

# Abnormal Electrocardiogram Signal Detection Based on the BiLSTM Network

Husnain Asif<sup>1</sup> and Tae-Young Choe<sup>2,\*</sup>

<sup>1</sup> Dept. of Computer Engineering, Kumoh National Institute of Technology; Master; husnain@kumoh.ac.kr

<sup>2</sup> Dept. of Computer Engineering, Kumoh National Institute of Technology; professor; choety@kumoh.ac.kr

\* Correspondence

<https://doi.org/10.5392/IJoC.2022.18.2.068>

Manuscript Received 29 March 2022; Received 22 June 2022; Accepted 24 June 2022

**Abstract:** The health of the human heart is commonly measured using ECG (Electrocardiography) signals. To identify any anomaly in the human heart, the time-sequence of ECG signals is examined manually by a cardiologist or cardiac electrophysiologist. Lightweight anomaly detection on ECG signals in an embedded system is expected to be popular in the near future, because of the increasing number of heart disease symptoms. Some previous research uses deep learning networks such as LSTM and BiLSTM to detect anomaly signals without any handcrafted feature. Unfortunately, lightweight LSTMs show low precision and heavy LSTMs require heavy computing powers and volumes of labeled dataset for symptom classification. This paper proposes an ECG anomaly detection system based on two level BiLSTM for acceptable precision with lightweight networks, which is lightweight and usable at home. Also, this paper presents a new threshold technique which considers statistics of the current ECG pattern. This paper's proposed model with BiLSTM detects ECG signal anomaly in 0.467 ~ 1.0  $F_1$  score, compared to 0.426 ~ 0.978  $F_1$  score of the similar model with LSTM except one highly noisy dataset.

**Keywords:** Deep Learning; Bidirectional long Short-Term Memory Unit; Signal Processing; Electrocardiogram Signals; Anomaly Detection

## 1. Introduction

The process of recording the electrical signals from human heart during cardiac cycle is known as Electrocardiography (ECG). The ECG signals are obtained by placing the electrodes on human body. The electrodes perceive the minor electrical variations on the human body surface from cardiac muscle depolarization and repolarization within each cardiac cycle [1]. There are two major features of the ECG signals: a) the periodic waveform and b) the recordings of multiple signals from several positions of cardiac muscle corresponding with a cardiac cycle [2]. The standard ECG signal consists of PQRST waveforms known as five morphology segments. A PQRST waveform corresponds to a cardiac cycle which consists of electrical signals. In general, cardiologist or cardiac electrophysiologist manually examines time-sequences of ECG signals in order to identify any abnormality in a human heart.

The detection of abnormalities in the ECG signals is stated as *anomaly detection*. Comparing to symptom classification, anomaly detection requires lighter neural networks than symptom classification does. Thus, anomaly detection technique suits well with portable ECG embedded systems. Since any embedded system does not have enough power nor battery capacity, an edge computing technique should be used together. For the lightweight anomaly detection system with suitable precision, efficiently organized neural network should be used.

Huge volume of ECG signals and various cardiac cycle patterns make anomaly detection a big challenge. The fact that ECG signal patterns and time scales of cardiac cycle vary from person to person makes further difficulty for anomaly detection. ECG signal artifacts are another obstacle against anomaly detection. The artifacts are various types of noise occurs by patient movements, machine vibrations, and so on.

There have been many studies and approaches that classify ECG signals to various cardiac arrhythmia. The researches applied extensive preprocessing techniques in order to derive vectored features and modelled classifiers in order to identify abnormal ECG signals. Traditional anomaly detection techniques uses statistical measures i.e. cumulative sum or exponentially weighted moving average over signals in a fixed time period, which is called *time-step*, to detect changes in the underlying value distribution [3]. The length of time-step generally needs to be pre-determined and the results greatly depend on this parameter. Some researchers have used time series novelty detection technique. However, these techniques require extensive computation to predict new series of ECG signal.

Long short-term memory units (LSTM) network [4] is a type of recurrent neural network (RNN) [5] and it is suitable to remember sequence patterns. Such property enables LSTM to classify sequence patterns and to predict the next patterns based on current patterns. Prediction of LSTM is used for anomaly detection by comparing the predicted patterns and detected signal patterns. One disadvantage of LSTM is one-directional processing. Although ECG signal is time sequential data, a symptom is diagnosed by an interval of ECG signal. Thus, multi-directional processing could detect any anomaly better. BiLSTM is a modification of LSTM such that it processes signals in forwards and backwards directions.

In this paper, we propose an ECG signal anomaly detection system based on BiLSTM. The proposed neural network model predicts the next time-step from the previous time-step. If the Euclidean distance between the predicted next timestep and actual timestep is larger than derived threshold, the proposed system decides the current ECG signal as anomaly. The threshold is carefully designed, since the threshold is an important parameter that decides accuracy or precision of the anomaly detection system.

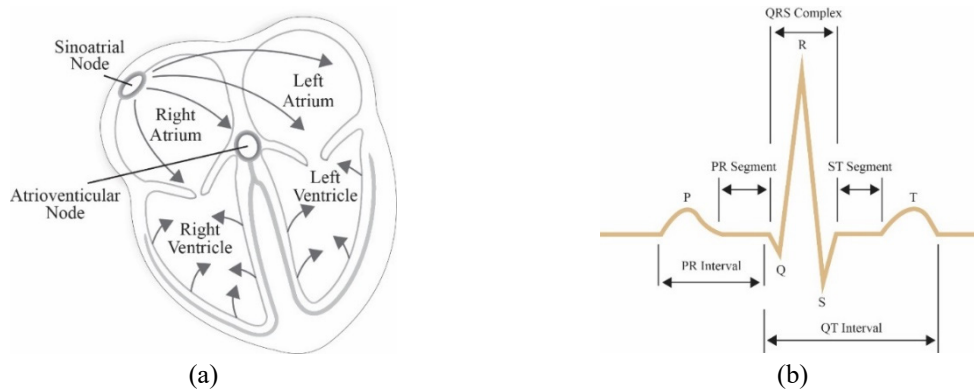
This paper makes two contributions in detecting ECG signal anomaly. First, the proposed anomaly detection system can be applied to any person quickly, since it trains normal ECG signals of a dedicated person within a short time duration and can start detection directly using the trained network. Second, the anomaly detection system can reside in a low-cost system like a personal computer near the patient, since it does not need heavy neural networks.

