IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 6, 2019, accepted August 21, 2019, date of publication September 6, 2019,
date of current version October 17, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2939947

A Deep Bidirectional GRU Network Model for Biometric Electrocardiogram Classification Based on Recurrent Neural Networks

HTET MYET LYNN[1], SUNG BUM PAN[2], AND PANKOO KIM[1], (Member, IEEE)
[1]Department of Computer Engineering, Chosun University, Gwangju 61452, South Korea
[2]Department of Electronics Engineering, Chosun University, Gwangju 61452, South Korea

Corresponding author: Pankoo Kim (pkkim@chosun.ac.kr)

This work was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant 2017R1A6A1A03015496, and in part by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIT) under Grant NRF-2019R1H1A2080262.

ABSTRACT In this paper, we propose a deep Recurrent Neural Networks (RNNs) based on Gated Recurrent Unit (GRU) in a bidirectional manner (BGRU) for human identification from electrocardiogram (ECG) based biometrics, a classification task which aims to identify a subject from a given time-series sequential data. Despite having a major issue in traditional RNN networks which they learn representations from previous time sequences, bidirectional is designed to learn the representations from future time steps which enables for better understanding of context, and eliminate ambiguity. Moreover, GRU cell in RNNs deploys an update gate and a reset gate in a hidden state layer which is computationally efficient than a usual LSTM network due to the reduction of gates. The experimental results suggest that our proposed BGRU model, the combination of RNN with GRU cell unit in bidirectional manner, achieved a high classification accuracy of 98.55%. Various neural network architectures with different parameters are also evaluated for different approaches, including one-dimensional Convolutional Neural Network (1D-CNN), and traditional RNNs with LSTM and GRU for non-fiducial approach. The proposed models were evaluated with two publicly available datasets: ECG-ID Database (ECGID) and MIT-BIH Arrhythmia Database (MITDB). This paper is expected to demonstrate the feasibility and effectiveness of applying various deep learning approaches to biometric identification and also evaluate the effect of network performance on classification accuracy according to the changes in percentage of training dataset.

INDEX TERMS 1D-convolutional neural networks, bidriectional recurrent neural networks, biometrics classification, ECG signals, gated recurrent unit, user identification, signal processing.

I. INTRODUCTION

Due to the unique characteristics of electrocardiogram (ECG) signal, it has drawn increasing attention from biometric researchers in broad fields of information security in recent years. Biometrics recognition study can be considered as identifying the individuals based on their unique physiological and behavioral attributes of a particular person which are encoded in a sequence of successive samples in time, by employing statistical methods. Biometrics human identification or authentication can be achieved using several unique human traits such as iris or retina, face, fingerprint, voice, written signature, etc. However, each of the discriminants has the limitations according to the particular hardware to operate, the sustainability of the security against spoofing attacks. However, leveraging the distinct and permanence to individuals, biometrics characteristics are reliable than traditional methods, some of the biometrics features mentioned above can easily be forged and replicated, and also unpleasant and invasive to acquire. More recently, the electrocardiogram which is the electrical activity of the heart on the body surface has gained momentum, and compared with other biometric characteristics has proven to be the most promising of them according to its hidden nature which makes it hard to replicate and inject into the system for spoofing purposes. Since ECG signal is an electrical current that is generated by the heart as it beats spread out not only within the heart, but also throughout the body, and the sinus node in the heart modulated by both

The associate editor coordinating the review of this manuscript and approving it for publication was Zhenhua Guo.

sympathetic and parasympathetic nerves, it is unique and permanent as based on the size and shape of oneś heart and the orientation of valves. Given the fact that ECG signals are unique for individuals, they have potential in biometric human identification. Recently, deep learning algorithms which are often used in the machine learning and artificial intelligence field, offer a structure in terms of both feature extraction and classification process, which is known as end-to-end learning, instead of using hand-crafted features [2].

In this study, several Recurrent Neural Networks (RNN) based models with different cell units, as well as, One-dimensional Convolution Neural Network (1D-CNN) model is also proposed for classifying ECG signals. For that purpose, above all the deep network models are designed, and the performance of these models classification was investigated.

The main contributions of this work are:

1) We demonstrate the effectiveness of performing unidirectional and bidirectional RNNs based on both Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) cells, as well as 1D-CNN for biometrics identification.
2) We designed an RNN-based bidirectional GRU (BGRU) model. To the best of our knowledge, this is the first work to do so.

The rest of this article is organized as follows. A brief review of previous works is reviewed in Section II. A background overview of RNNs with different types of cell units is discussed in Section III. The data argumentation and preprocessing, and proposed models are explained in Section IV and V, respectively. Experimental results and comparisons are presented in Section VI. Finally, concluding remarks are given in Section VII.

II. RELATED WORKS

Many researchers have been informed for ECG biometric identification, usually based on two approaches, i.e., fiducial and non-fiducial methods. The fiducial approach requires feature extraction process, points of interest within the heartbeat wave (P-QRS-T complex), which are then used to extract latency and amplitude features [1], [3]. Such approaches generally importantly rely on robust heartbeat segmentation and fiducial point detection techniques, and also much manual feature engineering efforts are required. The first deflection of a heartbeat is the P wave caused by right and left atrial depolarization of the heart as it beats. The second wave is QRS complex, which contains a series of three deflections associated with left and right ventricular depolarization. Finally, the T wave is the current of rapid phase three ventricular repolarization [3]. It is mandatory to be able to precisely estimate onset, offset and peak locations of the P, Q, R, S, and T waves of a signal. After detection of P-QRS-T complex, commonly, the amplitude and time-interval between corresponding points are measured, such process is considered as preprocessing step in many applications, namely fiducial features extraction, see Fig. 1.

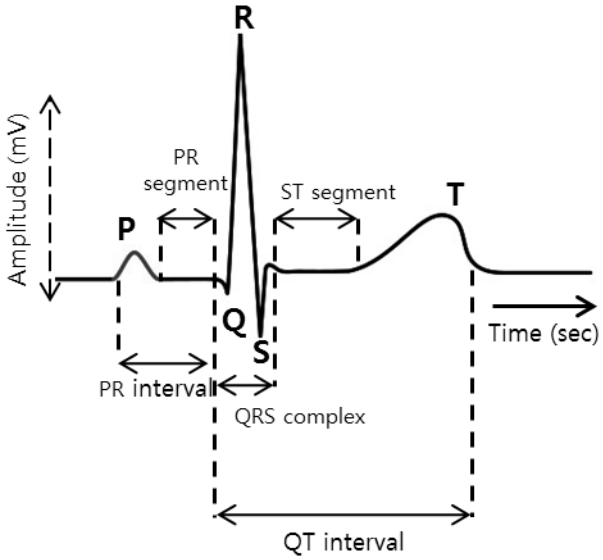

FIGURE 1. A single cardiac cycle of the ECG signal.

