECG signals using RNNs with LSTM architectures is desirable.

This paper proposes an RNNs with LSTM architecture for CHF and NSR classification. The well-known PhysioNet database is used for classification based on short waveforms, namely, the first five and fifteen minutes of ECG signals. Both times are compared and employed as a feature for the classifier without considering static or dynamic HRV. Furthermore, 24 LSTM models are fine-tuned to obtain the best model performance, as indicated by accuracy, sensitivity, specificity, precision, and F1-score. Hence, the design and development of an automated algorithm for CHF and NSR classification via DL need to be investigated in detail. The contributions and novelties of this paper are as follows.

- We generate a DL model with a short ECG waveform for automated long-term CHF.
- We propose an RNN with LSTM architecture and specific time features.
- We evaluate the proposed model using one-to four seconds to determine the window size with optimal efficiency for CHF classification.

2. Materials and methods

This paper proposes an ECG signal processing method that defines and learns feature representations directly from the Beth Israel Deaconess Medical Center Congestive Heart Failure (BIDMC CHF) and MIT-BIH Normal Sinus Rhythm (MIT-BIH NSR) databases. The ECG signal processing consists of data preparation and preprocessing; (i) ECG noise removal using discrete wavelet transform (DWT); (ii) ECG signal segmentation based on the first five and fifteen minutes of the signal; (iii) ECG signal normalization, and classification based on the RNN with an LSTM architecture. Finally, model performance was evaluated using accuracy, sensitivity, specificity, precision, and F1-score (Fig. 1).

2.1. ECG signal dataset

ECG signal raw data were obtained from the BIDMC CHF and MIT-BIH NSR databases. The BIDMC CHF (NYHA classes 3–4) database consists of long-term ECG recordings in two ECG signals are each about 20 hours in duration from 15 subjects. The NSR database includes 10 ECG recordings that were referred to the Arrhythmia Laboratory and found to have had no significant arrhythmias. Both databases have a frequency sampling of 250 Hz. The ECG sample records of CHF and NSR are presented in Fig. 2.

2.2. Noise removal

During acquisition, ECG signals become contaminated due to different forms of objects and interferences, such as muscle contraction, baseline drift, electrode touch noise, and power line interference [27–29]. In initial processing, ECG signals must be improved by eliminating various kinds of noise and artifacts. ECG preprocessing is a significant task performed to remove noise so that abnormalities in heart condition can be easily interpreted. DWT is widely used to analyze

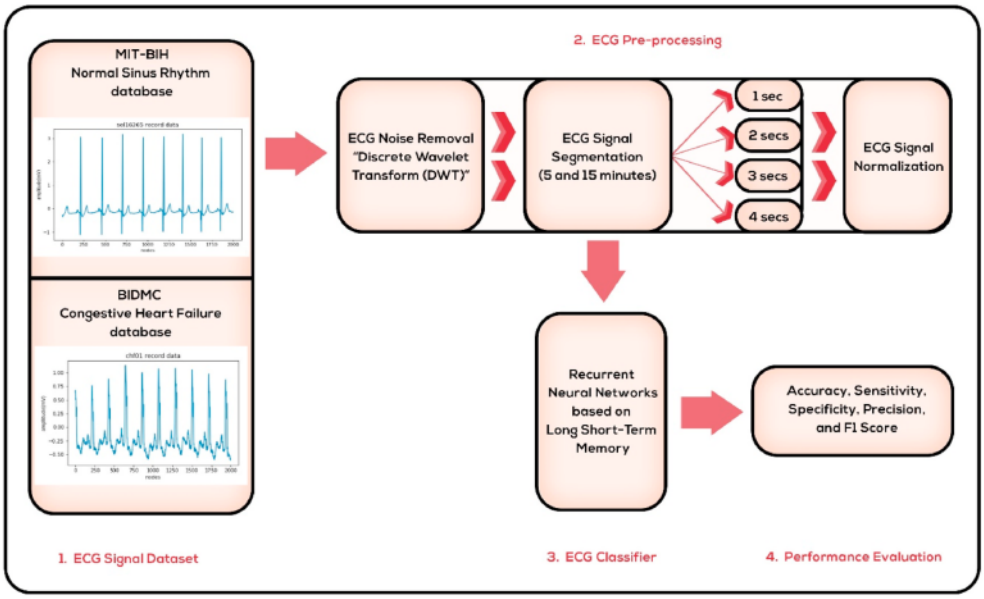
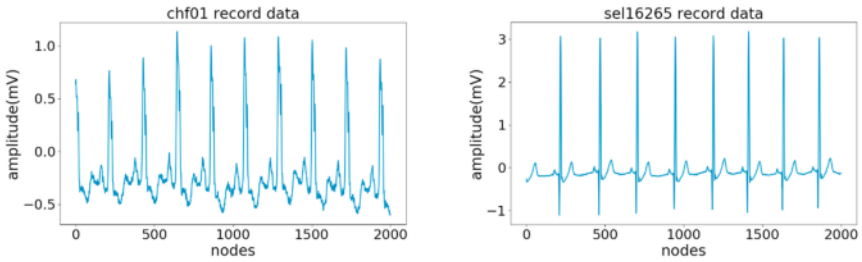


Fig. 1. The workflow of CHF classification process in ECG signal processing.