The rest of this paper is organized as follows. In section 2, basic concepts and the previous works are listed. Section 3 explains the proposed system in detail. In section 4, performance of the system is compared with a well-known previous system. Finally, section 5 concludes the paper with some discussions and future works.

## 2. Concepts and Previous Works

### 2.1 ECG Signals and ECG Leads

The depolarization is a change in cardiac muscle, during which the cardiac muscle experiences a shift in electrical charge, resulting in less amount of negative charge inside cardiac muscle. The repolarization refers to the change in cardiac muscle potential that returns it to a resting potential after the depolarization phase and changed the negative charge to a positive value. The cardiac depolarization happens in the four chambers of the heart, shown in Figure 1 (a). The cardiac depolarization happens in both chamber of atrium first and then both chambers of ventricles. A 1-lead ECG signal example is illustrated in Figure 1 (b). The *P* wave of an ECG signal is generated through contraction caused by depolarization in right and left atrium and depolarization is initiated by sinoatrial node at the wall of right atrium. The *QRS* complex in ECG signal is generated because of contraction in both ventricles after both atrium finish contracting. The atriums contraction is occurred because atrioventricular node received the depolarization wave from sinoatrial node. At the same time the atriums get repolarized and achieve resting potential. The *T* wave in ECG signal is generated when ventricles are repolarized and are obtained resting potential. *PR* interval initiates from the beginning of *P* wave and end before the start of *QRS* complex. *QT* interval initiates from the beginning of *QRS* complex and ends till the finish of *T* wave. The time delay of ventricular and atrial activation, from the end of *P* wave till the beginning of *QRS* complex is known as *PR* segment. The *ST* segment symbolizes period between depolarization and repolarization of ventricles. It connects the *QRS* complex and the *T* wave in ECG signal. This routine continues recurrently in each cardiac cycle, except when there is any cardiac abnormality [6-8]. An ECG signal is usually plotted as a voltage or amplitude in millivolt (mV) with time in second (sec). The 1-lead ECG provides one electrical signal for cardiac muscle activity. As per the number of leads increases, the ECG provides the number of electrical signals for different parts of cardiac muscle.



**Figure 1.** (a) Cardiac chambers; (b) General ECG signal pattern in a cycle

## 2.2 Related Works

Sanchez and Bustos proposed a discord detection algorithm in time sequence based on HOT SAX algorithm [6]. The main aim of their research is to decrease the time complexity of HOT SAX algorithm without presenting parameters configuration. Leng et al. proposed an anomaly detection algorithm aimed to improve detection rate in time sequence data [7]. In the respective research they focused only on clean signals without concerning noise in signals.

Lin C. et al. and Li K. et al. proposed a machine learning algorithm in order to detect anomaly from ECG signals [8,9]. They used unstructured ECG artifacts with uncertain waveform shapes. They considered a situation that the waveform provided by each separate lead for each patient can be different. Hence, it is impractical and difficult to gather the large number of multi-lead ECG signals with ECG artifacts. There is an alternative proposed work based on bit representation clustering, known as Bit Cluster Discord algorithm [10]. This research is aimed to improve the anomaly detection algorithm for ECG signals without needing a trained model. Although, this algorithm requires the length of abnormal ECG signal as an input from users, which is unfeasible for real-time or in the case that abnormal ECG signals are rare.

Many research have been proposed in order to deal with the difficulty of finding optimal input length to detect abnormal ECG signals. The Minimum Description Length (MDL) principle is used to automate the discovery of essential features [11]. Yingchareonthawornchai et al. utilized MDL for anomaly detection in ECG signals [12,13]. Chase C. et al. used brute force discord discovery (BFDD), as known as adaptive window-based discord discovery (AWDD) in order to detect abnormality in ECG signals [14,15]. They used  $R$  peaks of ECG signals in order to extract cycles with variable-lengths.

There are numerous researches which receive the fixed length of input parameter from users [15-18]. However, it is very difficult to identify the appropriate length. Thus, some researches had presented their algorithm with variable lengths [7],[12,13],[18-21]. In the research, the variable-length of results may not be consistent with actual cardiac cycle length. Hence, the result may not correctly cover a cardiac cycle and it could be crucial in diagnosis of abnormalities in ECG signals.

Li et al. used transfer learning to classify unlabeled signals from target users by transferring knowledge from supervised source signals [9]. However, the technique requires hand-coded features and relies on the availability of labeled data for all the different types of abnormalities. However, it is hard to prepare a variety of patients and their different waveforms generated by the different abnormalities for the technique. Polat et al [22] used a least square support vector machine (SVM) to classify normal and abnormal signals, Researchers have also used time series novelty detection techniques [9], [23,24]. However, these require extensive computation to predict on any new series.

## 2.3 Long Short-Term Memory Networks

The Long Short-Term Memory unit (LSTM) network [4] is a descendant of the Recurrent Neural Network (RNN) [5]. It was proposed as a solution to the gradient explosion problem which occurs by largely accumulated backpropagation error gradients during the learning process on the RNN model network. A LSTM network predicts an activity label at each time. There could be multiple ECG signal data combined to predict an activity label. Also, the label can be a multi-dimensional value. The LSTM network is known to be an influential model in the past and shows learning capabilities from sequential data. It can capture long term dependencies and efficiently learn from fluctuating length sequences.

Chauhan et al. proposed an anomaly detection system on ECG data using a LSTM network [25], which is composed of stacked LSTM layers and one output layer. Among various types of layer construction, they selected a two stacked LSTM layer having 20 LSTM units in L1 and L2 layers. Since their network was tested on ECG signals without any artifact signals, they do not guarantee to work in real situation.

