IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 1, 2021, accepted June 22, 2021, date of publication June 28, 2021, date of current version July 5, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3092631

Beat-to-Beat Electrocardiogram Waveform Classification Based on a Stacked Convolutional and Bidirectional Long Short-Term Memory

SITI NURMAINI, (Member, IEEE), ANNISA DARMAWAHYUNI, MUHAMMAD NAUFAL RACHMATULLAH, JANNES EFFENDI, ADE IRIANI SAPITRI, FIRDAUS FIRDAUS, AND BAMBANG TUTUKO

Intelligent System Research Group, Universitas Sriwijaya, Palembang 30139, Indonesia

Corresponding author: Siti Nurmaini (sitinurmaini@gmail.com)

This work was supported by Basic Research Grants (096/SP2H/LT/DRPM/2019) from the Ministry of Research and Technology, Indonesia and Unggulan Profesi Grants 2020 from Universitas Sriwijaya, Indonesia.

This work did not involve human subjects or animals in its research.

ABSTRACT Delineating the electrocardiogram (ECG) waveform is an important step with high significance in cardiology diagnosis. It refers to extract the ECG morphology in start, peak, end points of waveform. Due to various shapes and abnormalities presented in ECG signals, several conventional computer algorithms always fail to extract the essential feature of heart information. Thus, it is critical to investigate an automated ECG signal delineation with its result accuracy. In this study, we propose the delineation process by using bidirectional long short-term memory (BiLSTM) classifier. Such process was conducted as one beat to the next (beat-to-beat), that means the ECG waveform classification is start of P-wave$_1$ to start of P-wave$_2$. However, such classifier lack of feature extraction process, reducing the classification accuracy result. To improve the classifier performance, convolutional layers as facture extraction are stacked with BiLSTM named ConvBiLSTM. We conducted the experimental based on seven-class ECG waveform using a publicly available QT Database with annotation of the main waveforms to produce high accurate classifier, i.e., $P_{start} - P_{end}$, $P_{end} - QRS_{start}$, $QRS_{start} - R_{peak}$, $R_{peak} - QRS_{end}$, $QRS_{end} - T_{start}$, $T_{start} - T_{end}$, and $T_{end} - P_{start}$. It was found that the proposed model showed remarkable results with overall average performances of 99.83% accuracy, 98.82% sensitivity, 99.90% specificity, 98.86% precision, and 98.84% F1 score. Based on these promising results, the efficacy of the proposed stacked ConvBiLSTM model in classifying ECG waveform provides a great opportunity to help cardiologists in diagnosis decision-making for faster assessment.

INDEX TERMS ECG delineation, stacked convolutional layers, bidirectional LSTM, waveform classification.

I. INTRODUCTION

Medical practitioners acquire information about the electrical function of the heart via electrocardiogram (ECG) signals. Each heart cycle of one normal ECG beat contains three main waveforms, i.e., P-wave, QRS complex, and T-wave. These waveforms have a standard amplitude and time duration, indicating various heart conditions [1]. Relevant information from the ECG waveform must be extracted from the physiological signal to diagnose various heart abnormalities [2]. To achieve high diagnostic accuracies, ECG analysis requires the knowledge to extract the morphology of the ECG waves or/and segments (delineation) [1], [3], [4]. For example, to atrial fibrillation (AF) detection, at least we have to know P-wave absence, which it is one of the ECG signal important features. This makes P-wave delineation of great importance for AF detection. However, to delineate and get the knowledge about the location and morphology to detect start, peak, and end point of three main ECG waveforms is quite troublesome.

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

In addition, to diagnose such conditions, ECG signal analysis is time-consuming and requires years of training for specialized knowledge acquisition [5]. Moreover, it suffers from intraindividual variability, such as electrodeposition in common multi-lead ECG, and noise influencing the signal waveform [5]. Therefore, the process of ECG delineation that related to ECG waveform classification with using approach of beat-to-beat segmentation is still needed in clinical practice.

Advanced computing systems can reduce such limitations by permitting the automatic delineation of ECGs. Many researchers have proposed various conventional computer algorithms for ECG delineation. Pan and Tompkins [6] first introduced an automated algorithm to periodically adjust thresholds and parameters to adapt QRS morphology and heart rate. Their algorithm successfully detects 99.3% of QRS complexes. Laguna *et al.* [7] presented an automated algorithm for locating the waveform (the start and ends of the P-wave, QRS complex, and T-wave) in multi-lead ECG signals. Li *et al.* [8] improved and proposed an algorithm based on wavelet transform (WT) to detect ECG characteristic points with the detection rate of QRS complexes above 99.8%. Lin *et al.* [9] proposed a discrete wavelet transform (DWT) for ECG waveform classification. They reconstructed eight-level decompositions of DWT, i.e., P-wave (levels 5 and 6), QRS complex (levels 2 to 4), and T-wave (levels 4 and 5). Although these conventional ECG waveform technique as managed to achieve accuracy above 99.8%, their approaches rely heavily on the accuracy of ECG segmentation and its feature analysis. There can be a high degree of uncertainty and variability due to the subjective aspect of the measurements in the segmentation and measurement phases.

Machine learning (ML) with intelligent processing approaches has been implemented for ECG waveform classification and also obtained promising results, such as Bayesian [10], k-means algorithm [11], hidden Markov model [12], neural network [13], adaptive thresholding [14], and particle swarm optimization [15]. Unfortunately, in a conventional ML, the features are always extracted heuristically and hand-crafted. Deep learning (DL) can automatically extract a hierarchical representation of the data and then utilize the rest of the stacked layers to learn complex features from simpler ones [16], [17]. In contrast to conventional ML, DL may not require extensive human interaction and knowledge for feature design [16], [18]. One of the outstanding DL approaches is long short-term memory (LSTM). LSTM is a variant of recurrent networks used to overcome the gradient problem of recurrent neural networks (RNNs) by multiplicative gates that enforce constant error flow through the internal states of memory cells [19]. LSTM has successfully learned long-term correlations in a sequence of ECG [5], [20].

Due to the superiority of LSTM architecture in prediction capability for both past and future inputs [21], this study proposing a bidirectional phase of LSTM, named bidirectional LSTM (BiLSTM). Such a method is appropriate for sequential data processing based on forward and backward time steps [19], [22]. BiLSTM architecture was concerned for classifying the start and end of beat-to-beat ECG waveform in seven-class i.e., $P_{start} - P_{end}$, $P_{end} - QRS_{start}$, $QRS_{start} - R_{peak}$, $R_{peak} - QRS_{end}$, $QRS_{end} - T_{start}$, $T_{start} - T_{end}$, and $T_{end} - P_{start}$. The ECG waveform is dynamically changed every time depending on the frequency setting and the signal length. However, the BiLSTM technique lacks of feature extraction to recognize the dynamic ECG waveform [23]. Due to this problem, the aim of this study to improve the BiLSTM architecture stacked with a convolutional neural networks (CNNs) architecture as feature extraction to increase the classifier performance. CNNs can generate local features of the ECG signal sequence to recognize regional patterns in the convolution window [24], [25]. The convolution layer of CNNs helps to extract and learn by weight-sharing and modify the low-level hierarchical and invariant features from the raw data [26]. Previous studies about an automated delineation with DL with accurate and precise results are still limited. Besides, stacking of two DL architectures, i.e., convolutional layer and BiLSTM (ConvBiLSTM), has not been explored in depth. Hence, it is imperative to investigate the DL model improvement to increase the delineation waveform ECG result.

To best our knowledge, some previous studies that used DL implementation only limited to three main ECG waveform delineation. Therefore, in this study, we give the contributions and novelty are as follows:

- Developing stacked ConvBiLSTM architecture as feature extraction and classifier models;
- Proposing accurate BiLSTM classifier based on seven-class of ECG waveform;
- Implementing the proposed model for ECG waveform classification that conducted by using beat-to-beat segmentation to simplify the process with a highly accurate result;
- Validating the proposed stacked model with an ECG waveform which manually annotated by experts to insure its performance.

The rest of this paper is organized as follows: Section II presents the related work of this study. Section III describes the theory and background of proposed deep learning architecture. Section IV presents material and method comprising ECG raw data and the proposed stacked ConvBiLSTM architecture, and Section V presents the results and discussion. Finally, the conclusions are presented in Section VI.

