

An Attention Based Hierarchical LSTM Architecture for ECG Biometric System

This paper was downloaded from TechRxiv (<https://www.techrxiv.org>).

LICENSE

CC BY 4.0

SUBMISSION DATE / POSTED DATE

16-06-2021 / 18-06-2021

CITATION

Jyotishi, Debasish; Dandapat, Samarendra (2021): An Attention Based Hierarchical LSTM Architecture for ECG Biometric System. TechRxiv. Preprint. <https://doi.org/10.36227/techrxiv.14790786.v1>

DOI

[10.36227/techrxiv.14790786.v1](https://doi.org/10.36227/techrxiv.14790786.v1)

An Attention Based Hierarchical LSTM Architecture for ECG Biometric System

Debasish Jyotishi and Samarendra Dandapat

Abstract—The electrocardiogram (ECG) based biometric system has recently gained popularity. Easy signal acquisition and robustness against falsification are the major advantages of the ECG based biometric system. This biometric system can help automate the subject identification and authentication aspect of personalised healthcare services. In this paper, we have designed a novel attention based hierarchical long short-term memory (LSTM) model to learn the biometric representation corresponding to a person. The hierarchical LSTM model proposed in this paper can learn the temporal variation of the ECG signal in different abstractions. This addresses the long term dependency issue of the LSTM network in our application. The attention mechanism of the model learns to capture the ECG complexes that have more biometric information corresponding to each person. These ECG complexes are given more weight to learn better biometric representation. The proposed system is less complex and more efficient as it does not require the detection of any fiducial points. We have evaluated the proposed model for both the person verification and identification problems using two on-the-person ECG databases and two off-the-person ECG databases. The proposed framework is found to perform better than the existing fiducial and non-fiducial point based methods.

Index Terms—Biometrics, Electrocardiogram, off-the-person, Hierarchical LSTM, Temporal variation, Authentication

I. INTRODUCTION

Recently the need for information security has increased due to the significant progress in the field of information acquisition and processing. This has led to the exploration of different biometric modalities, e.g., iris, gait, fingerprint, etc. Depending upon the type of applications and requirements, one out of these available modalities becomes more suitable. Recently, ECG based biometric systems are gaining researchers' attention. The ECG based biometric systems can be used in a wide range of applications, e.g., mobile devices [1], healthcare services [2], and continuous authentication [3], etc. The increased use of telehealth monitoring has increased the requirement of proper security and privacy management of sensitive medical data. The personalised medicine system also requires person authentication [4]. Since the ECG signal is less prone to deformation than other biometric modalities in a hospital environment, it can be used as a practical choice for information security in healthcare services. The ECG based biometric system can also be extended to the personalised automated cardiovascular disease diagnosis systems. Ince *et al.* [5] have shown that a person specific cardiovascular disease diagnosis model performs better than a global model. This

requires the identification of the patient. In this case, an ECG based biometric system can fulfil two purposes, i.e., maintaining the security and privacy of the patient data [2] and improving the system's diagnostic performance.

The practical application of a biometric system has certain challenges, i.e., acceptability, circumvention, and performance [6]. An easy ECG acquisition system is necessary for the biometric system to be widely accepted. The ECG signal can be acquired through *off-the-person*, *on-the-person*, or *in-the-person* recording setup [7]. The *off-the-person* recording setup provides the simplest and the easiest way of acquiring the ECG signal. The quality of the ECG signal recorded in this setup is close to the ECG data acquired during telehealth monitoring. The *on-the-person* ECG recording setup requires conductive gel to be applied between the electrode and skin. This type of recording setup is generally used in the medical scenario. Although the *off-the-person* recording setup is the most suitable for ECG based biometric, the quality of the data is a great challenge. Thus designing a robust biometric system in the *off-the-person* recording scenario has become a necessity. Another requirement for a biometric system to be widely accepted is less time for enrolment as well as verification. However, most of the works available in the literature require 30 seconds to more than a minute of ECG data for enrolment [8], [9].