#### 2.4 Bi-directional LSTM

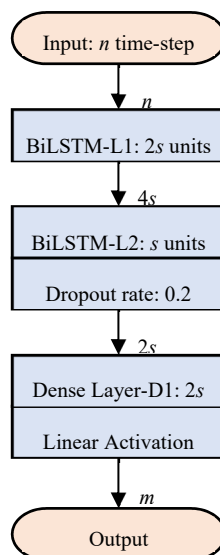
The limitation of the conventional LSTM network is that it uses only the previous sequence, that is one way direction. The Bi-directional Long Short-Term Memory unit (BiLSTM) network processes data in both directions with two separate hidden layers, which are then feed forwards to the same output layer [26], [27]. While a LSTM network only preserves patterns from the past to the current, BiLSTM runs inputs in two opposite ways, one from the past to the current and another from the future to the current. Recently, BiLSTM have been used in a lot of real-world sequence processing problems such as phoneme classification, continuous speech recognition, and speech synthesis. One disadvantage of BiLSTM is that it requires some time delay or more data sequence from the current point because the future data is necessary in order to process the current context or pattern.

BiLSTM has been used in multiple areas including ECG. Zhu et al. proposed a BiLSTM-CNN GAN model in order to generate synthetic EDG data for any patient [28]. Although the research is not directly related with anomaly detection, it shows that BiLSTM is utilized for ECG signal processing. Mostayed et al. proposed a deep neural network system to classify ECG symptom using BiLSTM networks [29]. In order to overcome small number of training data set for classification, 12-leads ECG signals are used.

### 3. Proposed Scheme

#### 3.1 Structure of the Anomaly Detection System

In order to predict anomalies in ECG sequential signals, we propose a deep learning model based on stacked BiLSTM layers as shown in Figure 2. Since stacked BiLSTM reduces input noise but excessive number of stacks distort signal properties, two level stacked BiLSTM is selected [30]. A neural network of the proposed model is trained and memorizes sequence patterns by training a patient's normal ECG signal sequences. Since training phase uses only normal signal sequences, abnormal signal sequence is removed from training dataset. It is explained in Section 3.2.



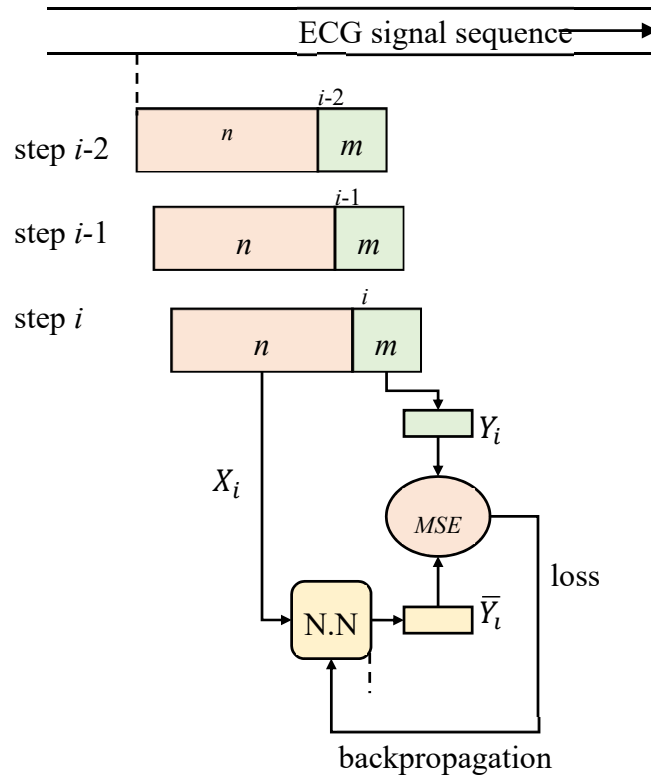
**Figure 2.** The proposed two-layer BiLSTM network

We use the neural network to expect the next some signal sequence of length  $m$  given the current signal sequence of length  $n$ . If the expected signal sequence is quite different with the original signal sequence, we decide the sequence is abnormal. We call  $n$  as time-step and  $m$  as expected time-step. In general, time-step  $n$  is

recommended to be the same or greater than a period of ECG signal in order to cover an ECG cycle. The number of BiLSTM units in each layer depends on the time-step  $n$ .

Given an input signal sequence  $x_k$  ( $k \geq 0$ ), let us call notations training input  $X_i$  or  $(x_{i-n}, x_{i-n+1}, \dots, x_{i-1})$  as an input at step  $i$ ,  $\bar{Y}_i$  or  $(y_i, y_{i+1}, \dots, y_{i+m-1})$  as the output or expectation of the neural network at step  $i$ , and  $Y_i$  as the label at step  $i$  as shown in Figure 4. Then,

**Figure 3.** Training structure of the proposed anomaly detection system



$$Y_i = (x_i, x_{i+1}, \dots, x_{i+m-1}),$$

because the proposed neural network is trained in order to expect the next signal sequence. If the neural network is trained well and ECG signal is normal sequence, the expectation  $\bar{Y}_i$  and the label  $Y_i$  will be almost the same. So, the following  $MSE$  (mean squared error) will be near to zero:

$$MSE_i = \frac{1}{m} \sum_{j=0}^{m-1} (y_{i+j} - x_{i+j})^2 \tag{1}$$

$MSE$  is used as a loss function of the neural network during the training process. The training process is explained in Figure 3. In order to reduce effects of ECG signal artifacts and to cover various ECG signal patterns of patients, input data areas are overlapped. Also, labels are overlapped likewise. By the training process, the neural network is constructed in order to minimize  $MSE$  value.

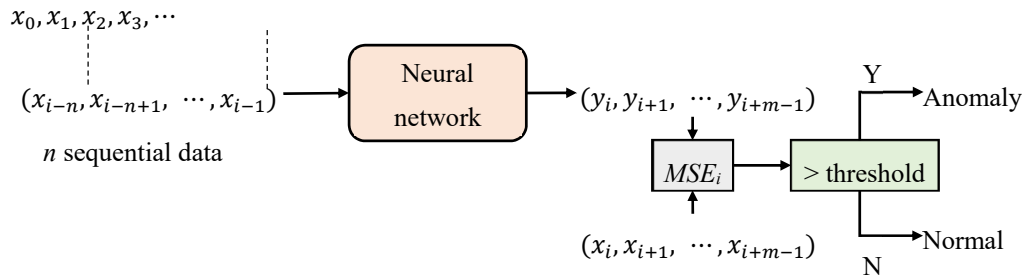
After the training phase, the trained neural network is used to detect anomalies in patients' ECG signal as shown in Figure 4. An anomaly at time  $i$  is decided by comparing the current  $MSE_i$  and predefined threshold value. Deciding the threshold value is described in Section 3.4. If  $MSE_i$  is smaller than the threshold value, the system decides the current value as a normal state. Otherwise, it is considered as an anomaly.