For classification task, deep learning algorithms have been successfully applied to many areas such as human handwriting recognition [9], face recognition [10], image classification [11], [12], and object recognition [13]. Moreover, there are many studies have been done in the area of bioinformatics signals by employing deep learning methods [14]–[19]. Zhang *et al.* [20] used a multiresolution parallel network based on CNN for biometric human identification in smart health applications, and they extended the raw input data into multiple versions of waveforms as to improve the context representations of signal for training but auto-correlation of raw signal and transformation of wavelet are required. Zhang *et al.* [21] proposed RNN based models with various types of cell unit in hidden layers, and reported that both LSTM and GRU cells were not significantly different in classification accuracy while both methods could achieve high accuracy. Rahhal *et al.* [15] proposed a method by applying stacked denoising autoencoders with sparsity constraint, and softmax regression layer is applied on the top of the hidden representation layer to form a deep neural network. Zihlmann *et al.* [22] presented two deep neural network models, CNN and a hybrid approach of combining CNN with RNN network with LSTM cell unit. Warrick and Homsi [19] also presented an approach in the same fashion as previous work [22] which automatically detect and identify cardiac arrhythmias in ECG signals by deploying CNN and LSTM techniques.

III. BACKGROUND

A. RECURRENT NEURAL NETWORKS

Recently, RNN model has been a highly preferred method [27], especially for sequential data and typical RNN is illustrated as show in Fig. 2(a). Every node at a time step consists of an input from the previous node and it proceeds

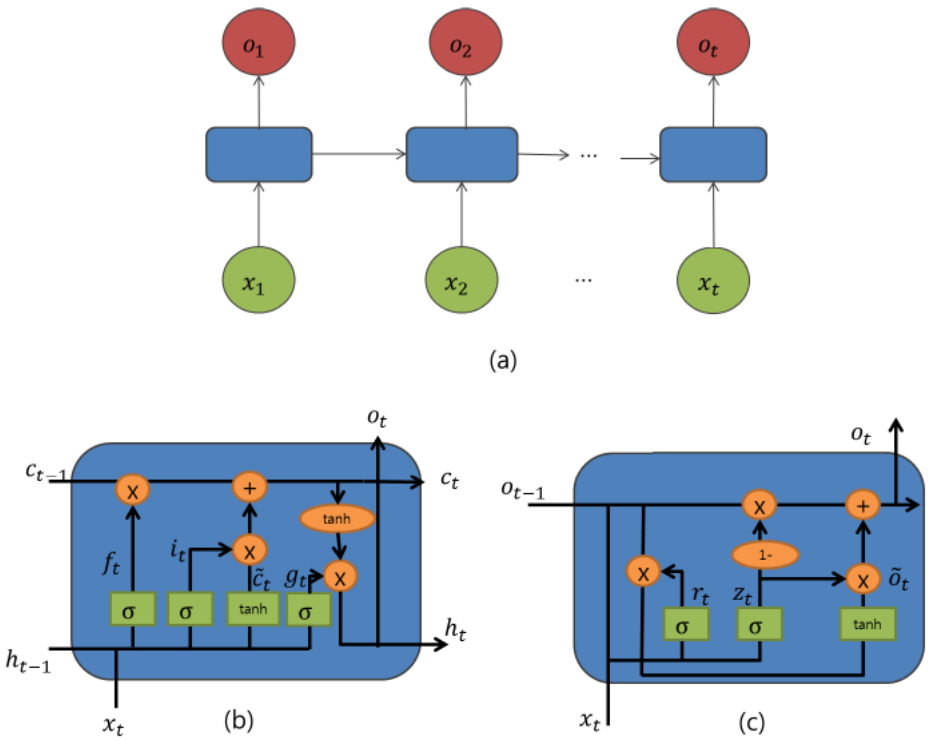

FIGURE 2. Recurrent neural networks and different cell units of its hidden layer (a) Conventional RNN model, (b) LSTM cell unit, (c) GRU cell unit.

using a feedback loop. Each node produces a current hidden state and output by using current input and previous hidden state as follows:

$$h_t = f(W_h h_{t-1} + V_h x_t + b_h). \quad (1)$$
$$o_t = f(W_o h_t + b_o). \quad (2)$$

where h_t denotes the hidden block of each time step (t), W and V are the weights for the hidden layers in recurrent connection, while b denotes the bias for hidden and output states as f represents an activation function applied on each node throughout the network.

B. LONG SHORT-TERM MEMORY (LSTM)

The common drawback of conventional RNN model is that as the time steps increase, the network fails to derive context from time steps of previous states much far behind, such phenomenon is known as long term dependency. Due to the deep layers of a network and recurrent behavior of typical RNN, exploding and vanishing gradients problems are also encountered quite often. Moreover, to address this problem, LSTM models are introduced by deploying memory cells with several gates in a hidden layer [28]. Fig. 2(b) illustrates the block of a hidden layer with LSTM cell unit, and three functions of gate controllers are demonstrated as follows:

- Forget gate f_t decides which part of long-term state c_t should be omitted.
- Input gate i_t controls which part of $\tilde{c}_t$ should be added to long-term state c_t.
- Output gate g_t determines which part of c_t should be read and outputs to h_t and o_t.

The following equations show the long-term and short-term states of the cell and the output of each layer in time step in Figure 2(b).

$$f_t = \sigma(W_{xf}^T \cdot x_t + W_{hf}^T \cdot h_{t-1} + b_f). \quad (3)$$
$$i_t = \sigma(W_x^T i \cdot x_t + W_h^T i \cdot h_{t-1} + b_i). \quad (4)$$
$$o_t = \sigma(W_x^T o \cdot x_t + W_h^T o \cdot h_{t-1} + b_i). \quad (5)$$
$$g_t = tanh(W_x^T g \cdot x_t + W_h^T g \cdot h_{t-1} + b_i). \quad (6)$$
$$c_t = f_t \otimes c_{t-1} + i_t \otimes \tilde{c}_t. \quad (7)$$
$$o_t, h_t = g_t \otimes tanh(c_t). \quad (8)$$

where $W_x f$, $W_x i$, $W_x o$, $W_x g$ denote the weight matrices for the corresponding connected input vector, $W_h f$, $W_h i$, $W_h o$, $W_h g$ represent the weight matrices of the short-term state of the previous time step, and b_f, b_i, b_o, b_g are bias.