II. RELATED WORKS

Deep learning (DL) has been successfully used in various biomedical signal processing, especially for automated ECG waveform classification in several past year. Some intelligent processing methods have been proposed for DL algorithms, such as autoencoder (encoder-decoder), CNNs, and LSTM framework. Londhe and Atulkar [1] conducted a concept of image segmentation for ECG wave segmentation, called semantic segmentation. They proposed a hybrid model

based on ConvBiLSTM to attend the semantic segmentation of ECG waveforms. The input model was 5,000 × 1 time series for P-wave, QRS complex, and T-wave classification. In the convolution layer, batch normalization and ReLU were implemented as activation functions. They proposed two layers of BiLSTM with 250 and 125 units. Overall, they achieved an average accuracy of 97.64%. Wang *et al.* [2] introduced domain knowledge to delineate the fiducial points of ECG signals (P_{on}, P_{off}, QRS_{off}, QRS_{on}, T_{on}, and T_{off}) under the encoder-decoder framework. They allowed the encoder-decoder model to extract ECG features. Sodmann *et al.* [27] implemented CNNs for ECG rhythm annotation. They generated a model with 9-layered CNNs, comprising convolution and pooling layers, batch normalization, activation function, and dropout. The ECG segmentation comprised 450 samples, or 1,500 milliseconds.

Malali *et al.* [28] proposed convolutional LSTM (ConvLSTM) to segment the ECG waves. The input model of a 700 × 4 was fed into the input and convolution layers. They compared each wave segment using the ConvLSTM and BiLSTM models. The results of ConvLSTM outperformed the BiLSTM model, with higher accuracy, sensitivity, and F1-score. They used a QT database from PhysioNet, in which the sample is divided into rhythm segments each of length 700 data points (2.8 seconds). They only focused on segmenting the P-wave, QRS complex, and T-waves and achieved above 92.73% accuracy. Peimankar and Puthusserypady [29] proposed a combination of CNNs and LSTM, named the DENS-ECG algorithm, to detect onset, peak, and offset of P-wave, QRS complex, and T-wave, and most of the incorrect cases in all three classes are classified into No-wave. They generated three convolution layers of CNNs and two layers of BiLSTM. They achieved the performance of the proposed DENS-ECG model in raw and filtered signals above 76.80% and 96.53% sensitivity, respectively. Finally, in our previous work [23], we implemented BiLSTM for P-wave, QRS complex, T-wave, and isoelectric line classification. The proposed method outperformed a unidirectional LSTM with an overall average of 99.64% accuracy. Based on the works mentioned above, DL has successfully proven its ability for ECG delineation.

Based on the previous studies, DL has proposed for ECG delineation. However, those studies were only limited to three ECG waveforms (P-wave, QRS complex, and T-wave) and the performance results under 99% accuracy. With such approaches, this study improved the ECG delineation by using beat-to-beat segmentation into seven-class of ECG waveform classification based on stacked ConvBiLSTM structure. The stacked convolutional layers can learn temporal information from ECG signals followed by BiLSTM to learn short- and long-term dependencies.

III. THEORY AND BACKGROUND

The stacked ConvBiLSTM architecture was conducted to increase the performance of ECG waveform classification. The convolution layers were used only for feature extraction, and ECG waveform classification was processed using BiLSTM.

A. CONVOLUTIONAL LAYERS

A convolution layer, as a part of CNNs, is an automatic extraction of the input model, which can extract deep features from ECG signal data points [1]. The convolution process can be expressed as follows [30]:

$$a_{ij}^{m} = \varphi(b_i + \sum_{k=1}^{M} w_{ik} x_{j+k-1}) = \varphi(b_i + w_i^T x_j) \tag{1}$$

where a_{ij}^{m} is the activation of the jth neuron of the ith filter for the mth convolutional layer, $M$is the kernel size, φ is the neural activation function, b_i is the shared bias of the ith filter, $w_i = [\, w_{i1} \;\; w_{i2} \;\; \ldots w_{iM}\,]^T$ are the shared weights of the ith filter, and $x_j = [\, x_j \;\; x_{j+1} \;\; \ldots \; x_{j+M-1}\,]^T$ are the corresponding M inputs.

B. BIDIRECTIONAL LSTM

To learn short- and long-term dependencies, BiLSTM is a part of RNNs. RNNs in the backward pass often encounter gradient problems, i.e., vanishing or exploding gradients. The gradient problems are caused by the RNNs's iterative nature, whose gradient is essentially equal to the recurrent weight matrix raised to high power. These iterated matrix powers cause the gradient to grow or shrink exponentially in the number of time steps [31]. LSTM tends to overcome this problem by multiplicative gates that enforce constant error flow through the internal states of memory cells (c_t). LSTM learns long-term correlations in a sequence and obviates the need for a prespecified time window [19]. Feedback loops at hidden layers of RNNs are unidirectional. Unidirectional means the process from left to right, in which the flow of the information is only in the forward direction [32]. Schuster presented new concepts of sequence learning in which the information flow is in forward and backward feedback [21]. The connections in the forward direction help us learn from previous representations, and those going backward help us to learn from future representations, called "bidirectional RNNs (BiRNNs)." BiRNNs can be learned to use all available input data for a specific timeframe in the past and future [33]–[35]. With the BiRNNs approach, the BiLSTM equations are used to calculate two parallel directions; forward (f_t) and backward (b_t) passes are given below [36], [37]:

$$\vec{LSTM}_{ft}^{1} = \tanh(W_{i\vec{h}}^{1} x_t + W_{\vec{h}\vec{h}}^{1} \vec{LSTM}_{t-1}^{1} + b_{\vec{h}}^{1} \tag{2}$$

$$\overleftarrow{LSTM}_{bt}^{1} = \tanh(W_{i\overleftarrow{h}}^{1} x_t + W_{\overleftarrow{h}\overleftarrow{h}}^{1} \overleftarrow{LSTM}_{t+1}^{1} + b_{\overleftarrow{h}}^{1}. \tag{3}$$

From (4) and (5), the output of the BiLSTM layer at a time t is:

$$y_t^1 = \tanh(W_{\vec{h}\vec{o}}^{1} \vec{LSTM}_t^1 + W_{\overleftarrow{h}\overleftarrow{o}}^{1} \overleftarrow{LSTM}_t^1 + b_0^1, \tag{4}$$

where the output depends on $\vec{LSTM}_t$ *and* $\overleftarrow{LSTM}_t$; h_0 is initialized as a zero vector.

IV. MATERIAL AND METHOD

A. DATASET

The PhysioNet: QT Database (QTDB) comprised 105 fifteen-minute excerpts of two-channel ECG Holter recordings. The database includes various QRS and ST-T morphologies. The 105 records were chosen primarily from existing ECG databases, including the MIT-BIH Arrhythmia Database, MIT-BIH Normal Sinus Rhythm (NSR DB), the European Society of Cardiology ST-T Database (courtesy of Prof. Carlo Marchesi), and several other ECG databases. However, for this study, we are only used ten records of QTDB: MIT-BIH Normal Sinus Rhythm as baseline for training and validating ECG waveform classification. QTDB has segmented beat by beat that consists of three main ECG waveforms (P-wave (p), QRS-complex (N), and T-wave (t)) in all records. For all experiments conducted in this study, we are only concerned in one complete beat. If there is an incomplete beat, (one of the waveforms is missing) the record is excluded.

QTDB records have record names, e.g. record.dat (contain the signal file), record.hea (describe the format of signal file), record.atr (include the original annotation), record.ari (contain QRS annotation), record.pu0 and.pu1 (contain the automatic waveform start and end), and record.man, record.qtn, and record.qnc (contain manual annotation). In this study, we are only conducted the automatic QRS annotations obtained by record.pu0, which contain the automatic waveform onsets and end in signals 0 as detected by *ecgpuwave*. The QTDB supplied the input to the waveform-database (WFDB) *ecgpuwave* function. The *ecgpuwave* provided the exact location of all waveforms in the signal. The *ecgpuwave* output was written as a standard WFDB-format annotation file related to the specified annotator that utilized as label or 'ground truth' for ECG waveforms.