### 3.2 Dataset Pre-processing

In order to train and to validate the anomaly detection system, ECG dataset provided by MIT-BIH Arrhythmia is used [31], [32]. Among the dataset, the following 5 ECG datasets are selected: sel100, sel221, sel223, sel14157, and sel15184. Each dataset has 15 minute two-lead ECG recordings with 250Hz. Thus 450,000 signals are contained in each dataset. Only the single-lead channel is used for experiments.

Table 1 shows normal / anomaly cycles in total dataset and test set. Sel221 and sel223 have lots of abnormal cycles annotated by experts. Sel14157 and sel15814 have relatively small number of abnormal cycles. Sel100 is a challenging dataset. Although sel100 has small number of abnormal cycles, its ECG patterns are not stable, which makes it hard to train the neural networks.

**input signal sequence:**



**Figure 4.** The structure of the anomaly detection system

**Table 1.** The number of cycles in dataset

Sel Name	Total Cycles	No. Normal Cycles	No. Abnormal Cycles	Test set	
				normal	abnormal
Sel100	1134	1115	19	212	8
Sel221	1270	1007	263	199	55
Sel223	1327	967	360	188	74
Sel14157	1081	1055	26	212	3
Sel15814	1036	950	86	183	22

Since the range of values in an input data varies widely especially by artifacts, there is a probability that the objective function does not work properly without normalization. In order to normalize the values, *feature scaling* is applied. Feature scaling is a method used to normalize the range of independent variables or features of data. In data processing, it is also known as data normalization and scales the range of features between [0, 1] or [-1, 1]. We select [-1, 1] as the target range with min-max normalization because ECG signals distribute around zero in millivolts scale with some peaks. The general formula of min-max normalization is given as follows:

$$x_i' = \frac{x_i - \min_k x_k}{\max_k x_k - \min_k x_k} \quad (2)$$

where  $x_i$  is an original value,  $x_i'$  is the normalized value,  $\max_k x_k$  is the maximum value of data signals, and  $\min_k x_k$  is the minimum value of data signals. After the scaling, a dataset is divided into a training set and a test set. In order to make a training set, front 80% of the dataset is selected and anomaly cycles are trimmed off. Remaining 20% of each dataset is used as a test set. 25% of the training set is used for validation. Thus, the proportion of training set, validation set, and test set is 60%, 20%, and 20%, respectively.

For each dataset front 80% signal data is used for training and verification. Since normal data is used for training process, data with abnormal annotations should be trimmed off. An annotation is labeled to each ECG

signal peak which is the point with the largest value in an ECG signal cycle. Point  $R$  in Figure 1 (b) is a peak. Given an ECG signal peak of index  $i$  with an abnormal annotation, we set a range  $[i - \alpha_{i-1} \sim i + \alpha_i]$  as a rough duration of an anomaly ECG cycle where  $\alpha_{i-1}$  is the medium between peak of index  $i-1$  and peak of index  $i$ . If an input data with size  $n$  overlapped with any anomaly ECG cycle, it is extracted from the training set.

### 3.3 BiLSTM Network Configuration

After being pre-processed, a training set is fed to the proposed BiLSTM network. The neural network is composed of two-level stacked BiLSTM layers (L1 and L2) and a dense layer as shown in Figure 2. The input size is decided by time-step parameter  $n$ . The input of one feature along with the previous  $n$  time-steps data is provided to  $2s$  BiLSTM units with  $\tanh$  activation function. After BiLSTM L1 layer, the output shape changed to  $(4s, 1)$  because one BiLSTM unit outputs two values and one feature is used. Output of L1 layer is provided to L2 BiLSTM layer with  $s$  units with  $\tanh$  activation function. The output shape after L2 is changed to  $(2s)$ . The L2-regularization (ridge regression technique) is used as kernel regularizer, recurrent regularizer, and bias regularizer in L2 in order to prevent overfitting of model. The dropout rate of 0.2 is applied after L2. The output of the L2 is provided to dense layer D1. There are  $2s$  neurons used in the dense layer D1 along with linear activation function. The neuron in the dense layer predicts the next  $m$  expected time-step signals based on the previous  $2s$  signals.

### 3.4 Threshold Selection

Threshold value is used to decide normal or abnormal signal by comparing with  $MSE_i$ . The maximum MSE is a frequently used threshold in anomaly detection. Unfortunately, QTDB ECG signal data is quite noisy, which makes the maximum MSE be not feasible as a threshold. In order to decide the threshold, we decided to use three-sigma rule of thumb used in empirical sciences as the base threshold selection rule. That is, a value exceeds three times of standard deviation, the value is quite special. After the neural network is trained by the training set, the network is applied to the training set in order to get mean square error  $MSE_i$  of  $i^{\text{th}}$  data.  $MSE_i$  is calculated from predicted value  $\bar{Y}_i$  generated by the neural network and label or true value  $Y_i$  in each step  $i$ . We assume that  $MSE_i$  has normal distribution.

Let  $\mu$  be the mean and  $\sigma$  be the standard deviation of the distribution. Probability that random variable  $X$  is within the  $3\sigma$ , that is,  $Pr(\mu - 3\sigma \leq X \leq \mu + 3\sigma)$  is about 99.73%. We assume that anomaly is in the range above the upper  $3\sigma$  boundary, that is,  $Pr(X > \mu + 3\sigma)$ , which is about 0.135%. The threshold  $\mu + 3\sigma$  is computed from *percentile()* function in *Numpy* package as follows:

$$\text{Threshold} = \text{numpy.percentile}(\text{MSE\_training\_set}, 99.865) \quad (3)$$

where  $\text{MSE\_training\_set}$  is a list that contains MSE loss of all training data.

## 4. Experimental Results and Performance Analysis

We present detailed experiments with various parameter configurations. The experimental results show effect of parameters on LSTM and BiLSTM.