(a) Congestive Heart Failure (b) Normal Sinus Rhythm

Fig. 2. ECG sample records of BIDMC CHF and MIT-BIH NSR databases.

nonstationary signals, that is, ECG signal denoising [30,31]. DWT is realized by passing the signal, where $x(n)$ is the discrete input signal with length n, through a series of low-pass ($g[n]$) and high-pass filters ($h[n]$). DWT is applied to analyze the signals at different resolution levels; the number of decomposition levels is decided wavelet coefficients to do a series of signal processing [32]. Denoising performance is evaluated using the signal-to-noise ratio (SNR or S/N). SNR gives information about the quality of a signal. The input SNR (SNR_i) is defined as,

$$SNR_i = 10\log_{10}\left[\frac{\sum_n x^2(n)}{\sum_n r^2(n)}\right] \quad (1)$$

The output SNR (SNR_o) is given by the following equation:

$$SNR_o = 10\log_{10}\left[\frac{\sum_n x_d^2(n)}{\sum_n x_d n) - x(n))^2}\right] \quad (2)$$

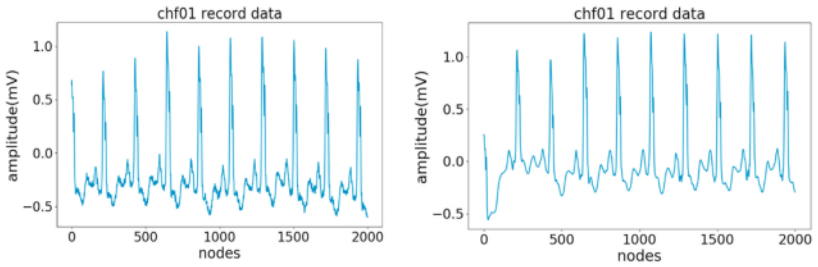
where $x(n)$ is the original signal, $r(n)$ is the added noise signal, and $x_d(n)$ is the denoised signal.

A wavelet denoising algorithm mainly contains three steps: wavelet decomposition, coefficient processing, and wavelet reconstruction [33]. Common wavelet families, such as Daubechies, biorthogonals, coiflets, and symlets, can be used for ECG signal denoising [34]. This study applies the symlet wavelet sym5 because it is good at denoting noisy ECG signals (Fig. 3 (a)). Fig. 3 presents the wavelet denoising algorithm, which assumes that the "clear" signal is correlated with some of the decomposed coefficients. The other signals are related to the mean noise value. Thus, with the unimportant noise-related coefficient removed, the signal can be restored without loss of data (Fig. 3 (b)).

In this study, we obtain the SNR value of each record for the BIDMC HF and NSR databases (Table 1). Table 1 presents the SNR values of sym5 for the BIDMC and NSR databases. As we can see in the table, the average SNR values of BIDMC and NSR are −0.001663 and −0.000271 decibel (dB), respectively. A ratio greater than 0 dB or higher than 1:1 indicates that there is more signal than noise. For the BIDMC HF and NSR databases, there are nine and seven records, respectively, that have SNR values greater than 0. Therefore, sym5 can deal with ECG signal noise well.

2.3. ECG signal segmentation

For the ECG signal in this study, the first five- and fifteen minutes recordings of BIDMC CHF are used as features because the dataset has a



(a) ECG signal before denoising (b) ECG signal after denoising using sym5

Fig. 3. The sample of ECG noise removal processing for CHF classification in BIDMC CHF databases.

Table 1
The SNR value of BIDMC HF and NSR databases.

Records	SNR (sym5)	
	BIDMC	NSR
1	−0.040526	−0.003116
2	0.001411	0.000835
3	−0.006769	0.002295
4	0.001104	0.000486
5	−0.002702	0.001657
6	−0.000363	−0.001300
7	0.000191	−0.006883
8	0.000303	0.001965
9	−0.001464	0.000338
10	0.001843	0.001015
11	0.016934	
12	0.001490	
13	−0.002656	
14	0.000256	
15	0.005998	
Average	**−0.001663**	**−0.000271**

long record. The ECG signals are segmented rhythm by rhythm from one to four seconds (Fig. 4). For the NSR, because the data are not the same length as that of the BIDMC CHF database, a zero-padding process is performed. Its functions are used to increase the signal length, which is lacking by considering the maximum signal length as a reference with zero value (0).

After ECG signal segmentation, normalization is used to scale the data of an attribute in a smaller range with a normalizing bound. This is required when dealing with characteristics on different scales. Thus, the ECG signal is normalized to bring all the attributes on the same scale. The normalizing bound changes the values of the lower limit (lb) and upper limit (ub) on the amplitude of a signal to the desired range without changing the pattern or shape of the signal itself. In this study, the data are acquired in the preprocessing with a lower limit of 0 and an upper limit of 1, respectively. The mathematical function of the normalization with the normalizing bound is as follows:

$$f(x) = x^{*}coef - (x_{mid} * coef) + mid \tag{3}$$

where

$$coef = \left(\frac{ub - lb}{x_{\max} - x_{\min}}\right) \tag{4}$$

$x_{mid}$is the midpoint of the input signal:

$$x_{mid} = x_{\max} - \frac{x_{\max} - x_{\min}}{2} \tag{5}$$

$x_{\max}$ is the peak point of the input signal, $x_{\min}$ is the lower point of the input signal, and mid is the midpoint of the specified limit:

$$mid = ub - \frac{ub - lb}{2} \tag{6}$$

2.4. Recurrent neural networks

RNNs are a type of artificial neural network commonly designed to recognize the characteristics of sequential data [23,24]. RNNs as a part of DL architecture due to the automatic process of calculating features without determining some appropriate features. RNNs use patterns to predict the next likely scenario; its layers use for loop to iterate over the timesteps of a sequence [35]. RNNs encounter a problem caused by its iterative nature; its gradient is substantially equal to the recurrent weight matrix raised to a high power [23,24]. The architecture of RNNs consist of LSTM and GRU, which are implemented to solve the problem of the gradient with the gating mechanism, which controls how internal states are retained or discarded. Among them, LSTM architecture has been proposed, due to it achieved excellent performance in our previous study [23,24].