C. GATED RECURRENT UNIT (GRU)

In GRU cell unit, the two vectors in LSTM cell are combined into one vector o_t [29]. One gate controller controls both forget and input gates. When z_t outputs 1, the forget gate is opened and the input gate is closed, whereas z_t is 0, the forget gate is closed and the input gate is opened. In this fashion, the input of the time step is deleted every time the previous

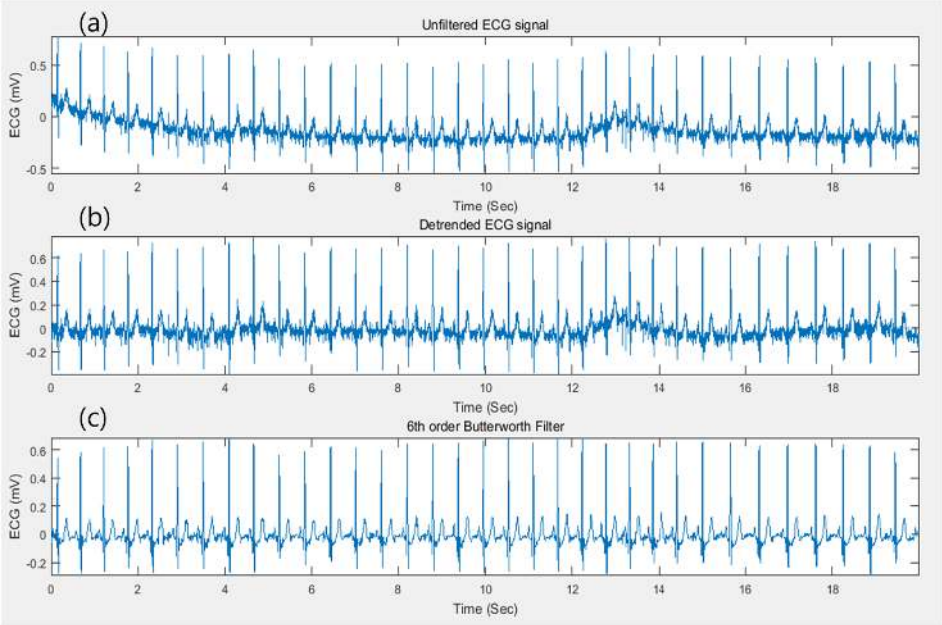

FIGURE 3. (a) A raw ECG signal with non-linear trend. (b) Detrended ECG signal. (c) Filtered ECG signal by applying 6*th* order Butterworth filter.

$(t-1)$ memory is stored. In the absence of an output gate, it can be said that the GRU is a different implementation of the delivery and combination of the information that LSTM wants to implement. Intuitively, the reset gate determines how to combine the new input with the previous memory, and the update gate decides how much of the previous memory information is retained to calculate the new state, see in Fig. 2(c). The differences from the notable LSTM, except for the differences already described, are as follows:

$$r_t = \sigma(W_{xr}^T \cdot x_t + W_{or}^T \cdot o_{t-1} + b_r). \quad (9)$$

$$z_t = \sigma(W_{xz}^T \cdot x_t + W_o^T z \cdot o_{t-1} + b_z). \quad (10)$$

$$\tilde{o}_t = tanh(W_{x\tilde{o}}^T \cdot x_t + W_{o\tilde{o}}^T \cdot (r_t \otimes o_{t-1}) + b_{\tilde{o}}). \quad (11)$$

$$o_t = z_t \otimes o_{t-1} + (1 - z_t) \otimes \tilde{o}_t. \quad (12)$$

where $W_x r$, $W_x z$, $W_{x\tilde{o}}$ denote the weight matrices for the corresponding connected input vector, $W_o r$, $W_o z$, W_o represent the weight matrices of the previous time step, and b_r, b_z, b are bias.

IV. DATA ARGUMENTATION AND PREPROCESSING

For our experiment, ECG-ID (ECGID) and MIT-BIH Arrhythmia Database (MITDB) from PhysioNet [25], [26] has been performed separately for all candidate models. ECG-ID database contains 310 ECG recordings, collected from 90 persons. Each recording contains ECG lead I signal version for 20 seconds (10, 000 samples) digitized at 500 Hz with 12-bit resolution over a nominal $+/-10\ mV$ range. The signals were recorded from volunteers of 44 men and 46 women aged from 13 to 75 years. The MIT-BIH Arrhythmia Database consists of ECG recordings obtained from 47 subjects collected by the BIH Arrhythmia laboratory between 1975 and 1979. The records were digitized at 360 samples per second with $11-bit$ resolution over a $10mV$ range. The signal processing step can also be considered as data pre-processing, including following three operations, i.e., detrending, noisy removal (filtering), and R-peak detection which is the act of locating the index numbers of corresponding R-peak points throughout the signal. The raw ECG signal is detrended in order to enable the approximation better when conducting with finite-length segments for signal analysis. To eliminate the nonlinear trend in a signal, fit a low-order polynomial to the signal and subtract it. In this case, the polynomial is of order 6 [24]. Accordingly, a 6-order Butterworth bandpass filter with $5Hz$ and $15Hz$ range is applied to remove the baseline wander (low frequency noise due to the perspiration that affects electrode impedance, respiration, body movements, for example finger movements on the electrode) while recording the signal. The result of detrending and filtering a raw signal can be found in Fig. 3. Then the filtered signals are normalized in the range of 0 and 1, and by subtracting from mean value to balance the contribution for training phase using (13) and (14), where x and $\tilde{x}$ are the original ECG signal and new processed ECG signal, respectively.

$$\tilde{x} = \frac{(x - min(x))}{(max(x) - min(x))}. \quad (13)$$

$$\tilde{x} = \tilde{x} - mean(\tilde{x}). \quad (14)$$

The traces of heartbeat consist of three complexes: P, QRS, and T. There are respectively defined by their corresponding fiducial points which are the peak of each complex. Time domain features such as amplitudes and intervals of these points are commonly used as the distinct unique features

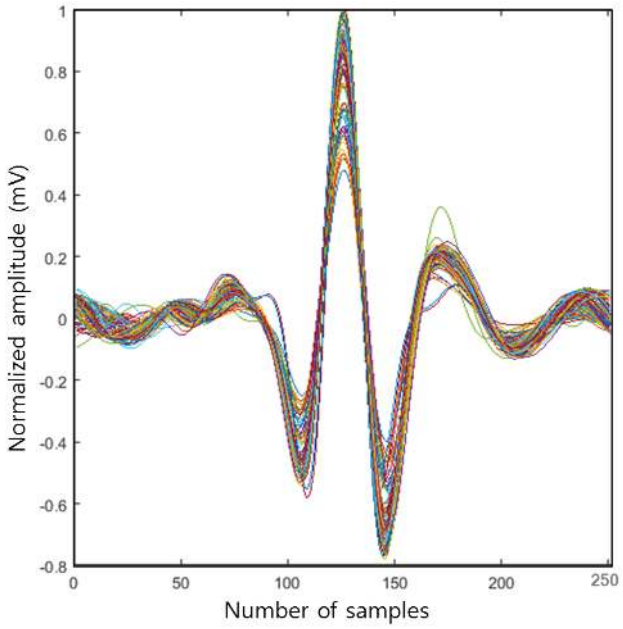

FIGURE 4. Multiple samples utilized for the extraction of a single heartbeat according to the respective R peak points on ECG signals of MITDB dataset.