### 4.1 Dataset

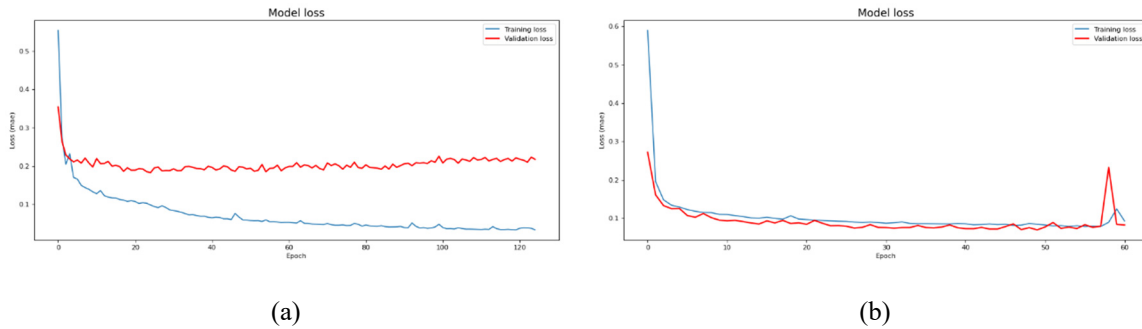
We use the following 5 datasets: sel100, sel221, and sel223 from MIT-BIH Arrhythmia Database, and sel14157 and sel15814 from MIT-BIH Long-Term ECG Database in PhysioBank. PhysioBank contains digital recordings of physiological signals for the purpose of biomedical research. We analyzed recordings of sel100, sel221, sel223, sel14157 and sel15814 till the end.

Table 1 shows major properties of the datasets including the number of normal and abnormal annotations. The annotations are notated by experts.

### 4.2 Hyper-parameters Selection

The proposed BiLSTM network is implemented using Keras library [33] in TensorFlow 2.0 [34] as a backend. There are some hyper-parameters which affect the performance of the system: time-step, expected time-step, batch size, epoch number, number of layers, and number of units per layer. We performed several experiments in order to select optimal and fair hyper-parameters for performance comparison. You can see that

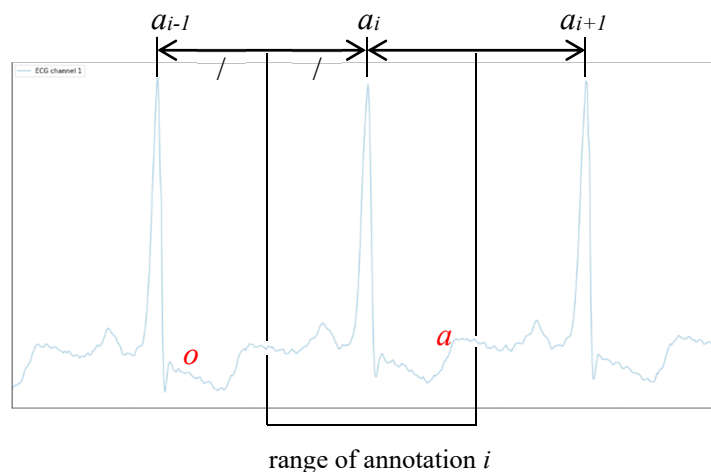
some parameter values are already applied in Figure 2 such that the number of layers as two in the BiLSTM network. The reason of selecting two layers is that deeper network layers tend to blur the characteristics of datasets. Other selected parameters are as follows: time-step 290, batch size 1160, epoch size 500 at max with early stop, which give us the best training and validation accuracy results for the LSTM and BiLSTM networks. The followings minor hyper-parameters are selected. Dropout rate 0.2 is selected only once after the second BiLSTM layer because adding more dropout layers disturbs accuracy and loss trends. Similarly, increasing or decreasing dropout rate also disturbs accuracy and loss trends. The Adam optimizer is used during training with a learning rate of 0.0001, since the smaller value for learning rate showed better results. Because the number of LSTM units per layer and threshold value highly affect performance of the detection system, they are carefully observed in the experiments.



**Figure 5.** Loss curve (a) 2-layered LSTM with expected time-step 60 on sel223 dataset, (L1-290, L2-145), (b) 2-layered BiLSTM with expected time-step 7, (L1-64, L2-32) on sel100 dataset

#### 4.3 The Training and Validation Results of BiLSTM

Figure 5 shows training loss and validation loss curves of two-layered LSTM with expected time-step 60, L1-290 units, L2-145 units on sel223 dataset and of two-layered BiLSTM with expected time-step 7, L1-64 units, and L2-32 units on sel100 dataset. Although the training loss curve of the LSTM model is quite smooth, the validation loss curve does not reduce to a suitable level. BiLSTM with the same configuration on sel223 shows similar loss curve. BiLSTM loss curves in Figure 5 (b) show flicks at the end of epoch, which look like that the BiLSTM structure is not suitable to sel100. In real, sel100 dataset itself is quite unstable and does not allow easy training. LSTM with the same configuration shows similar loss curves without any flick.



**Figure 6.** Examples of detections and its mapping: (a) detection  $a$  is within the range of annotation  $a_i$ , (b) detection  $o$  is not in the of the range of annotation  $a_i$ , but in the range of annotation  $a_{i-1}$ .

#### 4.4 Decision of Detections

In the test phase, the anomaly detection system gets a contiguous sequence of signal data with size (time-step  $n$  + expected time-step  $m$ ) and detects whether the expected signal data contains any anomaly or not. Since



such detections occur in every time step, there are lots of detections. For example, used test dataset has 45,000 data and labels, so there are 45,000 detections. But sell100 has 220 cycles in the test set as shown in

Table 1. Then there are 204.5 detection behaviors in a cycle in average. Anomaly detections do not require high density in general and an anomaly occurs rarely. Thus, we decide to notate an anomaly detection once at maximum in an ECG cycle. Let us explain the relation between the ECG cycle and range of an annotation using Figure 6. Let  $i$  be a peak index and it has annotation  $a_i$ . Range of annotation  $a_i$  starts from the middle of the previous peak and the current peak  $i$  and ends by the middle of the current peak  $i$  and the next peak. Thus, range of annotation  $a_i$  is defined as follows, where  $index(a_i)$  is a function that returns time step of annotation  $a_i$ :

$$\left[ \frac{index(a_{i-1}) + index(a_i)}{2}, \frac{index(a_i) + index(a_{i+1})}{2} \right) \quad (4)$$

For example, if  $index(a_{i-1}) = 2300$ ,  $index(a_i) = 2520$ , and  $index(a_{i+1}) = 2740$ , then the range of annotation  $a_i$  is  $[2410, 2630)$ . If one or more anomalies are detected in the range, the detection system announces that there is an anomaly in the range. Otherwise, it is announced as normal. If point  $a$  in Figure 6 is detected as an anomaly, the range of annotation  $a_i$  is decided as anomaly. After that, another anomaly detection in the same range does not change the decision. Since point  $o$  in Figure 6 locates within the range of annotation  $a_{i-1}$ , it is related with anomaly of annotation  $a_{i-1}$  not of annotation  $a_i$ .