LSTM was introduced by Hochreiter and Schmidhuber [36]. In our previous work [23], LSTM captured an extended context with input or forget gates by passing gradients and showed outstanding results. Each node in LSTM is a cell that comprises input, forget, and output gates. Common problems of standard RNNs are caused by iterative behavior; their gradient is significantly equal to the recurrent weight matrix elevated to high power. The gradients increase or decrease at rates that are exponential in terms of the number of timesteps. With the gating mechanism, which controls how internal states are retained or discarded, LSTM overcomes the gradient problem [23]. Mathematically, LSTM can be written as follows:

$$c_t = \sigma(W_f I_t) c_{t-1} + \sigma(W_i I_t)\tanh(W_{in} I_t) \tag{7}$$

$$h_t = \sigma(W_o I_t)\tanh(c_t) \tag{8}$$

where $\sigma$is a sigmoid function, $c_t \in R^N$is a column vector, and $I_t \in R^{(M+N)}$ is a concatenation of the current input, $X_t \in R^M$, and the previous output, $h_{t-1} \in R^N$. Under the assumption $c_o = 0$, the hidden state vector of LSTM can be derived by:

$$c_t = \sum_{k=1}^{t}\left[\prod_{j=k+1}^{t}\sigma(W_f I_j)\right]\sigma(W_i I_k)\tanh(W_{in} I_k) \tag{9}$$

The comparison of output RNNs and the gating mechanism in LSTM is given by:

$$h_t = \tanh\left(\sum_{k=1}^{t} W_e^{t-k} W_{in} X_k\right) \tag{10}$$

$$h_t = \sigma\left(W_o I_t\right)\tanh\left(\sum_{k=1}^{t}\left[\prod_{j=k+1}^{t}\sigma(W_f I_j)\right]\sigma\left(W_i I_k\right)\tanh\left(W_{in} I_k\right)\right) \tag{11}$$

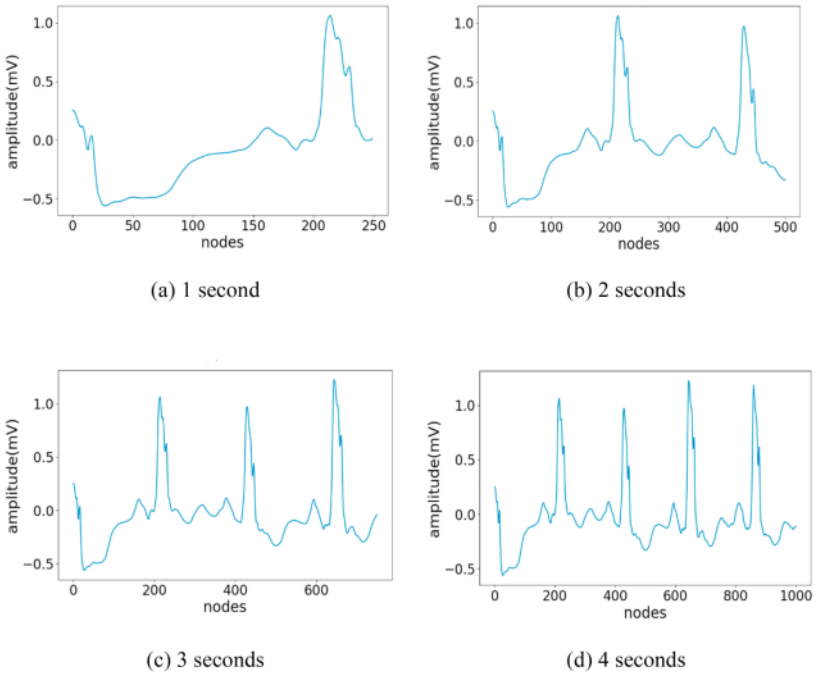


Fig. 4. The ECG signal segmentation samples from 1 to 4 seconds.

where W_i, W_f, W_o, W_{in} are weight matrices for the input gate, forget gate, output gate, and input, respectively. The total parameters in LSTM are equal to $4 \times (n^2 + nm + n)$, where n-dimensions are assumed as the cell state and m-dimensions are the input signal.

3. Results and discussion

The data are grouped and resampled for training and testing sets, namely, 80% for the training set and the rest for the testing set. In the proposed model, the LSTM input layer must have three dimensions, namely, samples, timesteps, and features. Here, we use the first five and fifteen minutesof BIDMC CHF; for the NSR database, we add zero padding. With a frequency sampling of 250 Hz, 1 second equals 250 nodes. Therefore, the first five and fifteen minutes are equal to 75,000 and 225,000 nodes, respectively. From both nodes, we use the segmentation of one to 4 seconds as a feature with timesteps of 1 second. The hyperparameters of 24 models are fine-tuned to generate the LSTM model (Table 2). Table 2 shows all the hyperparameters of the 24 LSTM

Table 2
The hyper-parameters tuning of 24 LSTM models.