B. NOISE REMOVAL

ECG signals become corrupted during acquisition due to different types of artifacts and interferences, such as muscle contraction, baseline drift, electrode contact noise, power line interference, etc. [38]–[40]. Generally, discrete wavelet transform (DWT) is used for ECG signal preprocessing (noise removal) because of the properties of a good representation non-stationary signal and the possibility of dividing the ECG signal into different bands of frequency [29], [38]–[40]. The 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 ($h[n]$) [41]. DWT is used to analyze signals at various resolution levels; wavelet coefficients calculate the number of decomposition levels to perform a sequence of signal processing operations [9]. The denoising efficiency is measured using the signal-to-noise ratio (SNR or S/N). SNR provides information about the signal quality. The input SNR (SNR_i) is defined as:

$$SNR_i = 10\log_{10}\left[\frac{\sum_n x^2(n)}{\sum_n r^2(n)}\right] \tag{5}$$

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

$$SNR_o = 10\log_{10}\left[\frac{\sum_n x_d^2(n)}{\sum_n (x_d(n) - x(n))^2}\right], \tag{6}$$

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

In this study, ECG raw data preprocessing was implemented for DWT for noise removal. For each of the mother functions, such as haar, db4, db6, coif4, bior6.8, sym5, sym8, and bior3.5, we calculated SNR values (refer to Table 1). A ratio exceeding 0 decibel (dB) or above 1:1 means more signal than noise. Table 1 shows ten records of QTDB: MIT-BIH Normal Sinus Rhythm, in which the maximum SNR value was reached by bior6.8, with 15.53 dB. Hence, we implemented bior6.8 as the mother function for ECG noise removal.

TABLE 1. The SNR value of QTDB: MIT-BIH normal sinus rhythm database.

Wavelet Mother Functions	An average of SNR (all records)
haar	14.762
db4	15.474
db6	15.502
coif4	15.513
bior6.8	15.537
sym5	15.490
sym8	15.506
bior3.5	15.379

C. STACKED CONVBILSTM

A total of 370 nodes, which was the beat-to-beat (start of P-wave$_1$ to start of P-wave$_2$) was used for input. In our experiments based on QTDB, the maximum node of $P_{start1} - P_{start2}$ is 370. If the beat of $P_{start1} - P_{start2}$ is less than 370 nodes, we perform zero padding technique. This technique was done by adding the value 0 (zero) until the signal has a length of 370 nodes. Every one-dimensional CNNs filter kernels have size of 3 with stride 1. The rectified linear unit (ReLU) function was adopted with four convolution layers (8, 16, 32, and 64 filters). By setting the negative value of the neuron to zero to accelerate the training speed, the ReLU activation function eliminates redundancy. The proposed stacked ConvBiLSTM architecture can be seen in Fig. 1. In addition, Table 2 lists the details of the proposed model, which all the processes can be summarized as follow;

- The ConvBiLSTM architecture consists of two main parts; the ECG waveform feature extraction with convolutional layers and BiLSTM classifier.
- The input of ConvBiLSTM is the ECG waveform which bounded by a vector label indicating the class of each node. The class label is formed in vector with size of (370, 1);
- The input timesteps with dimension (370, 1) are fed into the convolutional layer equipped with ReLU activation

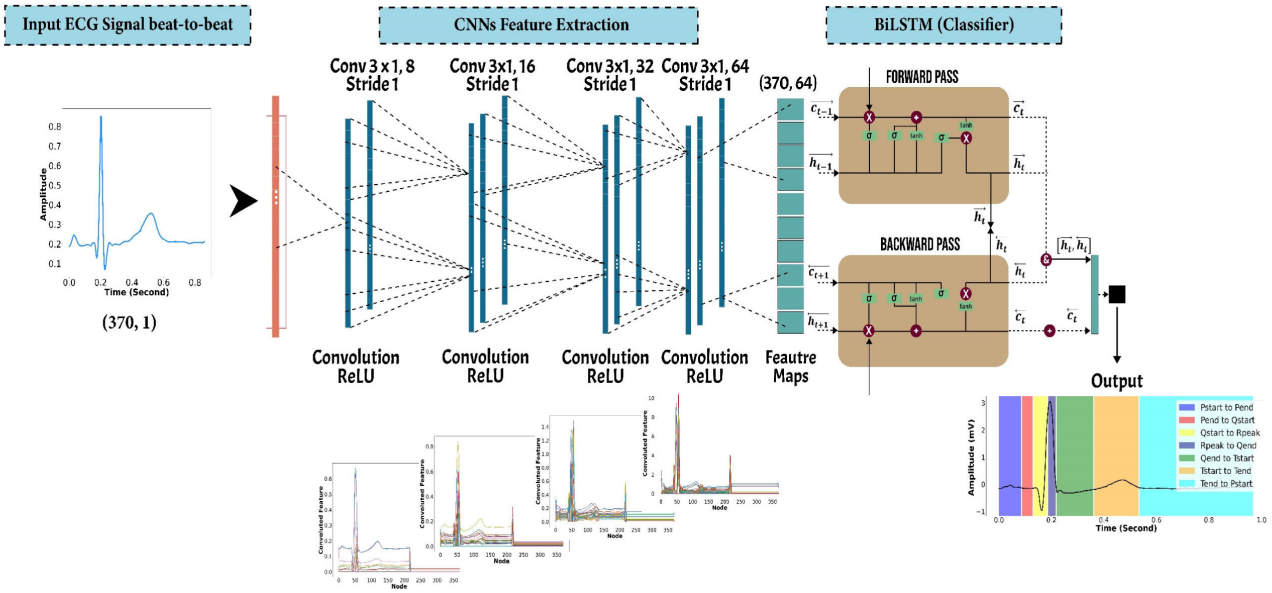

FIGURE 1. The proposed stacked ConvBiLSTM architecture for ECG waveform classification.

function. This process results feature maps with the size of 370, 64;

- The features output from the last convolutional layer was passed into a BiLSTM layer. However, the features need to be adjusted so it can feed into the BiLSTM classifier. Hence, the features were set into 370 timesteps and each timestep contains 64 features.
- The BiLSTM architecture was constructed with 64 BiLSTM cells and 512 nodes for each forward and backward direction (1024 nodes in total). This architecture produces an output size of (370, 8) (including zero-padding).
- This output contains the probability class for each node. One class that has the highest probability value is selected as the output prediction from the model.
- At the end of Fig 1, a vector with size of (370, 1) was formed as the predicted class of each node, from start of P-wave$_1$ to start of P-wave$_2$

We are also provided the informal language to represent pseudocode that can be seen in Algorithm. 1.

For the experiments, this study generated the stacked ConvBiLSTM models for ECG waveform classification in two cases: (1) four-class classification comprised P-wave (P_{start} – P_{end}), QRS complex (QRS_{start} – QRS_{end}), T-wave (T_{start} – T_{end}) and No-wave classification, and (2) seven-class classification. We compared the experiment of the ECG delineation in our previous work [23], which limited to three main ECG waveform (P-wave, QRS-complex, T-wave and other waveforms belong to No-wave class) to this current study. The current study is not only focus to delineate the start and end of three main ECG waveforms points, but also the ECG segments, e.g., P_{end} – QRS_{start}(PR- segment), QRS_{end}– T_{start}(ST-segment), and T_{end} – P_{start}. For the labeling process of four and seven-class cases, we used one-hot label encoding (categorical encoding) to represent each class

Algorithm 1 ConvBiLSTM

Parameters: input x (370,1), output y_t (370,8)

For each epoch **do:**
CNNs Feature Extraction
    For each convolutional layer **do:**
        for each sample in X **do:**
            Calculate a_{ij}^m from X by Eq.1
        End for
        #Dimension of a is $(370 - KernelSize + 1, FilterSize)$
        If $a_{ij}^m length < 370$ **do**:
            Apply zero-pad to a_{ij}^m
            #Dimension of a is (370, $FilterSize$)
        End if
    End for
    #Dimension of a is (370,64)
    #BiLSTM Classifier
    For each sample in a **do:**
        Calculate Forward Pass of a by Eq.2
        Calculate Backward Pass of a by Eq.3
        # Dimension of the output a is (370,2*$NeuronSize$)
        Calculate y_t by Eq.4
    End for
End for

TABLE 2. The description of stacked ConvBiLSTM features.