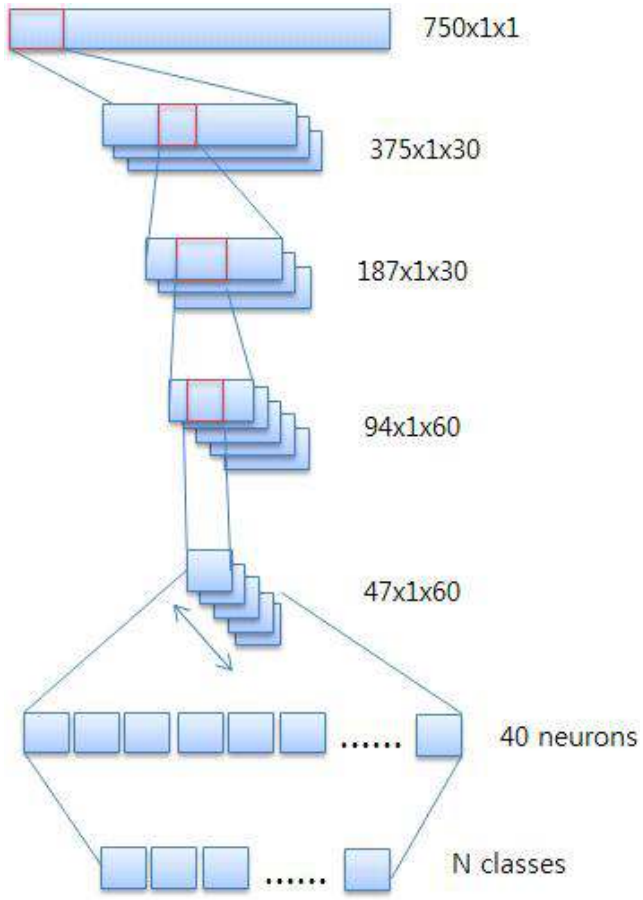

FIGURE 5. Proposed 1D-CNN network architecture.

to individuals. However, due to the non-fiducial approach with deep learning technique such fiducial extraction is omitted. The R-peak detection, thereafter, is performed using the Pan-Tompkins algorithms [25] for indexing the corresponding peaks throughout the signal since R peak is the most prominent peak, and it can be used as an index of given heartbeat waveform to segment the raw signal into individual heartbeat waveforms. Once indices of R peak of the signal are detected, a certain number of samples before and after a given R peak are extracted to form a vector which illustrates the heartbeat waveform. By concatenating 125 samples before and after of R peak indices of MITDB dataset, and 150 samples before and after of R peak points for ECG-ID dataset since both datasets have different sampling rates, form a vector which decodes the heartbeat waveform given in Fig. 4.

For each recording, from MITDB dataset, approximately 45 heartbeat waveforms can be withdrawn with 251 samples, whereas 51 heartbeats are extracted with 301 samples from ECG-ID dataset. For CNN model, blind segmentation with two-second window was used to enrich the training the dataset and each window includes 750 samples where at least one or more heartbeats can be extracted randomly from a signal.

V. MODELS OVERVIEW

In this section, several deep neural networks were designed to perform classification task on biometircs ECG signals. The first of these models is based on 1-D CNN which enables to learn the hierarchical distinct features in order to compose a new feature set of representation of a high level abstraction, which can then be fed into a classifier such as fully connected layer for further category identification process. The second and third models are based on RNN with modified cell units, LSTM and GRU, in their hidden states while the training procedure is conducted in bidirectional manner. In a conventional RNN, the hidden state at a given time step is computed as a linear combination of the previous hidden state and the current input. GRU and LSTM networks have similar block diagram, however the update of the hidden state is more complex in both approaches. Fig. 5 and 6 illustrate the proposed 1D-CNN model, and various models of RNN networks with different cell unit used in this study, respectively.

A. PROPOSED 1-D CNN MODEL

CNNs were initially developed in the 1980s, and it is designed to be trained robustly by means of the stochastic gradient descent algorithm for its layers [10]. The CNNs have been widely used for feature extraction and signal classification problems. In this work, a deep 1-D CNN is proposed to perform the biometrics based classification for non-fiducial approach. CNNs address the classification tasks without exploiting the temporal correlations between sequential samples. The overall network structure of proposed CNN model is implemented according to Fig. 4, and its parameters of the network used in this study are listed in Table 1.

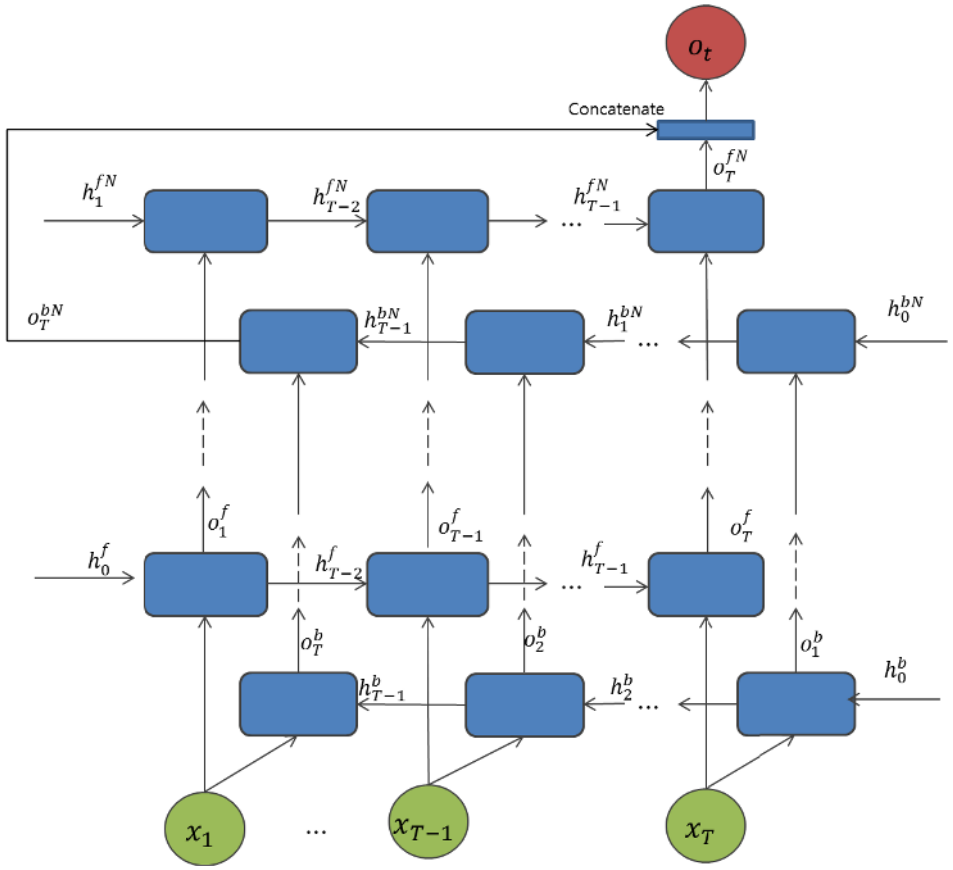

FIGURE 6. Proposed bidirectional RNN based models where the hidden blocks can be used as either LSTM or GRU cells.