Model	Data	Segmentation	Total Hidden Layer	Nodes Input Layer	Nodes Hidden Layer	Timesteps	Epochs
1	15 minutes	1 second	1	250	100	1	139
2	5 minutes		1	250	100	1	172
3	15 minutes	2 seconds	1	500	100	1	110
4	5 minutes		1	500	100	1	120
5	15 minutes	3 seconds	1	750	100	1	150
6	5 minutes		1	750	100	1	311
7	15 minutes	4 seconds	1	1000	100	1	174
8	5 minutes		1	1000	100	1	166
9	15 minutes	1 second	2	250	50	1	91
10	5 minutes		2	250	50	1	139
11	15 minutes	2 seconds	2	500	50	1	138
12	5 minutes		2	500	50	1	72
13	15 minutes	3 seconds	2	750	50	1	123
14	5 minutes		2	750	50	1	89
15	15 minutes	4 seconds	2	1000	50	1	151
16	5 minutes		2	1000	50	1	77
17	15 minutes	1 second	3	250	25	1	85
18	5 minutes		3	250	25	1	99
19	15 minutes	2 seconds	3	500	25	1	80
20	5 minutes		3	500	25	1	106
21	15 minutes	3 seconds	3	750	25	1	91
22	5 minutes		3	750	25	1	90
23	15 minutes	4 seconds	3	1000	25	1	89
24	5 minutes		3	1000	25	1	88

models, namely, eight as the batch size, Adam as the optimizer (learning rate = 0.001), and binary cross-entropy as the loss function. Moreover, all the models are fine-tuned with different hidden layers and nodes. For epochs, we use early stopping to obtain the best iteration results. Table 2 presents the variant epochs, where the minimum and maximum epochs are 72 (model 12) and 311 (model 6), respectively.

Table 2 also presents the 24 LSTM models in the first five and fifteen minutes of the ECG signal. For all hidden layers of the LSTM, 250, 500, 750, and 1000 nodes of the input layer are used. We decrease the total nodes of the hidden layer from one to three, 100, 50, and 25, respectively. The results of the 24 LSTM models are listed in Table 3. The table shows that the higher the segmented signal, the lower the accuracy of the results obtained for one hidden layer of the LSTM (models 1–8). For instance, for both the first five and fifteen minutes, the accuracy in 1 second was 99.86% and 99.66%, respectively. In 2 and 3 seconds, the accuracy was 99.68% and 99.06%, and 99.66% and 98.20%, respectively. However, in 4 seconds, the accuracy decreases—only 99.28% and 96.26%—for the first fifteen and five minutes of ECG signal. For two hidden layers, the accuracy of the first fifteen minutes in 1 second decreases from 99.86% to 99.84%. Then, an additional decrease is observed in three hidden layers, where the accuracy is only 99.80%. It also applies to the other metric measurements, namely, sensitivity, specificity, precision, and F1-score. Therefore, the addition of hidden layers does not improve the result. Then, we compare the first five and fifteen minutes of the ECG signal; the accuracy, sensitivity, specificity, precision, and F1-score for the first fifteen minutes are higher than those for the first five minutes. This is because the LSTM model learns additional features. Among the 24 LSTM models, model 1, which covers the first fifteen minutes, yields the best performance; it has the highest accuracy, sensitivity, specificity, precision, and F1-score of 99.86%, 99.85%, 99.85%, 99.87%, and 99.86%, respectively. Therefore, we propose model 1 as the best model for CHF and normal heart ECG signal classification.

From the results in Table 3, the accuracy, sensitivity, specificity, precision, and F1-score values of models 1, 3, 5, 9, and 17 are above 99.60%. In this study, the accuracy and loss curves are presented to validate five of the best LSTM models (Fig. 5). Fig. 5 (a)–(j) show that the error is decreased and the accuracy increased from the training (red line) and testing data (blue line) with the increase in epochs. All curves are presented as a well-fit model. The plot of the testing accuracy and loss toward a point of stability shows a small gap between the training accuracy and loss. All five of the best LSTM models show promising results for CHF classification in fifteen minutes of the ECG signal. However, model 1 performs better than do other models. As seen in Fig. 5 (a), the accuracy in the initial epoch is above 93%, and it then increases to perfect performance. Moreover, the error decreases, becoming almost close to zero (0).

Table 3
The results of 24 LSTM models hyper-parameters tuning.

Model	Performance Evaluation (%)				
	Accuracy	Sensitivity	Specificity	Precision	F1-Score
1	99.86	99.85	99.85	99.87	99.86
2	99.66	99.50	99.50	99.64	99.57
3	99.68	99.62	99.62	99.72	99.67
4	99.06	98.81	98.81	99.21	99.03
5	99.66	99.60	99.60	99.69	99.65
6	98.20	98.14	98.14	98.19	98.17
7	99.28	99.20	99.20	99.33	99.26
8	96.26	96.21	96.21	96.21	96.21
9	99.84	99.81	99.81	99.86	99.83
10	99.46	99.31	99.31	99.54	99.42
11	99.60	99.52	99.52	99.64	99.58
12	98.53	98.54	98.54	98.35	98.45
13	98.60	98.39	98.39	98.66	98.52
14	95.80	95.65	95.65	95.71	95.68
15	98.66	98.56	98.56	98.70	98.63
16	96.26	96.27	96.27	96.13	96.20
17	99.80	99.75	99.75	99.82	99.79
18	99.46	99.34	99.34	99.55	99.44
19	99.64	99.55	99.55	99.70	99.62
20	98.66	98.66	98.66	98.55	98.60
21	99.13	98.98	98.98	99.19	99.08
22	95.20	95.16	95.16	94.86	95.00
23	98.57	98.40	98.40	98.66	98.53
24	94.40	94.53	94.53	93.80	94.13