Layer	Input Nodes	Filter Number	Kernel Size/Pool Size	Output Nodes	Feature Interpretation
Input	370,1	-		-	ECG waveform for one beat
Convolution 1	370 x1	8	3x1, stride 1	370 x 8	8 feature maps
Convolution 2	370 x 8	16	3x1, stride 1	370 x 16	16 feature maps
Convolution 3	370 x 16	32	3x1, stride 1	370 x 32	32 feature maps
Convolution 4	370 x 32	64	3x1, stride 1	370 x 64	64 feature maps
BiLSTM Input	370, 64	-	-	-	Output feature from CNNs, with 370 timesteps and 64 features for each step.
BiLSTM	370 x 64	-	-	370 x 1024	Two direction feature data (512 nodes for both forward and backward directions)
Fully Connected	-	-	-	370 x 8	370 nodes containing seven-class probability of ECG waveforms (include zero-padding)
Output	-	-	-	370, 1	370 nodes containing the maximum probability of seven-class ECG waveforms (including zero-padding)

using scikit-learn Python library. Label encoding is a technique for handling categorical variables which assigned a unique integer.

Each fine-tuned hyperparameters of the two cases is listed in Table 3. We fine-tuned followed by varying learning rates ranging from 10^{-1} to 10^{-5} (Models 1–5), and then the hidden layers number of LSTM (Models 6–9) for four-class classifications. The learning rate is the most vital hyperparameter, which controls how quickly the model is adapted to the learning problems. The selected learning rate based on the result performance is presented in Table 4. As we can see in Table 4, model 5 outperformed models 1-4 in accuracy, sensitivity, specificity, precision, and F1-score measurement. The selected learning rate was implemented to fine-tune the hidden layer from one to five hidden layers. The best model of nine models for four-class classification cases that used unidirectional LSTM was compared with the bidirectional LSTM (Model 10). We also conducted the convolution layers and BiLSTM classifier for Model 11. For seven-class classification, we generated four models of stacked ConvBiLSTM structure. The difference between the hyperparameter models lies only in the batch size value (1, 2, 4, and 8).

D. PLATFORM

We conducted an experiment on a workstation with one Intel(R) Core (TM) i7-10700K CPU@ 3.80Ghz processor with 32 GB RAM, and one NVIDIA GeForce RTX 2070 Super 8GB GPU. All experiments were run on Windows 10 Pro 64 Bit. We implemented our Python codes in the Spyder 4.1.5 deep learning framework and libraries, i.e., tensorflow, keras, numpy, pandas, sklearn, matplotlib, and the native Python WFDB package. Our experiments compared the stacked ConvBiLSTM models using some metrics to evaluate the experimental performance: accuracy, sensitivity, specificity, precision, and F1-score.

V. RESULTS AND DISCUSSION

The ECG signal was segmented to achieve a fixed window size of 370 nodes, starting from P_{start1} to P_{start2} (start of P-wave$_1$ to start of P-wave$_2$), which contained one heartbeat at a normal heart rate. The sample of one normal heartbeat can be presented in Fig. 2. A total of 8,572 beats comprised 7,715 training and 857 validation sets. The model was trained over 300 epochs, with a learning rate of 0.0001 and categorical cross entropy as the loss metric. For classification tasks, we have implemented common performance metrics (accuracy, sensitivity, specificity, precision, and F1-score) that calculated and obtained based on validation (refer Tables 5 and 7) and testing dataset (Table 8).

As stated before, this study generated a two-case model (refer to Table 3): four and seven-class of ECG waveform classification. Table 4 shows the accuracy, sensitivity,

TABLE 3. The fine-tuned hyperparameters of stacked ConvBiLSTM architecture for four and seven-class classification.

Class Total	Model	Layers	Learning Rate	Hidden Layer	Batch Size
Four-class	1	LSTM	10^{-1}	1	8
	2	LSTM	10^{-2}	1	8
	3	LSTM	10^{-3}	1	8
	4	LSTM	10^{-4}	1	8
	5	LSTM	10^{-5}	1	8
	6	LSTM	10^{-5}	2	8
	7	LSTM	10^{-5}	3	8
	8	LSTM	10^{-5}	4	8
	9	LSTM	10^{-5}	5	8
	10	BiLSTM	10^{-5}	1	8
	11	Convolution 8 x 3, 16 x 3, 32 x 3, 64 x 3, strides= 1 + ReLU - BiLSTM	10^{-5}	1	8
Seven-class	1	Convolution 8 x 3, 16 x 3, 32 x 3, 64 x 3, strides= 1 + ReLU - BiLSTM	10^{-5}	1	1
	2	Convolution 8 x 3, 16 x 3, 32 x 3, 64 x 3, strides= 1 + ReLU - BiLSTM	10^{-5}	1	2
	3	Convolution 8 x 3, 16 x 3, 32 x 3, 64 x 3, strides= 1 + ReLU - BiLSTM	10^{-5}	1	4
	4	Convolution 8 x 3, 16 x 3, 32 x 3, 64 x 3, strides= 1 + ReLU - BiLSTM	10^{-5}	1	8

TABLE 4. The average performance of four and seven-class ECG waveform classification models.

Class Total	Model	Average Performance (%)				
		Accuracy	Sensitivity	Specificity	Precision	F1-score
Four-class	1	24.78	25	75	1.45	2.75
	2	24.71	25	75	1.46	2.75
	3	99.43	97.96	99.6	98.19	98.08
	4	99.64	98.71	99.75	98.8	98.75
	5	99.35	97.47	99.52	98.77	97.62
	6	99.68	98.84	99.78	98.96	98.9
	7	99.69	98.84	99.78	99.04	98.94
	8	99.67	98.81	99.77	98.92	98.87
	9	99.69	98.93	99.78	98.98	98.96
	10	99.68	98.84	99.78	98.96	98.90
	11	99.69	98.91	99.79	99.01	98.96
Seven-class	1	99.83	98.79	99.90	98.86	98.83
	2	99.83	98.82	99.90	98.86	98.84
	3	99.83	98.76	99.90	98.83	98.80
	4	99.82	98.75	99.90	98.80	98.78

specificity, precision, and F1-score for each case. In four-class classification, we obtained the worst performance in models 1 and 2, in which the models used a learning rate 10^{-1}. In order to enhance the performance, we tried a smaller learning rate from 10^{-3} to 10^{-5}. Consequently, the performances increased in terms of accuracy, sensitivity,

TABLE 5. The performance of the selected model for seven-class ECG waveform classification.

Model	Metrics	Class Performance (%)							Average
		P_{start}-P_{end}	P_{end}-Q_{start}	Q_{start}-R_{peak}	R_{peak}-Q_{end}	Q_{end}-T_{start}	T_{start}-T_{end}	T_{end}-P_{start}	
Proposed	Accuracy	99.89	99.82	99.93	99.97	99.69	99.61	99.90	99.83
	Sensitivity	98.96	98.29	98.50	99.64	97.94	98.58	99.81	98.82
	Specificity	99.94	99.89	99.97	99.98	99.85	99.76	99.93	99.90
	Precision	99.09	97.81	99.10	99.52	98.32	98.37	99.80	98.86
	F1-score	99.03	98.05	98.80	99.58	98.13	98.48	99.80	98.84

TABLE 6. The comparison results of the selected model (model 2) without the stacked convolutional layers.

Model	Performance (%)				
	Accuracy	Sensitivity	Specificity	Precision	F1-score
BiLSTM	99.81	98.68	99.90	98.72	98.70
Stacked ConvBiLSTM	99.83	98.82	99.90	98.86	98.84

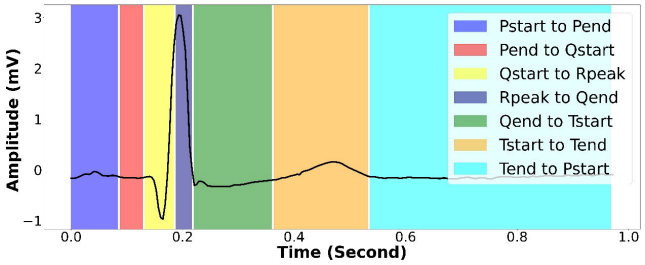

FIGURE 2. The sample of ECG beat at normal heart rate from $P_{start1} - P_{start2}$.

specificity, precision, and F1 score. All performances were achieved above 97.47% (Models 2–5). Still, to get a better performance, we also fine-tuned the models with different hidden layers, from one to five layers (Models 6–9). The performances obtained were above 98.81%. From the results, it can be concluded that the learning rate and hidden layers affected the performance of ECG waveform classification. Although smaller learning rates required more training epochs, but it can give much better performance results. It allowed the model to learn a more optimal or even globally optimal set of weights.