TABLE 1. Values of proposed 1-D CNN model parameters.

Layer#	1	2	3	4
Kernel size	5	2	5	2
Stride	2	2	2	2
Padding	2	0	2	0
Input size	750	375	187	94
Output size	375	187	94	47

The proposed CNN model consists of four hidden layers for feature learning followed by a fully-connected layer with 40 neurons before decision making classification layer with sigmoid function is deployed. The intution of the proposed model is that this problem by nature should allow a function which differentiate the patterns of all subjects based on their respective input signal for identification. Usually the grown truth distribution is expressed in terms of a one-hot distribution. The input is discrete sequential samples $(x_1, \ldots, x_T)$, where each data point x_t is a vector of individual samples aquired by the sensors at time t. These samples are blindly segmented into two-second windows which equivalent to 750 samples (750×1 dimensions) for each window randomly, in order to capture at least one or more heartbeats from the original signal. From the first layer to the final layer, the values of the parameters such as filter size, stride and padding values are applied according to Table 1, and perform convolutional operations with non-linear activation functions are applied between each layer. In first and second layer, 30 numbers of filters Softmax function is applied on the last layer to express the distribution of the corresponding class in terms of a one-hot distribution in the range of 0 to 1. The cross-entropy loss function for a signle instance with independent targets can be formulated as:

$$E = -\sum_{i=1}(\hat{y}_i log(y_i) + (1 - \hat{y}_i) log(1 - y_i)). \quad (15)$$

where $\tilde{y}_i$ is the ground truth target vector, y_i is the output vector of our model for i class. To obtain the output in distributional manner across all the subject, outputs $y_{1,\ldots,i}$ are computed by applying non-linear sigmoid function to the weight sums of activation function of previous layer.

B. PROPOSED RNN ARCHITECTURES

For RNN based models, the input training dataset can be set by $S = (X_n, O_n), n = 1, ..., N$, where sample $X_n = x_k^n, k = 1, .., m$ denotes the m numbers of samples in a signal which is 251(an R peak index point with 125 samples from left/right) in MITDB, while 301 (an R peak index point with 150 samples from left/right) in ECG-ID dataset. $O_n = o_i^n$,

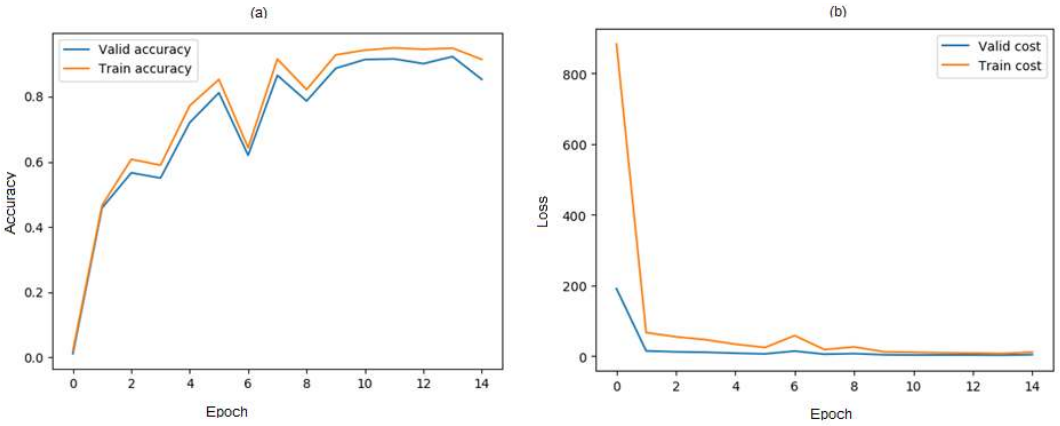

FIGURE 7. Accuracy and loss of the 1D-CNN model per epoch over MITDB dataset: (a) Training and validation accuracies ; (b) cross-entropy loss for training and validation per epoch.

where i denotes the number of subjects or classes. O_i^n denotes the corresponding ground truth binary score for each subject for nth input. The ground truth value for correct subject is determined as 1, and 0 for the rest of the subjects respectively. Formally, given a sequence input, a classifier is trained to learn the probabilities of N classes.

1) BIDIRECTIONAL RNN WITH LSTM CELL (BLSTM) AND GRU CELL (BGRU)

The second proposed model is based on a bidirectional RNN with LSTM cell unit in the hidden state layer, and is briefly called BLSTM, as shown in Fig.6 associated with the cell unit from Fig. 2(b). The segmented signal inputs $(x_1, \ldots, x_T)$ from Section IV, are fed into the network for each time step t $(t = 1, ..., T)$ for each LSTM cell. Each bi-directional cell contains a parallel of LSTM tracks, forward and a backward sequence for utilizing context from the past and future to predict its corresponding class. At the final time step, two parallel tracks of backward and forward LSTM tracks are concatenated into a single vector. In the first hidden layer, the forward cell states h_0^f, the backward cell state h_0^b are initialized to zeroes for every layer N. The input x_t at time t, and previous cell states h_{t-1} to produce the output of the corresponding layer o_t^n, at time t and at $n - th$ layer either the backward or forward stream given its parameter θ^n can be defined as:

$$o_t^n, h_t^n = LSTM^n(h_{t-1}^n, x_t; \theta^n). \quad (16)$$

$$o_t^n, h_t^n = GRU^n(h_{t-1}^n, x_t; \theta^n). \quad (17)$$

where θ^n denotes the parameters (b, U, W) of the respective cell unit for layer n.

For third proposed model, the only difference with BLSTM is the cell unit at the hidden layers, where gated recurrent units (GRUs) are substituted as shown in Figure 2(c). Moreover, to address one of the most important challenges, overfitting, in deep neural networks, the dropout layer is also applied in each cell. However, as the outputs at the last layer resulting from both forward and backward streams, the late-fusion for bidirectional networks is concatenated into a single vector and, it is followed by a softmax activation function to obtain N-dimensional output. The forward track traces the input segment from left to right, whereas the backward track traces back the input from right to left in both BLSTM and BGRU, and can be defined as follows:

$$o_t^f, h_t^f, c_t^f = LSTM^f(c_{t-1}^f, h_{t-1}^f, x_t; W^f). \quad (18)$$

$$o_t^b, h_t^b, c_t^b = LSTM^b(c_{t-1}^b, h_{t-1}^b, x_t; W^b). \quad (19)$$

$$o_t^f, h_t^f, c_t^f = GRU^f(c_{t-1}^f, h_{t-1}^f, x_t; W^f). \quad (20)$$

$$o_t^b, h_t^b, c_t^b = GRU^b(c_{t-1}^b, h_{t-1}^b, x_t; W^b). \quad (21)$$

VI. EXPERIMENTAL RESULTS

A. TRAINING

In order to speed up the training procedure, typically a bottleneck when running a deep network with multiple layers, our proposed network schemes are implemented using Tensorflow deep learning library written in Python, which can be executed on Graphics Processing Unit (GPU). GPU generally brings at least 5 to 10 times speedups compared with Central Processing Unit (CPU) and also can significantly accelerate the training procedure, and GeForce GTX 1080 GPU is used for our experiments.