#### 4.5 Performance measure

Since the anomaly detection is a binary decision, confusion matrix is used to measure the performance. The followings are the outcomes by actual classification (annotation) and predicted classification (detection):

- TP: true positive, predicted as anomaly on abnormal signal
- FP: false positive, predicted as anomaly on normal signal
- TN: true negative, predicted as normal on normal signal
- FN: false negative, predicted as normal on abnormal signal

If one or more anomalies are detected by the system in a range of anomaly annotation, it is counted as TP. If one or more anomalies are detected in a range of normal annotation, it is counted as FP. If any anomaly is not detected in a range of normal annotation, it is counted as TN. Otherwise, it is counted as FN. If the proportion of normal annotations and anomaly annotations is the similar, the following *accuracy* is enough to be used as a measure of a system detection performance.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (5)$$

In the ECG signal, most annotations are normal. As the result, although a detection system ignores all anomalies and announces them as normal, accuracy is high because TN is the large part in both a numerator and a denominator. In order to fix such misguided situation,  $F_1$  score is defined as follows:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F_1 \text{ score} = \frac{2 * (Recall \times Precision)}{Recall + Precision} \quad (6)$$

By excluding TN,  $F_1$  score concentrates on the proportion of TP over FP and FN.

#### 4.6 Performance comparison between LSTM and BiLSTM

By changing multiple parameters that effect performance of the anomaly detection system, we found that it is hard to find single optimal parameter configuration that covers the selected dataset. Parameters that cannot be fixed are the number of units in each layer and the expected time-step  $m$ . The greater number of units each layer has, the longer sequence patterns can be memorized. Thus stacked LSTM / BiLSTM with L1-64 layer and L2-32 layer (L1-64, L2-32) has expected time-step (expSize) 7, while stacked LSTM / BiLSTM with L1-290 layer and L2-145 layer (L1-290, L2-145) has expected time-step 60. Configuration (L1-64, L2-32) with time-step 7 represents a lightweight neural network can be executed in a low-cost personal computer, while configuration (L1-290, L2-145) with time-step 60 does a middleweight neural network suitable for a general personal computer.

Table 2 and Table 3. Test result of network model based on BiLSTM (L1-64, L2-32, expSize=7) show experimental results of stacked LSTM and stacked BiLSTM in the lightweight configuration. Table 2 and Table 3 show experimental results in the middleweight configuration. Bold numbers in each dataset are the best  $F_1$  score among various parameters. The best scores are distributed diversely over configurations, so we cannot easily decide the best parameter configuration. LSTM has the three best cases, so does BiLSTM.

If we consider  $F_1$  score itself, sel100 shows too small  $F_1$  score for all configurations. Sel100 has high FP values and low TP values, which means excessive noise obscures true detection. In the case of sel221, the middleweight configuration works better than the lightweight configuration. In the middleweight configuration of sel221, threshold is set to  $\mu + 4\sigma$  in order to distinguish high MSE valued noise from true anomalies. BiLSTM with the lightweight configuration shows the best  $F_1$  score in the case of sel223 if threshold  $\mu + 4\sigma$  is used. In the case of sel14157, middleweight configuration shows better performance, where LSTM and BiLSTM yield the same results. Sel14157 has only 3 anomalies in the test set. Thus, small number of wrong detections makes large score down. In the case of sel15814, middleweight configuration shows better performance. Especially, BiLSTM with middleweight configuration perfectly detects anomalies. Thus, BiLSTM shows better performance for almost stable ECG signals like Sel14157 and Sel15814 when a middleweight computer is available. BiLSTM would suit to portable ECG monitoring machine, because a patient with highly light symptom of ECG signal prefers his/her home with Wi-Fi home network and a computer system rather than a hospital.

Threshold is set to  $\mu + 3\sigma$  initially. But some datasets in some configurations prefer different values like  $\mu + 3.5\sigma$  or  $\mu + 4\sigma$ . Sel14157 shows better performance when  $\mu + 3.5\sigma$  or  $\mu + 4\sigma$  is used. MSEs of such datasets are highly dispersed and  $\mu + 3\sigma$  is not enough threshold. Thus, deciding good threshold needs more works.

**Table 2.** Test result of network model based on LSTM (L1-64, L2-32, expSize=7)

	Sel100	Sel221	Sel223	Sel14157	Sel15814
$\sigma$	3	3	3	3.5	3
TN	168	121	130	210	167
TP	5	36	73	3	21
FP	44	78	58	2	16
FN	3	19	1	0	1
Accuracy	0.786	0.618	0.775	0.991	0.917
Precision	0.102	0.316	0.557	0.600	0.568
Recall	0.625	0.655	0.986	1.000	0.955
$F_1$ score	<b>0.175</b>	0.426	0.712	0.750	0.712

**Table 3.** Test result of network model based on BiLSTM (L1-64, L2-32, expSize=7)

	Sel100	Sel221	Sel223	Sel14157	Sel15814
$\sigma$	3	3	4	3.5	3
TN	122	131	172	210	183
TP	7	37	69	2	10

FP	90	68	16	2	0
FN	1	18	5	1	12
Accuracy	0.586	0.661	0.920	0.986	0.941
Precision	0.072	0.352	0.812	0.500	1.000
Recall	0.875	0.673	0.932	0.667	0.455
F <sub>1</sub> score	0.133	0.467	<b>0.868</b>	0.571	0.625