Precision-recall (PR) curves are shown in Fig. 6 (a) to present detailed information about model 1, which is the proposed model for CHF and NSR classification. PR curves are typically generated from a confusion matrix to evaluate model performance on a given dataset [37]. Such curves, with precision plotted on the y-axis and recall on the x-axis, could expose differences between algorithms. They also summarize the trade-off between the positive predictive value (PPV, precision) and the true-positive rate (TPR, recall/sensitivity) of a predictive model. A curve that lies above another curve has a better performance level. This study also plots the TPR (sensitivity) versus the false-positive rate (1-specificity) across varying cutoffs. A curve is generated in the unit square, and it is called the receiver operating characteristic (ROC) curve. ROC evaluation is a standard approach for analyzing the reliability of medical diagnostic systems. The plot of the ROC curve for CHF and NSR classification is presented in Fig. 6 (b). Both curves obtain excellent visualization because their values are close to 1.

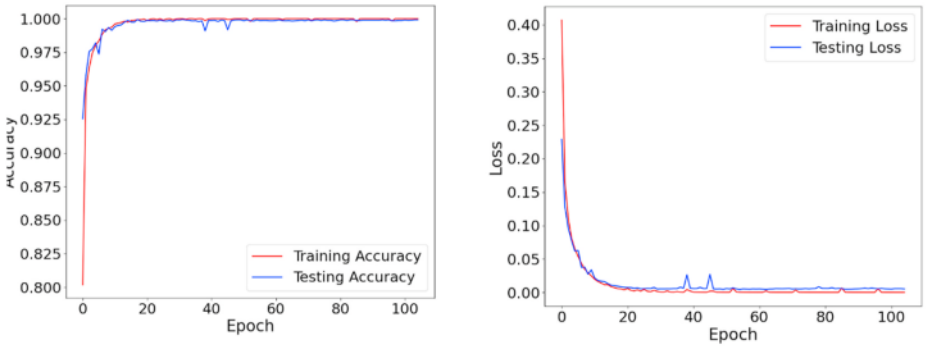
Some previous studies explored CHF classification using DL techniques, such as autoencoders, CNNs, and LSTM (refer to Table 4). Chen et al. [25] constructed a CHF detection model based onan SAE-based DL algorithm. SAE was applied to learn features from raw unsupervised RR data intervals automatically. An SAE was then trained to build a model that would discriminate against CHF. The DL algorithm first extracts unsupervised features from raw RR intervals using an SAE and then constructs a deep neural network model with various combinations of hidden nodes. Results showed that the model reached an accuracy of 72.41%. Therefore, RR intervals have potential in CHF detection but cannot completely represent dynamic changes in 24 hours. Acharya et al. [3] presented an 11-layer deep CNN model for CHF diagnosis. This proposed CNN model requires minimal preprocessing of ECG signals, and it does not need engineered features or classification. They generated the structure of the CNN model for four sets (A–D). Among the four sets, set B (using the BIDMC CHF database and the Fantasia database [FD]) achieved the maximum accuracy, specificity, and sensitivity of 98.97%, 99.01%, and 98.87%, respectively. However, there are limitations in the use of CNNs; when applied to ECG, a CNN cuts the window size of a fixed length, thereby eventually reducing classification performance. Wang et al. [6] combined LSTM and an inception module from GoogLeNet for CHF detection. Three RR segment-length forms (N = 500, 1000, and 2000) were used for comparison with other studies. They divided their database into two, namely, databases 1 and 2. The results showed that database 1 outperformed database 2. Using database 1 (BIDMC CHF, NSR, and FD), the proposed method can detect CHF through a short-term assessment of the heartbeat; the method achieved 99.22%, 98.85%, and 98.92% accuracy in a blindfold validation. LSTM performs well; however, in this case, excellent results were obtained without considering HRV analysis for CHF classification.

Although our results look promising, our study has certain limitations, such as the following.

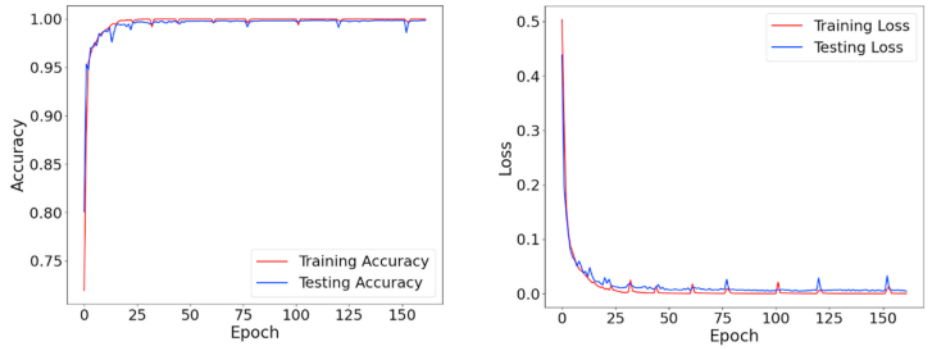
- We did not classify the CHF patients based on classes 1–4 and ignored the influence of HFrEF, HFpEF, and HFmrEF.
- We used a limited dataset to generate the model and did not validate the proposed model against hospital patient data.