In the training phase for 1D-CNN approach, the network self-learns hierarchical features by convolutional and pooling operations according to the parameters provided in Table 1. An example of the training process is given in Fig. 7, where the block on the left side shows the gradually decreasing training loss and the right part corresponds to the increasing training and validation accuracy per epoch. A stochastic gradient descent (SGD) training strategy is selected to further accelerate the training process, which allows for passing a subset of training data to the neural network each time. The batch size is chosen as 150 for all proposed methods including BGRU and BLSTM to trade off two considerations, i.e., a large size results in a short convergence time by reducing the variance of stochastic gradient updates, and a small size brings more power for SGD to jump out of shallow minima in the error function. The epoch size is set as 50 to balance under-fitting

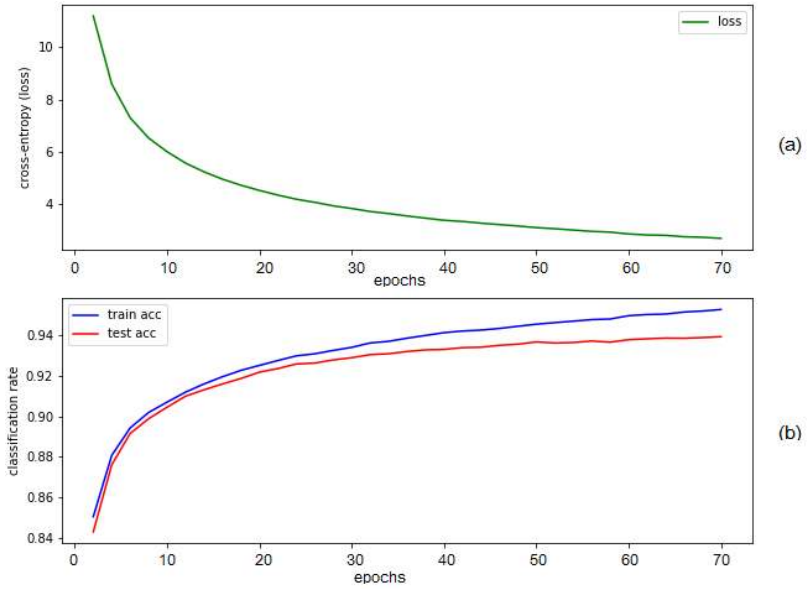

FIGURE 8. Accuracy and loss of the proposed BGRU model per epoch over MITDB dataset: (a) Training and test accuracies; (b) cross-entropy loss for training per epoch.

and over-fitting considerations, although the network can learn the hidden patterns of the signal and converges around 14 epochs.

For training RNN based models, the batch size is determined as 150 which yields better performance than other schemes, while the optimization is deployed by Adam optimizer as the learning rate is set to 0.001, and loss function was selected as the categorical cross-entropy as mentioned in (22), where $\tilde{o}_l$ is the ground truth target vector, and o_l is the output vector of the model for l class. The optimal window length of segmented signal is selected according to the previous works and after various lengths of trails, in order to capture a single heartbeat regarding to the corresponding dataset. Moreover, all the parameters of the networks are also conducted with various settings by trial-and-error approach, and chose the optimal setting for each network which yields the most with better performance results. The weights in the models were initialized randomly at the start of the training process, and progressively updated throughout the process. In addition, a dropout of 0.2 at the first layer outputs of the networks, and at the last layer inputs were deployed to avoid an overfitting problem typically encounter in deep neural networks. An example of the training and test process of the proposed BGRU model on MITDB dataset is shown in Fig. 8, where the top block presents the training cost when the percentage of subjects used for training is 50%

$$E = -\sum_{l=1}(\tilde{o}_l log(o_l) + (1 - \tilde{o}_l) log(1 - o_l)). \quad (22)$$

B. EVALUATION METRICS

For classification task, the models were evaluated by the classification accuracy which is defined by the confusion matrix which is one of the most intuitive metrics used for evaluating the performance and accuracy of the model in machine learning, commonly used for the classification problems where the output could be two or more types of classes as show in Table 2.

TABLE 2. Confusion matrix of evaluating classification accuracy.

Actual / Predicted	Positives (1)	Negatives (0)
Positives (1)	TP	FP
Negatives (0)	FP	TN

The terms associated with confusion matrix can be defined as: True positives (TP), when the actual class of the data point is 1 and the predicted outcome is also 1. True negatives (TN) are the cases when the actual class of the given data point is 0 and the predicted result is also 0. False positives (FP) are the cases when the actual class of the data point is 0 and the predicted outcome is 1, which can be assumed that the model predicts incorrectly as the actual class is positive. False negatives are the cases when the actual class should be 1 and the predicted outcome is 0, where the model predicts incorrectly as negative.

Since the accuracy in classification problems is the number of correct predictions made by the model over all kinds predictions made. The correct predictions True Positive (TP) and True Negative (TN) are divided by the kind of all predictions made by the model, and it can be formulated as below:

$$Classificationaccuracy = \frac{TP + TN}{TP + FP + FN + TN}. \quad (23)$$

C. RESULTS

For the experiment, the total of 5066 recordings from both datasets was randomly separated into training and test sets.

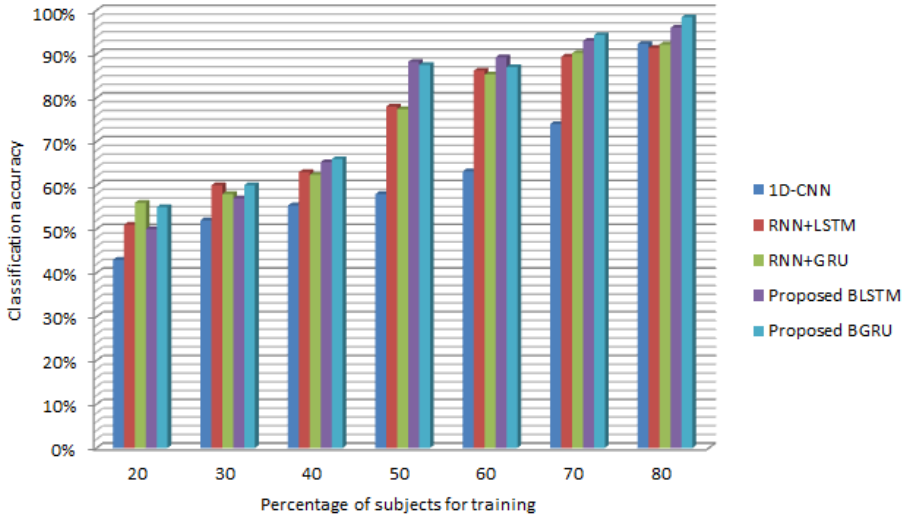

FIGURE 9. Performance comparison of proposed models with standard RNN models on ECGDB dataset.