**Table 4.** Test result of network model based on LSTM (L1-290, L2-145, expsize=60)

	Sel100	Sel221	Sel223	Sel14157	Sel15814
$\sigma$	3	4	3	3.5	3
TN	189	199	123	210	182
TP	0	43	51	3	22
FP	23	1	66	1	1
FN	8	12	22	0	0
Accuracy	0.859	0.949	0.664	0.995	0.995
Precision	0.000	0.977	0.436	0.750	0.957
Recall	0.000	0.782	0.699	1.000	1.000
F <sub>1</sub> score	None	<b>0.869</b>	0.537	<b>0.857</b>	0.978

**Table 3.** Test result of network model based on BiLSTM (L1-290, L2-145, expSize=60)

	Sel100	Sel221	Sel223	Sel14157	Sel15814
$\sigma$	2.5	4	3	4	3
TN	132	193	121	210	183
TP	3	46	52	3	22
FP	80	7	68	1	0
FN	5	9	21	0	0
Accuracy	0.614	0.937	0.660	0.995	1.000
Precision	0.036	0.868	0.433	0.750	1.000
Recall	0.375	0.836	0.712	1.000	1.000
F <sub>1</sub> score	0.066	0.852	0.539	<b>0.857</b>	<b>1.000</b>

## 5. Discussion and Conclusions

Electrocardiography (ECG) signals are extensively used to measure the condition of the heart, and the subsequent time series signal is frequently examined manually by a medical expert to identify any arrhythmia that the patient may have suffered. There has been a lot of work done in order to systematize the procedure of examining ECG signals, but most of the research involves extensive preprocessing. In this paper, we utilize a Bi-directional Long Short-Term Memory (BiLSTM) architecture to advance a model for observing anomalies in ECG signals. We further utilize the time-step and the expected time-step from the recurrent model to indicate normal or abnormal behavior. We propose an anomaly detection in ECG time signals using two-layer BiLSTM. Experimental results show that BiLSTM networks works better for patients who has almost stable ECG signals and a middleweight computer for detection.

The main contribution of the paper is the proposal of a suitable threshold for anomaly detection in ECG signal. Instead of the maximum MSE, MSE values above  $\mu + 3\sigma$  are decided as anomaly. The proposed detection system can be applied to any person quickly, and the system can be maintained in small system like a general Personal Computer. Another fact that we derived from the experiments is that each QTDB ECG signal has quite peculiar characteristics compared to others. Some signal datasets have excessive noise. Another signal dataset has clear signals and easily reveals anomaly. Thus, we need to investigate characteristics of ECG signal patterns in detail.

Although detecting anomaly in QTDB ECG signals is a challenging area, our research can be expanded to improve the performance. One extension is to classify the QTDB dataset according to properties of dataset like types of cardiac arrhythmia. Another extension is to use two lead ECG recordings. Since each lead recording has different patterns, it is not easy to apply two lead recording as two-dimensional datasets. If each

lead is applied to each LSTM network and results are combined using ensemble technique, the neural network would show another type of results.

**Acknowledgments:** This research was supported by Kumoh National Institute of Technology (202001250001)

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- [1] Lilly, Leonard S, *Pathophysiology of Heart Disease: A Collaborative Project of Medical Students and Faculty*, 6th ed. Lippincott Williams & Wilkins, 2016. [Online]. Available: <https://doi.org/10.1097/01823246-199506030-00013>
- [2] Yama Y., Ueno A., and Uchikawa Y., "Development of a wireless capacitive sensor for ambulatory ECG monitoring over clothes," *Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS '07)*, pp. 5727–5730, 2007.
- [3] M. Basseville and I. V. Nikiforov, *Detection of abrupt changes: theory and application*, vol. 104. prentice Hall Englewood Cliffs, 1993. [Online]. Available: [https://doi.org/10.1016/0005-1098\(96\)82332-6](https://doi.org/10.1016/0005-1098(96)82332-6)
- [4] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: <http://dx.doi.org/10.1162/neco.1997.9.8.1735>.
- [5] L. Medsker and L. C. Jain, *Recurrent neural networks: design and applications*. CRC press, 1999.
- [6] H. Sanchez and B. Bustos, "Anomaly detection in streaming time series based on bounding boxes," presented at the International Conference on Similarity Search and Applications, Springer, 2014. [Online]. Available: [http://dx.doi.org/10.1007/978-3-319-11988-5\\_19](http://dx.doi.org/10.1007/978-3-319-11988-5_19)
- [7] M. Leng, W. Yu, S. Wu, and H. Hu, "Anomaly detection algorithm based on pattern density in time series," in *Emerging Technologies for Information Systems, Computing, and Management*, Springer, 2013, pp. 305--311. [Online]. Available: [https://doi.org/10.1007/978-1-4614-7010-6\\_35](https://doi.org/10.1007/978-1-4614-7010-6_35)
- [8] C. C. Lin and C. M. Yang, "Heartbeat Classification Using Normalized RR Intervals and Wavelet Features," presented at the 2014 international symposium on computer, consumer and control, IEEE, 2014. [Online]. Available: <https://doi.org/10.1109/is3c.2014.175>
- [9] K. Li, N. Du, and A. Zhang, "Detecting ECG abnormalities via transductive transfer learning," presented at the ACM Conference on Bioinformatics, Computational Biology and Biomedicine, 2012. [Online]. Available: <https://doi.org/10.1145/2382936.2382963>
- [10] Y.-H. Noh, G.-H. Hwang, and D.-U. Jeong, "Implementation of real-time abnormal ECG detection algorithm for wearable healthcare," presented at the 6th International Conference on Computer Sciences and Convergence Information Technology (ICCIT), IEEE, 2011.
- [11] P. Grünwald, *The minimum description length principle*, MIT press, 2007. [Online]. Available: <https://doi.org/10.7551/mitpress/1114.003.0004>
- [12] S. Yingchareonthawornchai, H. Sivaraks, T. Rakthanmanon, and C. A. Ratanamahatana, "Efficient proper length time series motif discovery," presented at the IEEE 13th International Conference on Data Mining, IEEE, 2013. [Online]. Available: <https://doi.org/10.1109/icdm.2013.111>
- [13] B. Hu, T. Rakthanmanon, Y. Hao, S. Evans, S. Lonardi, and E. Keogh, "Using the minimum description length to discover the intrinsic cardinality and dimensionality of time series," *Data Min. Knowl. Discov.*, vol. 29, no. 2, pp. 358–399, 2015, doi: <http://dx.doi.org/10.1007/s10618-014-0345-2>.
- [14] C. Chase and W. J. Brady, "Artifactual electrocardiographic change mimicking clinical abnormality on the ECG," *Am. J. Emerg. Med.*, vol. 18, no. 3, pp. 312–316, 2000, doi: [http://dx.doi.org/10.1016/S0735-6757\(00\)90126-8](http://dx.doi.org/10.1016/S0735-6757(00)90126-8).
- [15] E. Keigh, J. Lin, and A. Fu, "HOT SAX: efficiently finding the most unusual time series subsequence," presented at the Fifth IEEE International Conference on Data Mining, IEEE, 2005. [Online]. Available: <https://doi.org/10.1109/icdm.2005.79>
- [16] E. Keogh, J. Lin, S.-H. Lee, and H. Van Herle, "Finding the most unusual time series subsequence: algorithms and applications," *Knowl. Inf. Syst.*, vol. 11, no. 1, pp. 1–27, 2007, doi: <http://dx.doi.org/10.1007/s10115-006-0034-6>.
- [17] G. Li, O. Braysy, L. Jiang, Z. Wu, and Y. Wang, "Finding time series discord based on bit representation clustering," *Knowl Based Syst.*, vol. 54, pp. 243–254, 2013, doi: <http://dx.doi.org/10.1016/j.knsys.2013.09.015>.
- [18] K. Buza, A. Nanopoulos, L. Schmidt-Thieme, and J. Koller, "Fast classification of electrocardiograph signals via instance selection," presented at the IEEE First International Conference on Healthcare