4. Conclusions

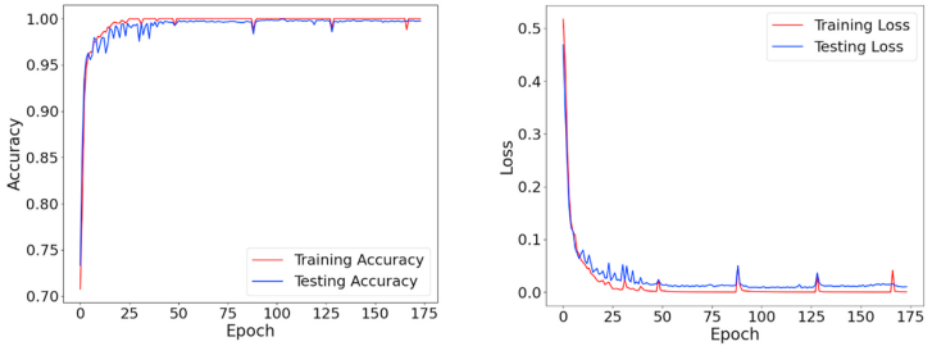
CHF is a complex clinical disease characterized by reduced heart pumping and blood-filling capacities. The use of ECG classification for CHF detection is challenging because ECG waveforms require long- or



(a) Accuracy Curve Model 1 (b) Loss Curve Model 1



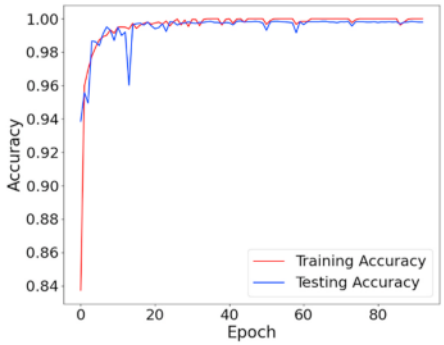
(c) Accuracy Curve Model 3 (d) Loss Curve Model 3



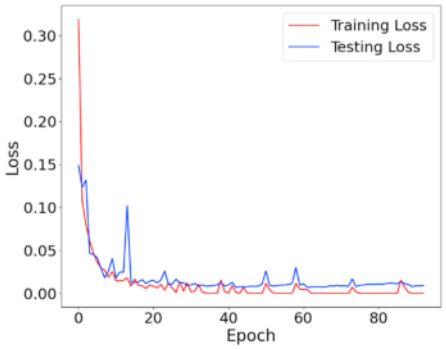
(e) Accuracy Curve Model 5 (f) Loss Curve Model 5

Fig. 5. The accuracy and loss curve of the five best LSTM models.

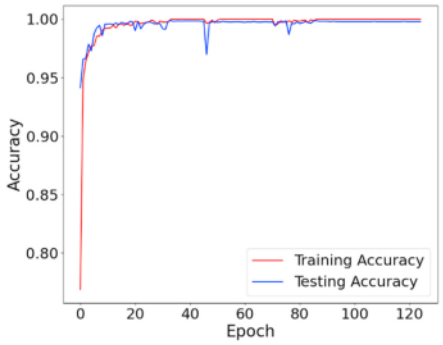
short-term RR interval data. In summary, this paper proposes a DL technique—RNN with LSTM architecture—to distinguish normal heart and CHF ECG signals. Using the first five and fifteen minutes of BIDMC CHF and NSR data as features, this paper compares the segmentation of ECG signals from one to four seconds for the learning process. The experiments show that the first fifteen minutes of both ECG signals yield better performance than the first five minutes, and one second ECG signal segmentation shows better performance than the other time segmentations. Moreover, we fine-tune the parameters of the hidden layers. The addition of more layers of LSTM does not produce better performance. We generate 24 LSTM models to achieve the best performance, and model 1 is the best of all models. The proposed LSTM



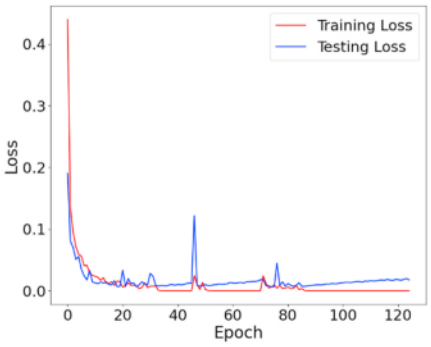
(g) Accuracy Curve Model 9



(h) Loss Curve Model 9

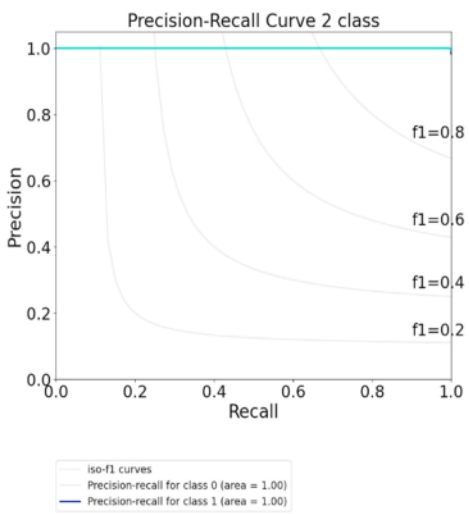


(i) Accuracy Curve Model 17

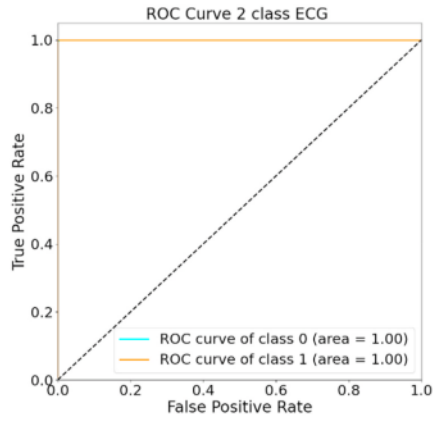


(j) Loss Curve Model 17

Fig. 5. (*continued*).