A unidirectional LSTM layer was implemented for a four-class classification model. Among the models, the best model was also compared to the BiLSTM and stacked ConvBiLSTM models represented in Models 10 and 11, respectively. Although the gap results from both models did not differ significantly, overall, Model 11 outperformed Model 10 in all metrics performance. We employed all fine-tuned hyperparameters for the seven-class classification model using this model. Besides learning rate and hidden layers, we also fine-tuned a batch size (1, 2, and 4). From Models 1–4, the results' accuracy and specificity performance did not differ considerably, around 99.83% and 99.90%, respectively. The delineation result of each model is almost the same as the *ecgpuwave* annotation (refer to Fig. 3). However, Model 2 achieved the highest sensitivity and F1-score: 98.82% and 98.84%. Overall, Model 2 had the best performance among other models, with an average of 99.83% accuracy, 98.82% sensitivity, 99.90% specificity, 98.85% precision, and 98.84% F1-score. Therefore, Model 2 was selected as the best model for ECG waveform classification, with the average result shown in Table 5. From Model 2, we obtained the highest performance of QRS complex. It can be stated that the selected models can detect QRS complex more accurately and faster. Our model could outperform some conventional methods for QRS complex detection, such as low-pass differentiator [42], Hilbert transform [43], multiple higher order moments [33], Phasor transform [34], etc.

In the experiment, we generated the BiLSTM model without the convolution layers of CNNs for seven-class classification. The results were quite decreased in accuracy, sensitivity, specificity, precision, and F1-score (see Table 6). As we suggested, the convolution layers will improve the performance results, showing the locations and strength of a detected function in an input due to the same filter to an input results in a map of activations (feature map). Its ability insures an automatic learning of many filters in parallel specific to a training dataset. In the experiments, the convolution layers contained a filter set whose parameters were required to be learned. To compute an activation map made of neurons, each filter was convoluted with the input volume.

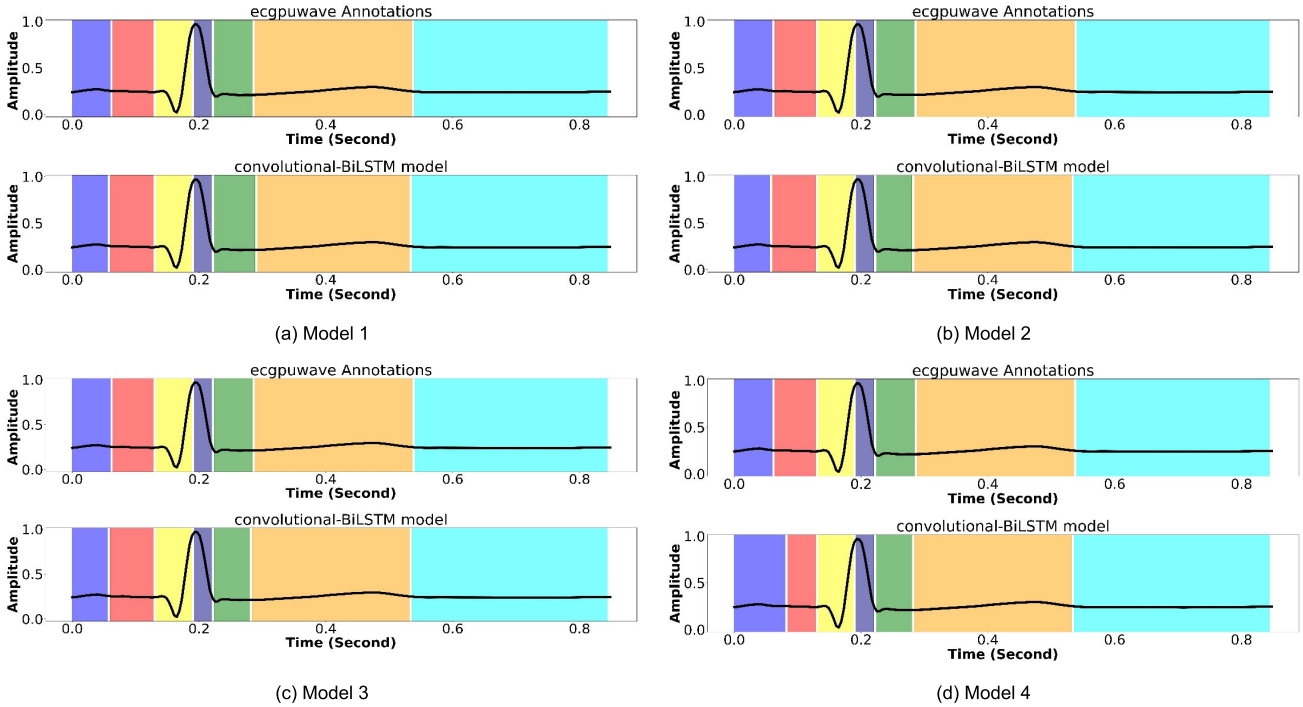

FIGURE 3. The results comparison ECG waveform of ecgpuwave and stacked ConvBiLSTM models for seven-class case.

TABLE 7. The results performance of the abnormal database from model 2 with stacked ConvBiLSTM architecture.

Database	Average Performance (%)				
	Accuracy	Sensitivity	Specificity	Precision	F1-score
MIT-BIH Arrhythmia	99.63	96.94	99.80	97.07	97.00
MIT-BIH Long-Term	99.09	94.85	99.47	95.22	95.03
MIT-BIH ST Change	99.74	97.56	99.86	97.54	97.55
MIT-BIH Supraventricular Arrhythmia	99.62	97.17	99.79	97.21	97.19
European ST-T	99.65	97.14	99.81	97.21	97.17

TABLE 8. The performance evaluation of unseen data testing (expert/.q1c).

Metrics Performance (%)	Class performance (%)							Average
	P_{start}- P_{end}	P_{end}-Q_{start}	Q_{start}-R_{peak}	R_{peak}-Q_{end}	Q_{end}-T_{start}	T_{start}- T_{end}	T_{end}-P_{start}	
Accuracy	99.31	98.67	98.47	98.67	96.51	96.44	99.46	98.22
Sensitivity	90.87	89.02	73.89	86.18	98.40	78.93	99.38	88.09
Specificity	99.95	99.23	99.44	99.01	96.35	99.72	99.50	99.03
Precision	99.28	86.91	83.68	70.25	69.13	98.14	98.91	86.62
F1-score	94.89	87.95	78.48	77.40	81.21	87.49	99.15	86.65

The proposed stacked ConvBiLSTM was also trained in some abnormal databases that contain the several heart disorders, such as MIT-BIH Arrhythmia, MIT-BIH long-term, MIT-BIH ST change, MIT-BIH supraventricular arrhythmia, and European ST-T (refer to Fig. 4). For MIT-BIH arrhythmia, the average accuracy, sensitivity, specificity, precision, and F1-score exceeded 93%. The database is more successful in detecting P-wave (blue bar) than QRS complex and