- Informatics, Imaging and Systems Biology, 2011. [Online]. Available: <https://doi.org/10.1109/hisb.2011.26>
- [19] M. C. Chuah and F. Fu, "ECG anomaly detection via time series analysis," presented at the International Symposium on Parallel and Distributed Processing and Applications, 2007. [Online]. Available: [https://doi.org/10.1007/978-3-540-74767-3\\_14](https://doi.org/10.1007/978-3-540-74767-3_14)
- [20] B. Raghavendra, D. Bera, A. S. Bopardikar, and R. Narayanan, "Cardiac arrhythmia detection using dynamic time warping of ECG beats in e-healthcare systems," 2011. [Online]. Available: <https://doi.org/10.1109/wowmom.2011.5986196>
- [21] G. Zhang, W. Kinsner, and B. Huang, "Electrocardiogram data mining based on frame classification by dynamic time warping matching," *Comput. Methods Biomech. Biomed. Engin.*, vol. 12, no. 6, pp. 701–707, 2009, doi: <http://dx.doi.org/10.1080/10255840902882158>.
- [22] K. Polat and S. Gunes, "Detection of ECG arrhythmia using a differential expert system approach based on principal component analysis and least square support vector machine," *Appl. Math. Comput.*, vol. 186, no. 1, pp. 898–906, 2007, doi: <http://dx.doi.org/10.1016/j.amc.2006.08.020>.
- [23] H. M. Rai, A. Trivedi, and S. Shukla, "ECG signal processing for abnormalities detection using multi-resolution wavelet transform and Artificial Neural Network classifier," *Measurement*, vol. 46, no. 9, pp. 3238–3246, 2013, doi: <http://dx.doi.org/10.1016/j.measurement.2013.05.021>.
- [24] A. P. Lemos, C. Tierra-Criollo, and W. Caminhas, "ECG anomalies identification using a time series novelty detection technique," presented at the IV Latin American Congress on Biomedical Engineering 2007, Bioengineering Solutions for Latin America Health, 2007. [Online]. Available: [https://doi.org/10.1007/978-3-540-74471-9\\_16](https://doi.org/10.1007/978-3-540-74471-9_16)
- [25] S. Chauhan and L. Vig, "Anomaly detection in ECG time signals via deep long short-term memory networks," presented at the *IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, 2015. [Online]. Available: <http://dx.doi.org/10.1109/DSAA.2015.7344872>
- [26] Savelie Cornegruta, R. Bakewell, S. Withey, and G. Montana, "Modelling Radiological Language with Bidirectional Long Short-Term Memory Networks," arXiv:1609.08409v1, Sep. 2016. [Online]. Available: <https://doi.org/10.18653/v1/w16-6103>
- [27] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Netw.*, vol. 18, no. 5, pp. 602–610, 2005, doi: <http://dx.doi.org/10.1016/j.neunet.2005.06.042>.
- [28] 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, pp. 1–11, 2019, doi: <https://doi.org/10.1038/s41598-019-42516-z>.
- [29] A. Mostayed, J. Luo, X. Shu, and W. Wee, "Classification of 12-lead ECG signals with bi-directional LSTM network," arXiv preprint, arXiv:1811.02090, 2018.
- [30] R. Azzam, Y. Alkendi, T. Taha, S. Huang, and Y. Zweiri, "A stacked LSTM-based approach for reducing semantic pose estimation error," *IEEE Trans. Instrum. Meas.*, no. 70, pp. 1–14, 2020, doi: <http://dx.doi.org/10.1109/TIM.2020.3031156>.
- [31] G. Moody and R. Mark, "MIT-BIH Arrhythmia Database," *MIT-BIH Arrhythmia Database*, Feb. 24, 2005. <https://physionet.org/content/mitdb/1.0.0/> (accessed Jan. 10, 2022).
- [32] G. Moody and R. Mark, "The impact of the MIT-BIH Arrhythmia Database," *IEEE Eng Med Biol*, vol. 20, no. 3, pp. 45–50, doi: <http://dx.doi.org/10.1109/51.932724>.
- [33] F. Chollet, "Keras: Deep Learning for humans," *Keras: Deep Learning for humans*. <https://github.com/keras-team/keras> (accessed Aug. 18, 2021).
- [34] Google Brain Team, "TensorFlow," *TensorFlow*. <https://github.com/tensorflow/tensorflow> (accessed Aug. 18, 2021).



© 2022 by the authors. Copyrights of all published papers are owned by the IJOC. They also follow the Creative Commons Attribution License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.