The ECG based biometric systems available in the literature can be broadly categorised into two categories, i.e., fiducial point based methods, and non-fiducial point based methods. Fiducial point based methods detect the R-peak or extract multiple fiducial points in the preprocessing stage. The methods which don't require the detection of any fiducial points are categorised as non-fiducial point based methods. In fiducial point based methods, the biometric system's efficiency depends entirely on the accurate detection of fiducial points. This becomes very challenging in the *off-the-person* recording scenario [8]. In [3] and [10], the wrongly detected R-peaks have been identified and discarded before further processing. They have also implemented an outlier removal algorithm that removes ECG beats that are not similar to the majority of ECG beats belonging to the respective person. However, these outlier removal algorithms cannot be implemented in a practical scenario. The non-incorporation of these algorithms in the practical scenario will severely degrade the performance of the biometric system. This has motivated us to develop a non-fiducial point based biometric system that can give a robust performance in *off-the-person* and *on-the-person* recording scenarios.

ECG signal is a cyclostationary signal with repetitive ECG

The authors are with the Department of Electronics and Electrical Engineering, Indian Institute of Technology Guwahati, Assam 781039, India. E-mail: {debasish.jyotishi07,samaren} @iitg.ac.in

beats. Each ECG beat comprises three primary complexes, i.e., P, QRS, and T. The temporal variation of an ECG signal is basically of two types; intra-beat variation and inter-beat variation. The relative variation of the three complexes represents the intra-beat variation, and the variation from beat to beat represents the inter-beat variation. Along with the specific morphology of the ECG complexes, these temporal variations can be utilised in a biometric system. In this work, we have designed an attention based hierarchical LSTM model to utilise these information for person verification and identification. The contributions made in the work has been summarised below.

- 1) A novel hierarchical LSTM model has been designed which can effectively model the intra-beat and inter-beat variation of the ECG signal for person identification and verification. This architecture resolves the difficulty in modelling the long term dependencies present in the ECG signal.
- 2) Chan et al. [11] have shown that different ECG complexes have different contributions to the identification process. So, an attention mechanism has been introduced to the hierarchical LSTM model, which learns to identify the ECG complex with more biometric information corresponding to each subject. These ECG complexes are given more weight to learn a better biometric representation. We observed that the biometric information present in an ECG complex varies from person to person.
- 3) The proposed biometric method can perform better with less amount of enrolment data, and it takes only 2 seconds of ECG data for the evaluation.

The rest of the paper is organised in the following manner. We have briefly reviewed the related works in section II. The proposed biometric system has been described in section III. The description of the experimental setup is given in section IV. Results obtained from the evaluation of the proposed model and a discussion on it have been presented in section V. Finally, a brief conclusion is made in section VI.

II. RELATED WORKS

Many research works have been reported in the literature using ECG as a biometric. A detailed review of the existing ECG based biometric systems has been presented in [8] and [9]. Pinto *et al.* have categorised the existing works based on the procedure used for ECG data acquisition, preprocessing methodology employed, and the feature extraction and classification methods [8]. Most of the works in literature have used *on-the-person* ECG databases in their work. A few works have used *off-the-person* ECG databases in their work. The low quality data recorded in the *off-the-person* scenario degrades the performance of the biometric systems [8]. This has motivated researchers to develop robust biometric systems using *off-the-person* ECG data. Arteaga *et al.* [1] have proposed an algorithm for an ECG based mobile authentication system. They have extracted biometric features based on seven fiducial points identified from the ECG signal, which are recorded using an ECG sensor attached to the mobile phone. They have shown an improved performance with only 30 seconds

of enrolment data and 4 seconds of data for authentication. A wavelet distance based person identification system has been proposed by Chan *et al.* [11]. They have recorded ECG data from 50 persons in an *off-the-person* recording setup. In their work, they have shown that the QRS complex of the ECG signal is primary for distinctiveness, followed by P wave and T wave. In [3], Louis *et al.* have proposed an ECG based continuous authentication system using the UofTDB database, which is recorded in an *off-the-person* recording setting. They have achieved an equal error rate of 7.89% for the same session analysis and 10.10% for across session analysis. Silva *et al.* have proposed a convolutional neural network (CNN) based authentication system which performs better in publicly available *off-the-person* ECG databases [10].