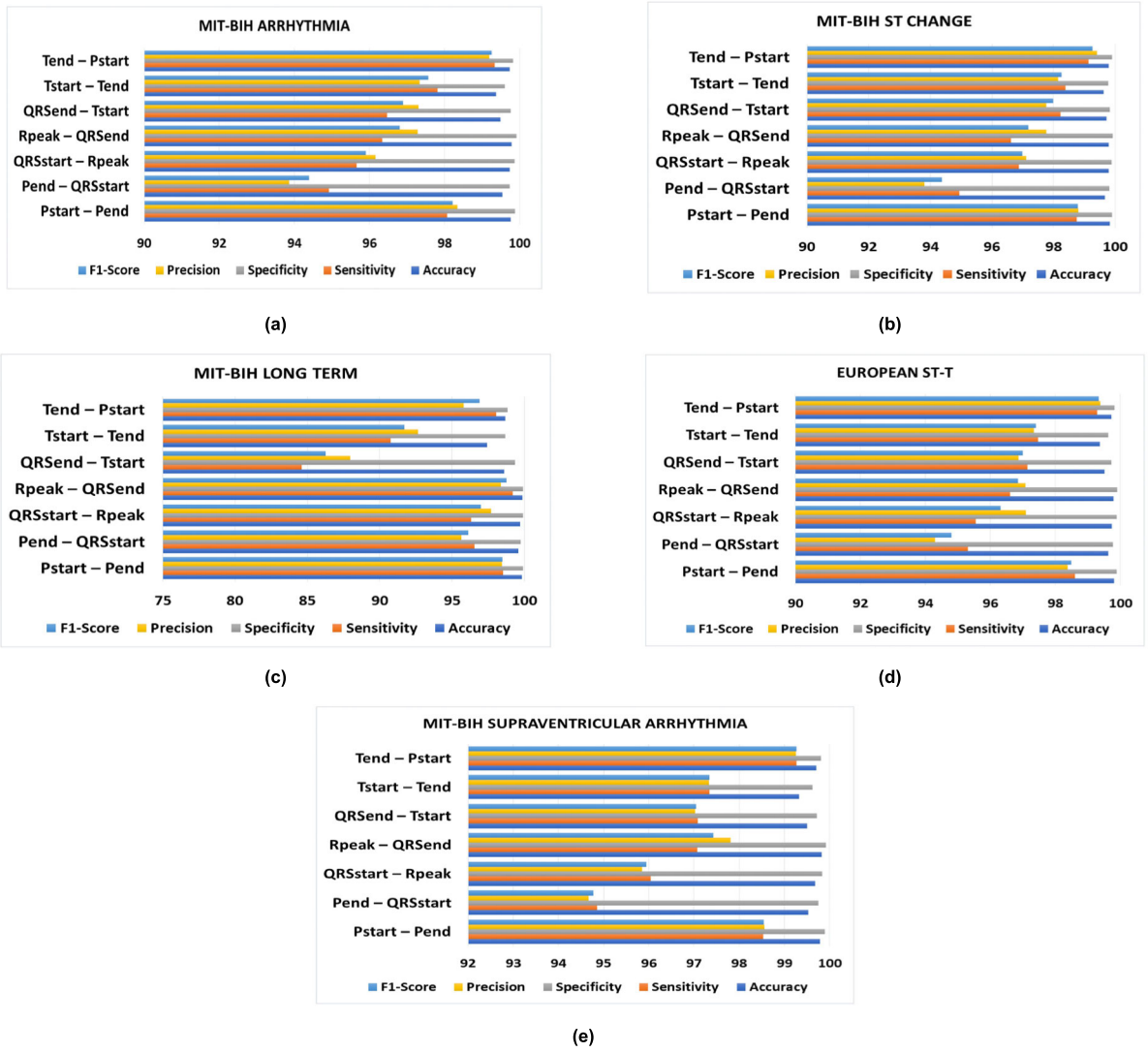

FIGURE 4. The results performance of seven-class classification for abnormal database.

T-wave. It can be happened due to the arrhythmia is associated with prolongation P-wave was reflected the atrial conduction. Thereof, our proposed model could be implemented in some arrhythmia cases, such as AF. AF can be related to irregular supraventricular tachycardia (SVT) without P-wave and irregularly irregular QRS complex [35].

The performance results decreased when it was trained on MIT-BIH long-term, which the worst case was around 85% sensitivity for Q_{end}-T_{start} (orange bar) class. The average performance exceeded 94.85% for accuracy, sensitivity, specificity, precision, and F1-score. The performance results of MIT-BIH ST change, MIT-BIH supraventricular arrhythmia, and European ST-T databases achieved good results and were above 99.62% accuracy. The worst case in the MIT-BIH ST change database was P_{end}-Q_{start} or PR-segment (yellow bar) with only 93.81% precision. This also applied to MIT-BIH supraventricular arrhythmia and European ST-T databases; the precision of the PR segment was only 94.67% and 94.30%, respectively. A moderately common ECG sign associated with clinically silent pericardial effusion was PR-segment depression, and it was an ECG predictor of inflammatory pericardial involvement. Changes in the PR-segment are relative to the baseline formed by the T-P segment.

As a result of average performance, the selected model was well trained on MIT-BIH ST change, with all performances exceeding 97.54% (refer to Table 7). Unfortunately, for MIT-BIH long-term, the sensitivity was only achieved at 94.85%. The other performances were below other abnormal databases. A too long signal recording could cause a normal T-wave to overlap with other T-wave classes: inverted, only upwards, only downwards, biphasic negative-positive, or biphasic positive-negative. Due to the raw data condition, in which the two signals differed, the long-term signal had more noise compared to the ST change.

TABLE 9. The benchmark study of ECG waveform classification.

Metrics Performance (%)	Class	CNNs-BiLSTM) [1]	CNNs-LSTM) [29]	CNNs [27]	CNNs-LSTM [28]	(BiLSTM) [23]	Current work (Stacked ConvBiLSTM)
Accuracy	P_{start}-P_{end}	98.20	-	-	94.87	99.84	99.89
	P_{end}-Q_{start}	-	-	-	-	-	99.82
	Q_{start}-R_{peak}	98.56	-	-	96.66	99.89	99.95
	R_{peak}-Q_{end}						
	Q_{end}-T_{start}	-	-	-	-	-	99.69
	T_{start}-T_{end}	96.17	-	-	92.73	99.54	99.61
	T_{end}-P_{start}	-	-	-	-	-	99.90
Sensitivity	P_{start}-P_{end}	88.66	96.53	92	93.30	98.38	98.96
	P_{end}-Q_{start}	-	-	-	-	-	98.29
	Q_{start}-R_{peak}	94.05	99.70	98	95.57	99.10	99.07
	R_{peak}-Q_{end}						
	Q_{end}-T_{start}	-	-	-	-	-	97.94
	T_{start}-T_{end}	90.50	96.81	88	94.56	98.47	98.58
	T_{end}-P_{start}	-	-	-	-	-	99.81
F1-score	P_{start}-P_{end}	89.86	93.01	-	94.24	98.69	99.03
	P_{end}-Q_{start}	-	-	-	-	-	98.05
	Q_{start}-R_{peak}	94.81	99.45	-	96.70	99.17	99.19
	R_{peak}-Q_{end}						
	Q_{end}-T_{start}	-	-	-	-	-	98.13
	T_{start}-T_{end}	92.15	96.12	-	93.04	98.20	98.48
	T_{end}-P_{start}	-	-	-	-	-	99.80

In order to test the generalization of proposed model, the stacked ConvBiLSTM was tested on unseen data that were manually annotated by experts. Unseen data is a set of testing data that is never seen or tested before. If the model produces a good generalization, it may make a correct prediction in unseen data. Although the QT database was provided to two experts for manual annotation of the ECG waveform classification, only one expert was used as unseen data testing in this study, i.e., record (.q1c). When the proposed stacked ConvBiLSTM was tested in all records (.q1c), the results were decreased and obtained an average of 88.09% sensitivity, 86.62% precision, and 86.65% F1-score (refer to Table 8). Overall, for the QRS complex, we obtained a poor performance with the worst class classification of 69.13% precision for Q_{end}-T_{start} (ST-segment). It can be affected by the elevation between depolarization and repolarization of ventricles of ST-segment. Unfortunately, for this study, we didn't concern of a displacement of ST segment.

Different to *ecgpuwave* annotation, all the records were manually annotated by experts, which only provides between 30 and 100 representative beats, not in all lengths of records. Only normal heart rate beats were annotated. In records with significant QRS morphology, around 20 beats of each non-dominant morphology were also annotated. Therefore, the proposed model obtained the poor performance for ST-segment classification.

Some previous studies has adopted record.q1c using deep learning algorithm [4], [44]. Abrishami *et al.*, [4] explored P-wave detection using fully-connected networks, LeNet-style convolutional networks (ConvNet) with dropout, and LeNet-style ConvNet without dropout. Costandy *et al.*, [44] had also proposed a fully convolutional networks with the use of segmentation via U-Net architecture. However, they experimented the ECG signals that converted to two-dimensional (2-D) ECG images. Our proposed stacked ConvBiLSTM outperformed the performance results of P-wave detection of previous mention studies with 99.31% accuracy (refer Table 8).