In order to demonstrate the effectiveness and performance of percentage of training examples on the models, different training ratios for datasets were conducted can be seen in Fig. 9. As the results were experimented, it shows that bidirectional network with LSTM and GRU models are more effective than conventional RNN models, and BGRU model obtains relatively higher performance than BLSTM model. For both datasets, for a given subject, the capability of classification accuracy relied on different number of training successive heartbeats from a particular session are also conducted for each subject.

TABLE 3. Performance of classification accuracy for selected input sequence length.

Type of model	Input sequence length (number of heartbeats)	Overall accracy
RNN + LSTM	3 9	0.935 0.961
RNN + GRU	3 9	0.955 0.963
Proposed BLSTM	3 9	0.951 0.964
Proposed BGRU	3 9	0.976 0.985

The reported overall classification accuracy ranged from 93.5% to 97.6% for 3-consecutive heartbeat training inputs whereas 9-consecutive heartbeat inputs resulting from 96.1% to 98.5% for RNN based models. The proposed BGRU models outperform conventional RNN models and other proposed models, see in Table 3. In addition, the effect of varying percentage of training data size for classification accuracy for each subject was also evaluated for both ECGID and MITDB dataset. The figures showed that the rate of accuracy performs better and well classified as the training size increases, and furthermore, the accuracy rate achieves at its peak when the percentage of subjects used for training is 80%. According to Fig. 9 and Fig. 10, although BLSTM surpasses the other models in training with 70% of dataset, BGRU yields the best results in overall experiments.

TABLE 4. Performance comparison with state-of-the-art models.

Methods	Datasets	Overall accracy (%)
Proposed BGRU	ECGID MITDB	98.60 98.40
Q. Zhang et al. [20]	ECGID MITDB	92.53 91.31
X. Zhang [21]	ECGID MITDB	98.34 97.82
Martis et al. [22]	ECGID MITDB	95.54 95.33
M. Zihlmann [23]	ECGID MITDB	90.70 91.15
Ö. Yildirim [24]	ECGID MITDB	99.32 98.74

The best performances of all proposed deep neural networks based models were evaluated as their convergences of classification accuracy for each epoch are illustrated in Fig. 11. The models achieve their corresponding accuracy convergence approximately in 60th epoch overall models. Based on the results, proposed BGRU network produced the best classification rate while time consumption is less than BLSTM network in the learning phase.

Given the fact that 80% of training inputs yields the best outcome, further investigations were conducted by comparing the proposed BGRU model with other state-of-the-art methods using 80% of training data for each subject. According to the Table 4, the proposed BGRU network outperforms other state-of-the-art methods by obtaining average 98% for both datasets, in fact, it fails to surpass the model proposed by Yildirim [23] which scored approximately 99%, as it requires more time compared to ours since it is based on LSTM approach. However, the only disadvantage of the network is the time cost of the training phase, which is a common problem of most deep neural networks, our

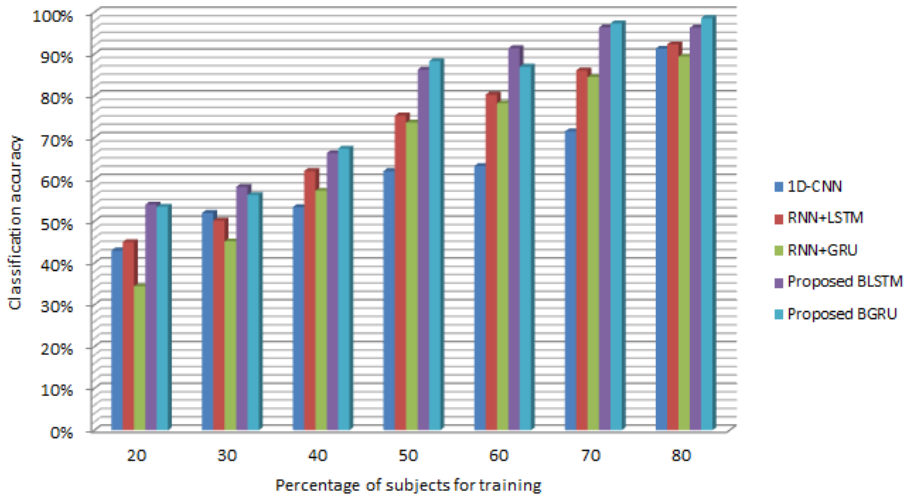

FIGURE 10. Performance comparison of proposed models with standard RNN models on MITDB dataset.

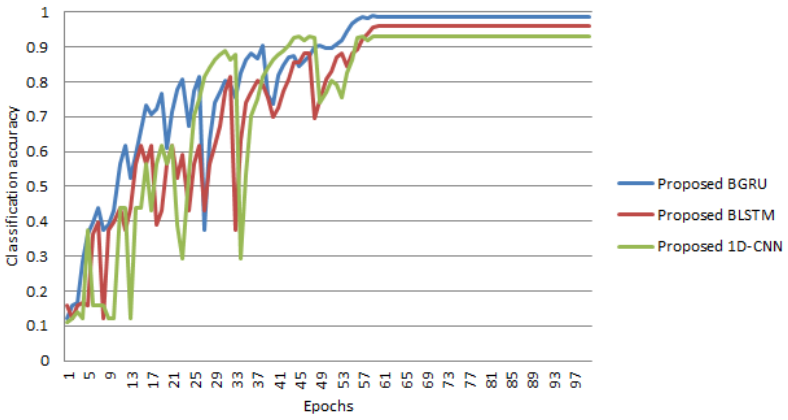

FIGURE 11. Performance of classification accuracy of proposed models for 80% of training size on ECGDB dataset.

proposed BGRU network is practically affordable and can be discarded since a fully trained network is used in real applications.

VII. CONCLUSION

In this study, the task of ECG based biometrics human identification was carried out based on deep RNN networks in bidirectional training manner with LSTM and GRU cell unit, which recently have performed a significant performance in the field of machine learning. Although being held a major issue in traditional RNN networks which they learn representations from previous time sequences, bidirectional networks are designed to learn the representations from future time steps which allows for better understanding of context, and eliminate ambiguity. In addition, GRU cell in RNNs deploys an update gate and a reset gate in a hidden state layer which is computationally efficient than a usual LSTM network due to the reduction of gates but still can perform as much as LSTM network does.

Experimental reports showed that the proposed models outperform most of the state-of-the-art methods by adapting the bidirectional learning trait in training process significantly improve the performance and able to extract more distinct features by feeding along the deep layers in order to capture the activity sequences in each time step of input signals. However, the disadvantage of the proposed method is the time cost of the training phase, which is a general problem of most deep networks, and the variation of the length of a segmented window for respective datasets should perform to investigate the optimal length of a window which capture a signle heartbeat of a signal for corresponding dataset. On the other hand, limitations of our proposal can be considered as data dependent acquistion where only ECG-I type data are suitable for pre-processing phase before training. Thus, our future work includes investigation on large-scale and diverse datasets, as well as domain independent training process with various versions of biometrics electrocardiogram waveforms for human identification. We will also further consider more

data representation methods and deep network architectures, to be able to learn the correlation between the data and distinct feature learning process.