The majority of the works reported in the literature employ the fiducial point based method. A few works have been reported based on non-fiducial point based methods. Wang et al. have proposed a non-fiducial point based sparse representation methodology for human identification [12]. They have shown an identification accuracy of 99.48% on 100 subjects taken from the PTB database. Wang *et al.* [13] have used discrete cosine transform (DCT) coefficients of the ECG signal's autocorrelation values. A CNN based architecture that utilises the multiresolution components of the autocorrelation signal of the ECG signal has been proposed by Zhang *et al.* [14]. Recently, Abdeldayem and Bourlai have proposed a CNN based architecture for person identification [15]. This architecture utilises the spectral correlation of the ECG signal. Their method has achieved an identification accuracy of 94.9% when evaluated using data from 290 subjects of the PTB dataset. However, we have not come across any non-fiducial point based methods using an *off-the-person* ECG database.

The works in the literature have explored various signal processing and machine learning methods for biometric feature extraction and classification, respectively. Chan *et al.* [11] have developed a wavelet based distance measure to identify the candidate subject by computing the similarity between ECG templates. In [16], a set of multiscale autoregressive features have been extracted for person identification and verification. Irvine et al. [17] have developed a principal component analysis (PCA) based method for biometric feature extraction. Some works have used sparse representation based methods for feature extraction and support vector machine (SVM) for the identification [12], [18]. A novel hybrid GMM-HMM based generative model has been developed by Lim *et al.* [19] for person identification. However, the method requires multiple fiducial point detection, which limits its applicability in the practical scenario. Recently, some deep learning based methods have been proposed for person identification and verification. Most of the works in literature have developed CNN based architectures [10], [14], [15], [20]. Salloum and Kuo have proposed a biometric system using a vanilla LSTM based architecture [21]. However, their method needs R-peak detection and beat segmentation. Except for a few works [3], [19], [21], most of the existing methods have not utilised the temporal variation of the time varying ECG signal.



Fig. 1. Block diagram of the proposed ECG based biometric system

III. PROPOSED METHOD

The proposed method primarily has three modules, i.e., (i) Preprocessing module, (ii) Segmentation module, and (iii) Identification/Verification module. Fig.1 shows the block diagram of the proposed method. First, the ECG signal is filtered and then normalised in the preprocessing module. The filtered ECG signal is then fed to the segmentation module. In the segmentation module, the training and testing templates are extracted from the filtered ECG signal. These extracted templates are then fed to the proposed attention based hierarchical LSTM model for learning and decision making. The proposed neural network model can operate both in identification mode as well as in verification mode.

A. Preprocessing

The noise present in the ECG signal varies depending on the type of recording procedure. The ECG signal recorded in an *on-the-person* recording setup has high SNR. But, the ECG data recorded in an *off-the-person* recording setup contains a significant quantity of high frequency noise and abrupt changes. So, we resampled the ECG signal to a sampling frequency of 200 Hz and then passed it through a band pass filter with a lower cutoff frequency of 0.5 Hz and a higher cutoff frequency of 40 Hz. This filters the baseline wander, power line interference and high frequency noises present in the ECG signal. Then the ECG signal is further filtered using the wavelet based method proposed by Sharma et al. [22] to remove the white gaussian noises and muscle artefacts. Fig. 2(a) shows two raw ECG samples, and Fig 2(b) shows the respective ECG signal after band pass filtering and wavelet based noise removal. From the figure, it can be observed that the sudden changes introduced primarily due to the loose contact between the electrode and the skin cannot be removed through the low pass filtering. These noises are very frequent in the case of the *off-the-person* recording setup. So, we have employed the baseline wander removal method proposed in [23] to remove the sudden changes in the baseline. Fig 2(c) shows the respective ECG signal after removal of sudden changes. The peak-to-peak amplitude and the mean of the filtered ECG signal are then normalised to one and zero, respectively.