The difference in ECG waveform results annotated by experts and the our proposed stacked ConvBiLSTM are shown in Fig. 5. A green block color (Q_{end}–T_{start}) was misclassified as an orange block color (T_{start}–T_{end}). Owing to the problem, the selected stacked ConvBiLSTM model had poor performance. However, this could happen because the record

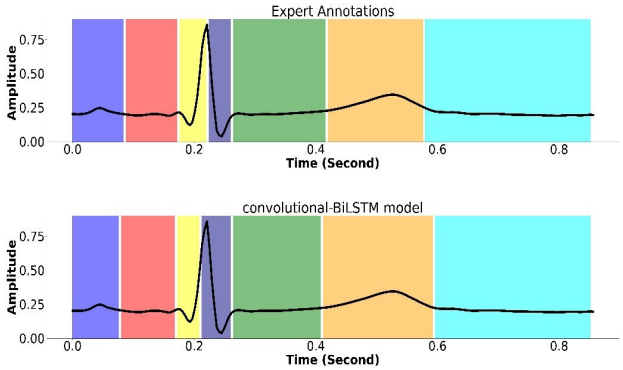

FIGURE 5. The sample results of manually annotated by expert and proposed stacked ConvBiLSTM model.

(.q1c) annotations of QT Database were audited to correct the inconsistencies detected (e.g., misplaced or missing annotations) and changed to the regular annotation symbols. The manual annotations were made by experts using a SUN workstation with waveform analyzer, viewer, and editor (WAVE) tools. The tools still have variance noises, such as baseline wander, electrode motion artifact, and muscle noise, which can affect the testing performance of experts. Unfortunately, this study was unconcerned about the aforementioned noises. Still, the selected model of stacked ConvBiLSTM that used *ecgpuwave* annotations still has many frictions, especially for P- and T-wave classification.

In this study, we have compared ECG waveform classification between our proposed model and other DL techniques such as encoder-decoder, CNNs, and LSTM (refer to Table 9). The previous research has classified the main ECG waveform, from P_{start} - T_{end}. Different to them, we added the isoelectric line from T_{end} – P_{start2}. In our previous work [23], we conducted a BiLSTM model for P-wave, QRS complex, T-wave, and isoelectric lines. We achieved a good performance with an average accuracy, sensitivity, and F1-score of 99.64%, 98.74%, and 98.78%, respectively. However, to achieve greater generalization than we stated in our previous study [23], we had to add automatic feature extraction.

Although the results look promising, there are some limitations to our study:

- To generate the proposed model, we only used the ground truths or label annotated by the *ecgpuwave* due to the limitation of manual annotation by an expert.
- The preprocessing steps for ECG delineation still need adjustment to differ ECG frequency sampling, leads, and various noises.
- The process of ECG delineation, we only concerned to determine the start and end of ECG waveform without considering the specific heart disorders.
- The proposed model was only validated by a limited expert. We did not validate the proposed model in other ECG delineation databases; more datasets could achieve greater generalization.

VI. CONCLUSION

To provide a heart diagnosis, ECG wave information has become a gold standard. Heart wave formation gives some fiducial points to represent the abnormality, such as P-wave, QRS complex, and T-wave. Automated ECG delineation is a crucial yet insufficiently addressed area in automated heart diagnostics. This study explored automated delineation without using the fiducial point, but it directly segmented beat-to-beat from ECG waveform signal recording. We proposed the stacked DL method by synergizing CNNs for spatial feature extraction and BiLSTM to classify each sample of the ECG waveform into one of the seven-class ECG waveform classification. Utilizing such feature extractors addresses the various heart wave formations that can be easily recognized for an accurate diagnosis. Feature vectors generated by the CNNs are the input timesteps to the sequence learner architecture through a time distributed layer in the BiLSTM architecture. BiLSTM captured the temporal attributes of ECG signals and classified them to produce an accurate result. Our proposed stacked model showed outstanding results, with all average performances ranging of 99.83% accuracy, 98.82% sensitivity, 99.90% specificity, 98.86% precision, and 98.84% F1 score. Our study demonstrated that the stacked ConvBiLSTM model is a powerful network able to capture the temporal attribute of the ECG signal using beat-to-beat as local features to yield ambitious results. The ability to segment, identify, and classify heart waveforms augments the possibility of impacting future research in cardiology. Also, it can be implemented in clinical practice to interpret some specific heart abnormalities that related to delineation process with real-time performance analysis.

ACKNOWLEDGMENT

The authors thank to Intelligent System Research Group (ISysRG) Universitas Sriwijaya, Indonesia, for full support in their research infrastructure.

REFERENCES

[1] A. N. Londhe and M. Atulkar, "Semantic segmentation of ECG waves using hybrid channel-mix convolutional and bidirectional LSTM," *Biomed. Signal Process. Control*, vol. 63, Jan. 2021, Art. no. 102162.
[2] J. Wang, R. Li, R. Li, and B. Fu, "A knowledge-based deep learning method for ECG signal delineation," *Future Gener. Comput. Syst.*, vol. 109, pp. 56–66, Aug. 2020.
[3] S. Kiranyaz, T. Ince, R. Hamila, and M. Gabbouj, "Convolutional neural networks for patient-specific ECG classification," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 2608–2611.
[4] H. Abrishami, M. Campbell, C. Han, R. Czosek, and X. Zhou, "P-QRS-T localization in ECG using deep learning," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Informat. (BHI)*, Mar. 2018, pp. 210–213.
[5] O. Faust, A. Shenfield, M. Kareem, T. R. San, H. Fujita, and U. R. Acharya, "Automated detection of atrial fibrillation using long short-term memory network with RR interval signals," *Comput. Biol. Med.*, vol. 102, pp. 327–335, Nov. 2018.
[6] 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.
[7] P. Laguna, R. Jané, and P. Caminal, "Automatic detection of wave boundaries in multilead ECG signals: Validation with the CSE database," *Comput. Biomed. Res.*, vol. 27, no. 1, pp. 45–60, Feb. 1994.
[8] C. Li, C. Zheng, and C. Tai, "Detection of ECG characteristic points using wavelet transforms," *IEEE Trans. Biomed. Eng.*, vol. 42, no. 1, pp. 21–28, Jan. 1995.