(a) Precision-Recall Curve



(b) ROC Curve

Fig. 6. The precision-recall and ROC curve in model 1 LSTM as the proposed model.

Table 4
The benchmark study of CHF classification.

Author	Method	Metrics (%)				
		Accuracy	Sensitivity	Specificity	Precision	F1-Score
Chen et al. [25]	A sparse auto-encoder (SAE)	72.41	–	–	–	–
Acharya et al. [3]	CNNs	98.97	98.87	99.01	–	–
Wang et al. [6]	LSTM	99.22	99.22	99.72	–	–
This paper	**LSTM**	**99.86**	**99.85**	**99.85**	**99.87**	**99.86**

architecture in model 1 consists of one hidden layer, a 250-node input layer, a 100-node hidden layer, one timestep, eight batch size, and 139 epochs (using early stopping). The proposed method exhibits good classification performance, yielding accuracy, sensitivity, specificity, precision, and F1-score values of 99.86%, 99.85%, 99.85%, 99.87%, and 99.86%, respectively. In future work, we will generalize the proposed LSTM model to classify other heart abnormalities using several databases. Our proposed model gives preliminary diagnoses for clinicians for further medical treatment. We conclude that the DL technique can be a useful predictive method for increasing the identification rate of CHF patients in ECG short-term recordings.

Author contributions

A.D. Formal analysis, resources, and writing—original draft; S.N. Conceptualization, funding acquisition, supervision, and writing—review & editing; M.Y. Software, Data curation; M.N.R. Software and writing—review & editing; F·F.Writing—review & editing; B.T. Writing—review & editing.

Funding

This work was funded by Hibah Profesi No. SP DIPA-023.17.2.677515/2020, Universitas Sriwjaya, Indonesia.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.imu.2020.100441.