REFERENCES

[1] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 3, pp. 230–236, Mar. 1985.

[2] Y. Bengio, "Learning deep architectures for AI," *Found. Trends Mach. Learn.*, vol. 2, no. 1, pp. 1–127, 2009.

[3] P. Raman and S. M. Ghosh, "Classification of heart diseases based on ECG analysis using FCM and SVM methods," *Int. J. Eng. Sci.*, vol. 6, pp. 6739–6744, Jun. 2016.

[4] S.-N. Yu and K.-T. Chou, "Integration of independent component analysis and neural networks for ECG beat classification," *Expert Syst. Appl.*, vol. 34, no. 4, pp. 2841–2846, May 2008.

[5] O. T. Inan, L. Giovangrandi, and G. T. A. Kovacs, "Robust neural-network-based classification of premature ventricular contractions using wavelet transform and timing interval features," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 12, pp. 2507–2515, Dec. 2006.

[6] S. Sahoo, B. Kanungo, S. Behera, and S. Sabut, "Multiresolution wavelet transform based feature extraction and ECG classification to detect cardiac abnormalities," *Measurement*, vol. 108, pp. 55–66, Oct. 2017.

[7] M. Thomas, M. K. Das, and S. Ari, "Automatic ECG arrhythmia classification using dual tree complex wavelet based features," *Int. J. Electron. Commun.*, vol. 69, no. 4, pp. 715–721, Apr. 2015.

[8] R. J. Martis, U. R. Acharya, and L. C. Min, "ECG beat classification using PCA, LDA, ICA and Discrete Wavelet Transform," *Biomed. Signal Process. Control*, vol. 8, pp. 437–448, Sep. 2013.

[9] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[10] M. Coskun, A. Uçar, Ö. Yildirim, and Y. Demir, "Face recognition based on convolutional neural network," in *Proc. Int. Conf. Modern Elect. Energy Syst. (MEES)*, Nov. 2017, pp. 376–379.

[11] Ö. Yildirim, A. Uçar, and U. B. Baloglu, "Recognition of real-world texture images under challenging conditions with deep learning," *J. Intell. Syst. Appl.*, vol. 1, no. 1, pp. 122–126, 2018.

[12] A. I. Krizhevsky, Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[13] A. Uçar, Y. Demir, and C. Güzelis, "Object recognition and detection with deep learning for autonomous driving applications," *Simulation*, vol. 9, pp. 759–769, Jun. 2017.

[14] S. Min, B. Lee, and S. Yoon, "Deep learning in bioinformatics," *Brief Bioinform.*, vol. 18, no. 5, pp. 851–869, 2016.

[15] M. M. Al Rahhal, Y. Bazi, H. AlHichri, N. Alajlan, F. Melgani, and R. R. Yager, "Deep learning approach for active classification of electrocardiogram signals," *Inf. Sci.*, vol. 345, pp. 340–354, Jun. 2016.

[16] J. H. Tan, Y. Hagiwara, W. Pang, I. Lim, S. L. Oh, M. Adam, R. S. Tan, M. Chen, and U. R. Acharya, "Application of stacked convolutional and long short-term memory network for accurate identification of CAD ECG signals," *Comput. Biol. Med.*, vol. 94, pp. 19–26, Mar. 2018.

[17] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 664–675, Mar. 2016.

[18] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, A. Gertych, and R. S. Tan, "A deep convolutional neural network model to classify heartbeats," *Comput. Biol. Med.*, vol. 89, pp. 389–396, Oct. 2017.

[19] P. Warrick and M. N. Homsi, "Cardiac Arrhythmia Detection from ECG Combining Convolutional and Long Short-term Memory Networks," in *Proc. Comput. Cardiol. (CinC)*, Sep. 2017, pp. 1–4. doi: 10.22489/CinC.2017.161-460.

[20] Q. Zhang, D. Zhou, and X. Zeng, "HeartID: A multiresolution convolutional neural network for ECG-based biometric human identification in smart health applications," *IEEE Access*, vol. 5, pp. 11805–11816, 2017.

[21] X. Zhang, Y. Zhang, L. Zhang, H. Wang, and J. Tang, "Ballistocardiogram based person identification and authentication using recurrent neural networks," in *Proc. 11th Int. Congr. Image Signal Process., BioMed. Eng. Inform. (CISP-BMEI)*, Beijing, China, Oct. 2018, pp. 1–5.

[22] M. Zihlmann, D. Perekrestenko, and M. Tschannen, "Convolutional recurrent neural networks for electrocardiogram classication," in *Proc. Comput. Cardiol. (CinC)*, Sep. 2017. doi: 10.22489/CinC.2017.070-060.

[23] Ö. Yildirim, "A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification," *Comput. Biol. Med.*, vol. 96, pp. 189–202, May 2018.

[24] (2018). *Remove Trends from Data. MathWorks*. [Online]. Available: https://www.mathworks.com/help/signal/ug/remove-trends-from-data.html

[25] P. S. Hamilton and W. J. Tompkins, "Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database," *IEEE Trans. Biomed. Eng.*, vol. 33, no. 12, pp. 1157–1165, Dec. 1986.

[26] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, Jun. 2000. [Online]. Available: http://circ.ahajournals.org/content/101/23/e215.full

[27] J. L. Elman, "Finding structure in time," *Cognit. Sci.*, vol. 14, no. 2, pp. 179–211, Mar. 1990.

[28] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[29] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *Proc EMNLP*, Oct. 2014, pp. 1724–1734.

HTET MYET LYNN received the B.E. and M.E. degrees from the Department of Computer Engineering, Chosun University, Gwangju, South Korea, in 2014 and 2016, respectively, where he is currently pursuing the Ph.D. degree. His research interests include automatic keyword extraction and automatic summarization from text, ECG-based biometrics human identification, deep learning, and natural language processing.

SUNG BUM PAN received the B.S., M.S., and Ph.D. degrees in electronics engineering from Sogang University, South Korea, in 1991, 1995, and 1999, respectively. He was a Team Leader with the Biometric Technology Research Team of ETRI, from 1999 to 2005. He is currently a Professor with Chosun University. His current research interests include biometrics, security, and VLSI architectures for real-time image processing.

PANKOO KIM received the B.E. degree from Chosun University, in 1988, and the M.S. and Ph.D. degrees in computer engineering from Seoul National University, in 1990 and 1994, respectively. He is currently a Full Professor with Chosun University. His research interests include wireless sensor networks, semantic web techniques, semantic information processing and retrieval, multimedia processing, semantic web, and system security. He is also the Editor-in-Chief of the *IT Convergence Practice* journal (INPRA).