[9] H.-Y. Lin, S.-Y. Liang, Y.-L. Ho, Y.-H. Lin, and H.-P. Ma, "Discrete-wavelet-transform-based noise removal and feature extraction for ECG signals," *IRBM*, vol. 35, no. 6, pp. 351–361, Dec. 2014, doi: 10.1016/j.irbm.2014.10.004.
[10] C. Lin, C. Mailhes, and J.-Y. Tourneret, "P- and T-wave delineation in ECG signals using a Bayesian approach and a partially collapsed Gibbs sampler," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 12, pp. 2840–2849, Dec. 2010.
[11] S. S. Mehta, D. A. Shete, N. S. Lingayat, and V. S. Chouhan, "K-means algorithm for the detection and delineation of QRS-complexes in electrocardiogram," *IRBM*, vol. 31, no. 1, pp. 48–54, Feb. 2010.
[12] M. Akhbari, M. B. Shamsollahi, O. Sayadi, A. A. Armoundas, and C. Jutten, "ECG segmentation and fiducial point extraction using multi hidden Markov model," *Comput. Biol. Med.*, vol. 79, pp. 21–29, Dec. 2016.
[13] J. Duan and Y. Hao, "ECG P wave based on wavelet transform and neural network," *J. Biomed. Eng. Res.*, vol. 24, no. 4, pp. 209–211, 2005.
[14] R. Gutiérrez-Rivas, J. J. Garcia, W. P. Marnane, and A. Hernández, "Novel real-time low-complexity QRS complex detector based on adaptive thresholding," *IEEE Sensors J.*, vol. 15, no. 10, pp. 6036–6043, Oct. 2015.
[15] S. Jain, A. Kumar, and V. Bajaj, "Technique for QRS complex detection using particle swarm optimisation," *IET Sci., Meas. Technol.*, vol. 10, no. 6, pp. 626–636, Sep. 2016.
[16] S. Khan and T. Yairi, "A review on the application of deep learning in system health management," *Mech. Syst. Signal Process.*, vol. 107, pp. 241–265, Jul. 2018.
[17] S. Nurmaini, R. U. Partan, W. Caesarendra, T. Dewi, M. N. Rahmatullah, A. Darmawahyuni, V. Bhayyu, and F. Firdaus, "An automated ECG beat classification system using deep neural networks with an unsupervised feature extraction technique," *Appl. Sci.*, vol. 9, no. 14, p. 2921, Jul. 2019, doi: 10.3390/app9142921.
[18] S. Min, B. Lee, and S. Yoon, "Deep learning in bioinformatics," *Brief Bioinform.*, vol. 18, no. 5, pp. 851–869, 2016.
[19] P. Malhotra, L. Vig, G. Shroff, and P. Agarwal, "Long short term memory networks for anomaly detection in time series," in *Proc. Eur. Symp. Artif. Neural Netw., Comput. Intell. Mach. Learn. (ESANN)*, Bruges, Belgium, Apr. 2015, pp. 89–94. [Online]. Available: http://www.i6doc.com/en/
[20] S. Chauhan and L. Vig, "Anomaly detection in ECG time signals via deep long short-term memory networks," in *Proc. IEEE Int. Conf. Data Sci. Adv. Analytics (DSAA)*, Oct. 2015, pp. 1–7.
[21] M. Schuster, "On supervised learning from sequential data with applications for speech recognition," Ph.D. dissertation, Nara Inst. Sci. Technol., Ikoma, Japan,1999.
[22] A. Darmawahyuni, S. Nurmaini, and Sukemi, "Deep learning with long short-term memory for enhancement myocardial infarction classification," in *Proc. 6th Int. Conf. Instrum., Control, Autom. (ICA)*, Jul. 2019, pp. 19–23, doi: 10.1109/ICA.2019.8916683.
[23] S. Nurmaini, A. E. Tondas, A. Darmawahyuni, M. N. Rachmatullah, J. Effendi, F. Firdaus, and B. Tutuko, "Electrocardiogram signal classification for automated delineation using bidirectional long short-term memory," *Informat. Med. Unlocked*, vol. 22, Jan. 2021, Art. no. 100507.
[24] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1D convolutional neural networks and applications: A survey," 2019, *arXiv:1905.03554*. [Online]. Available: http://arxiv.org/abs/1905.03554
[25] S. Nurmaini, A. E. Tondas, A. Darmawahyuni, M. N. Rachmatullah, R. U. Partan, F. Firdaus, B. Tutuko, F. Pratiwi, A. H. Juliano, and R. Khoirani, "Robust detection of atrial fibrillation from short-term electrocardiogram using convolutional neural networks," *Future Gener. Comput. Syst.*, vol. 113, pp. 304–317, Dec. 2020, doi: 10.1016/j.future.2020.07.021.
[26] G. B. Kshirsagar and N. D. Londhe, "Weighted ensemble of deep convolution neural networks for single-trial character detection in Devanagari-script-based P300 speller," *IEEE Trans. Cognit. Develop. Syst.*, vol. 12, no. 3, pp. 551–560, Sep. 2020.
[27] P. Sodmann, M. Vollmer, N. Nath, and L. Kaderali, "A convolutional neural network for ECG annotation as the basis for classification of cardiac rhythms," *Physiol. Meas.*, vol. 39, no. 10, Oct. 2018, Art. no. 104005.
[28] A. Malali, S. Hiriyannaiah, G. M. Siddesh, K. G. Srinivasa, and N. T. Sanjay, "Supervised ECG wave segmentation using convolutional LSTM," *ICT Exp.*, vol. 6, no. 3, pp. 166–169, 2020.
[29] A. Peimankar and S. Puthusserypady, "DENS-ECG: A deep learning approach for ECG signal delineation," 2020, *arXiv:2005.08689*. [Online]. Available: http://arxiv.org/abs/2005.08689
[30] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep Learning*, vol. 1. Cambridge, MA, USA: MIT Press, 2016.
[31] R. Jozefowicz, W. Zaremba, and I. Sutskever, "An empirical exploration of recurrent network architectures," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 2342–2350.
[32] H. Zen and H. Sak, "Unidirectional long short-term memory recurrent neural network with recurrent output layer for low-latency speech synthesis," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 4470–4474.
[33] A. Ghaffari, M. R. Homaeinezhad, M. Khazraee, and M. M. Daevaeiha, "Segmentation of holter ECG waves via analysis of a discrete wavelet-derived multiple skewness–kurtosis based metric," *Ann. Biomed. Eng.*, vol. 38, no. 4, pp. 1497–1510, Apr. 2010.
[34] A. Mart'inez, R. Alcaraz, and J. J. Rieta, "Application of the phasor transform for automatic delineation of single-lead ECG fiducial points," *Physiol. Meas.*, vol. 31, no. 11, p. 1467, 2010.
[35] L. S. B. Johnson, A. P. Persson, P. Wollmer, S. Juul-Möller, T. Juhlin, and G. Engström, "Irregularity and lack of p waves in short tachycardia episodes predict atrial fibrillation and ischemic stroke," *Heart Rhythm*, vol. 15, no. 6, pp. 805–811, Jun. 2018.
[36] A. Graves, "Supervised sequence labelling," in *Supervised Sequence Labelling With Recurrent Neural Networks*. Berlin, Germany: Springer-Verlag, 2012, pp. 5–13.
[37] F. Zhu, F. Ye, Y. Fu, Q. Liu, and B. Shen, "Electrocardiogram generation with a bidirectional LSTM-CNN generative adversarial network," *Sci. Rep.*, vol. 9, no. 1, p. 6734, Dec. 2019.
[38] R. Sameni, M. B. Shamsollahi, C. Jutten, and G. D. Clifford, "A nonlinear Bayesian filtering framework for ECG denoising," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 12, pp. 2172–2185, Dec. 2007, doi: 10.1109/TBME.2007.897817.
[39] B. H. Tracey and E. L. Miller, "Nonlocal means denoising of ECG signals," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 9, pp. 2383–2386, Sep. 2012, doi: 10.1109/TBME.2012.2208964.
[40] J. Wang, Y. Ye, X. Pan, and X. Gao, "Parallel-type fractional zero-phase filtering for ECG signal denoising," *Biomed. Signal Process. Control*, vol. 18, pp. 36–41, Apr. 2015, doi: 10.1016/j.bspc.2014.10.012.
[41] A. Darmawahyuni, S. Nurmaini, M. Yuwandini, M. N. Rachmatullah, F. Firdaus, and B. Tutuko, "Congestive heart failure waveform classification based on short time-step analysis with recurrent network," *Informat. Med. Unlocked*, vol. 21, 2020, Art. no. 100441.
[42] P. Laguna, N. V. Thakor, P. Caminal, R. Jané, H.-R. Yoon, A. B. de Luna, V. Marti, and J. Guindo, "New algorithm for QT interval analysis in 24-hour holter ECG: Performance and applications," *Med. Biol. Eng. Comput.*, vol. 28, no. 1, pp. 67–73, Jan. 1990.
[43] S. K. Mukhopadhyay, M. Mitra, and S. Mitra, "Time plane ECG feature extraction using Hilbert transform, variable threshold and slope reversal approach," in *Proc. Int. Conf. Commun. Ind. Appl.*, Dec. 2011, pp. 1–4.
[44] R. N. Costandy, S. M. Gasser, M. S. El-Mahallawy, M. W. Fakhr, and S. Y. Marzouk, "P-wave detection using a fully convolutional neural network in electrocardiogram images," *Appl. Sci.*, vol. 10, no. 3, p. 976, Feb. 2020.

SITI NURMAINI (Member, IEEE) received the master's degree in control system from the Institut Teknologi Bandung (ITB), Indonesia, in 1998, and the Ph.D. degree in computer science from the Universiti Teknologi Malaysia (UTM), in 2011. She is currently a Professor with the Faculty of Computer Science, Universitas Sriwijaya. Her research interests include biomedical engineering, deep learning, machine learning, image processing, control systems, and robotic.

ANNISA DARMAWAHYUNI is currently a Lecturer and a Researcher with the Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Indonesia. Her research interests include biomedical signal processing, deep learning, and machine learning.

ADE IRIANI SAPITRI is currently a Research Assistant with the Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Indonesia. Her research interests include medical imaging, deep learning, and machine learning.

MUHAMMAD NAUFAL RACHMATULLAH is currently a Lecturer and a Researcher with the Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Indonesia. His research interests include medical imaging, biomedical signal processing, deep learning, and machine learning.

FIRDAUS FIRDAUS is currently a Lecturer and a Researcher with the Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Indonesia. His research interests include text processing, deep learning, and machine learning.

JANNES EFFENDI is currently pursuing the bachelor's degree with the Faculty of Computer Science, Universitas Sriwijaya, Indonesia. His research interests include signal processing, deep learning, and machine learning.

BAMBANG TUTUKO is currently a Lecturer and a Researcher with the Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Indonesia. His research interests include robotics, deep learning, and machine learning.