References

[1] Ponikowski P, et al. Heart failure: preventing disease and death worldwide. ESC Heart Fail 2014;1(1):4–25. https://doi.org/10.1002/ehf2.12005.
[2] Sakata Y, Shimokawa H. Epidemiology of heart failure in Asia. Circ J 2013. https://doi.org/10.1253/circj.cj-13-0971. CJ–13.
[3] Acharya UR, et al. Deep convolutional neural network for the automated diagnosis of congestive heart failure using ECG signals. Appl Intell 2018:1–12. https://doi.org/10.1007/s10489-018-1179-1.
[4] Yancy CW, et al. ACC/AHA/HFSA focused update of the 2013 ACCF/AHA guideline for the management of heart failure: a report of the American college of cardiology/American heart association task force on clinical practice guidelines and the heart failure society of amer. J Am Coll Cardiol 2017;70(6):776–803. https://doi.org/10.1161/CIR.0000000000000509. 2017.
[5] Jahmunah V, et al. Computer-aided diagnosis of congestive heart failure using ECG signals–A review. Phys Med 2019;62:95–104. https://doi.org/10.1016/j.ejmp.2019.05.004.
[6] Wang L, Zhou X. Detection of congestive heart failure based on LSTM-based deep network via short-term RR intervals. Sensors 2019;19(7):1502. https://doi.org/10.3390/s19071502.
[7] Gladding PA, et al. Open access integrated therapeutic and diagnostic platforms for personalized cardiovascular medicine. J Personalized Med 2013;3(3):203–37. https://doi.org/10.3390/jpm3030203.
[8] Nurmaini S, Gani A. And others, "cardiac arrhythmias classification using deep neural networks and principle component analysis algorithm. Int J Adv Soft Comput Its Appl 2018;10(2).
[9] Melillo P, De Luca N, Bracale M, Pecchia L. Classification tree for risk assessment in patients suffering from congestive heart failure via long-term heart rate variability. IEEE J Biomed Health Informatics 2013;17(3):727–33. https://doi.org/10.1109/JBHI.2013.2244902.
[10] De Jong MMJ, Randall DC. Heart rate variability analysis in the assessment of autonomic function in heart failure. J Cardiovasc Nurs 2005;20(3):186–95. https://doi.org/10.1097/00005082-200505000-00010.
[11] Orhan U. "Real-time CHF detection from ECG signals using a novel discretization method," *Comput*. Biol Med 2013;43(10):1556–62. https://doi.org/10.1016/j.compbiomed.2013.07.015.
[12] Maśetic Z, Subasi A. Detection of congestive heart failures using c4. 5 decision tree. Southeast Eur J Soft Comput 2013;2(2). https://doi.org/10.21533/scjournal.v2i2.32.
[13] Kamath C. A new approach to detect congestive heart failure using detrended fluctuation analysis of electrocardiogram signals. J Eng Sci Technol 2015;10(2):145–59. https://doi.org/10.1016/j.medengphy.2012.03.001.
[14] Sudarshan VK, et al. Automated diagnosis of congestive heart failure using dual tree complex wavelet transform and statistical features extracted from 2 s of ECG signals. Comput Biol Med 2017;83:48–58. https://doi.org/10.1016/j.compbiomed.2017.01.019.
[15] Isler Y, Narin A, Ozer M, Perc M. "Multi-stage classification of congestive heart failure based on short-term heart rate variability," *Chaos*. Solitons & Fractals 2019;118:145–51. https://doi.org/10.1016/j.chaos.2018.11.020.
[16] Nurmaini S, et al. Robust detection of atrial fibrillation from short-term electrocardiogram using convolutional neural networks. Future Generat Comput Syst 2020. https://doi.org/10.1016/j.future.2020.07.021.
[17] Nurmaini S, et al. An automated ECG beat classification system using deep neural networks with an unsupervised feature extraction technique. Appl Sci 2019;9(14):2921. https://doi.org/10.3390/app9142921.
[18] Makino T, et al. Recurrent neural network transducer for audio-visual speech recognition,. In: 2019 IEEE automatic speech recognition and understanding workshop. ASRU; 2019. p. 905–12. https://doi.org/10.1109/asru46091.2019.9004036.
[19] Liu Q, Fang L, Yu G, Wang D, Xiao C-L, Wang K. Detection of DNA base modifications by deep recurrent neural network on Oxford Nanopore sequencing data. Nat Commun 2019;10(1):1–11. https://doi.org/10.1038/s41467-019-10168-2.
[20] Esteva A, et al. A guide to deep learning in healthcare. Nat Med 2019;25(1):24–9. https://doi.org/10.1038/s41591-018-0316-z.
[21] Limam M, Precioso F. Atrial fibrillation detection and ECG classification based on convolutional recurrent neural network. In: 2017 computing in cardiology. CinC; 2017. p. 1–4. https://doi.org/10.22489/CinC.2017.171-325.
[22] Feng K, Pi X, Liu H, Sun K. Myocardial infarction classification based on convolutional neural network and recurrent neural network. Appl Sci 2019;9(9):1879. https://doi.org/10.3390/app9091879.
[23] Darmawahyuni A, Nurmaini S, Caesarendra W, Bhayyu V, Rachmatullah MN. And others, "deep learning with a recurrent network structure in the sequence modeling of imbalanced data for ECG-rhythm classifier. Algorithms 2019;12(6):118. https://doi.org/10.3390/a12060118.
[24] Darmawahyuni A, Nurmaini S, and others. Deep learning with long short-term memory for enhancement myocardial infarction classification. In: 2019 6th international conference on instrumentation, control, and automation. ICA; 2019. p. 19–23. https://doi.org/10.1109/ICA.2019.8916683.
[25] Chen W, Liu G, Su S, Jiang Q, Nguyen H. A CHF detection method based on deep learning with RR intervals. In: 2017 39th annual international conference of the IEEE engineering in medicine and biology society. EMBC; 2017. p. 3369–72. https://doi.org/10.1109/EMBC.2017.8037578.
[26] Singh S, Pandey SK, Pawar U, Janghel RR. Classification of ECG Arrhythmia using recurrent neural networks. Procedia Comput Sci 2018;132:1290–7. https://doi.org/10.1016/j.procs.2018.05.045.
[27] Sameni R, Shamsollahi MB, Jutten C, Clifford GD. A nonlinear Bayesian filtering framework for ECG denoising. IEEE Trans Biomed Eng 2007;54(12):2172–85. https://doi.org/10.1109/TBME.2007.897817.
[28] Tracey BH, Miller EL. Nonlocal means denoising of ECG signals. IEEE Trans Biomed Eng 2012;59(9):2383–6. https://doi.org/10.1109/TBME.2012.2208964.
[29] Wang J, Ye Y, Pan X, Gao X. Parallel-type fractional zero-phase filtering for ECG signal denoising. Biomed Signal Process Contr 2015;18:36–41. https://doi.org/10.1016/j.bspc.2014.10.012.
[30] Aqil M, Jbari A, Bourouhou A. ECG signal denoising by discrete wavelet transform. Int J Online Eng 2017;13(9). https://doi.org/10.3991/ijoe.v13i09.7159.
[31] Martis RJ, Acharya UR, Min LC. ECG beat classification using PCA, LDA, ICA and discrete wavelet transform. Biomed Signal Process Contr 2013;8(5):437–48. https://doi.org/10.1016/j.bspc.2013.01.005.

[32] Lin H-Y, Liang S-Y, Ho Y-L, Lin Y-H, Ma H-P. Discrete-wavelet-transform-based noise removal and feature extraction for ECG signals. Irbm 2014;35(6):351–61. https://doi.org/10.1016/j.irbm.2014.10.004.
[33] Wang Z, Zhu J, Yan T, Yang L. A new modified wavelet-based ECG denoising. Comput Assist Surg 2019:1–10. https://doi.org/10.1080/24699322.2018.1560088.
[34] Yildirim Ö. A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. Comput Biol Med 2018;96:189–202. https://doi.org/10.1016/j.compbiomed.2018.03.016.
[35] Lui HW, Chow KL. Multiclass classification of myocardial infarction with convolutional and recurrent neural networks for portable ECG devices. Informatics Med Unlocked 2018;13:26–33. https://doi.org/10.1016/j.imu.2018.08.002.
[36] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997;9(8):1735–80. https://doi.org/10.1162/neco.1997.9.8.1735.
[37] Davis J, Goadrich M. The relationship between Precision-Recall and ROC curves. In: Proceedings of the 23rd international conference on Machine learning; 2006. p. 233–40. https://doi.org/10.1145/1143844.1143874.

Congestive heart failure waveform.pdf

ORIGINALITY REPORT

100%	28%	85%	100%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Sriwijaya University Student Paper	100%

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	On		