B. Segmentation

The segmentation module is the basic building block of the proposed model. In this module, the training and testing sequences are extracted using a rectangular window. The ECG sequences are extracted without any alignment with the R-peaks. This makes our model completely free of fiducial point detection. The ECG sequences are extracted using a 2s rectangular window, and then the z score standardisation

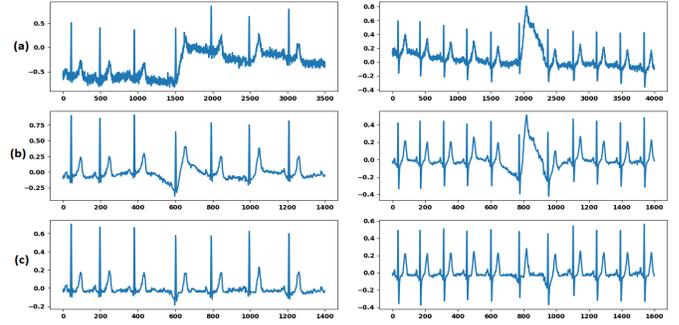


Fig. 2. (a) Raw ECG signal of two different persons. (b) ECG signal after band-pass filtering and high frequency noise removal. It can be noticed that the noise due to sudden changes has not been filtered out. (c) ECG signal after removal of baseline drifts due to sudden changes.

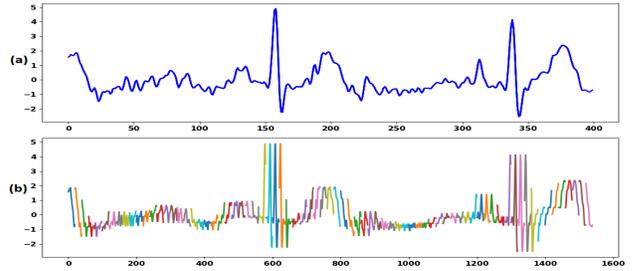


Fig. 3. (a) One complete ECG sequence of duration 2 s. (b) The ECG segments of length 0.1 s were obtained after applying the windowing technique to the ECG sequence. The ECG segments have been plotted sequentially.

is done. The training sequences are extracted by shifting the rectangular window with an overlap of 0.75 fractions. The testing sequences are extracted without any overlap of the window. This helps us extract more training templates with less amount of data, as well as in regularisation of the model through data augmentation. In this work, we have extracted 22 training templates, which amounts to 12.5 seconds of training data, and 6 testing templates, which amounts to 12 seconds of testing data. Fig. 3 shows an ECG sequence extracted from the filtered ECG signal. The ECG sequence extracted from the raw ECG signal is further segmented into smaller segments, which are given as input to the LSTM model at different time steps. In [24], we have shown that the smaller ECG segments allow an LSTM model to utilise the temporal variation present in the ECG signal fully. A rectangular window of length 100 ms and a shift with an overlap of 0.75 fractions of window length is applied to generate the ECG segments. Fig. 3 shows the ECG segments of an ECG sequence, aligned sequentially. The smaller ECG segments extracted with a shift of 75% overlap are essential to capture the intra-beat variation present in the ECG signal [24].

C. Proposed LSTM Model

We have designed a neural network model using LSTM cells to exploit the temporal variation present in the ECG signal. LSTM is one of the many variants of the recurrent neural network (RNN) proposed in the literature. A recurrent neural network is the recurrent connection of a neuron cell, which

takes past information as well as current information as its input. Thus the RNN cell utilises both the past information and the present data to make a decision. A recurrent connection of these cells utilise the temporal information present in a time-varying signal [25]. However, the basic RNN cells suffer from exploding gradient and vanishing gradient problem. This does not allow the model to learn long-term dependencies. This motivated researchers to develop different variants of RNN cells that can account for the long term dependencies, e.g., Skip-RNN [26], LSTM [25], GRU [25], Stochastic RNN [27], etc. Out of these, LSTM is widely used to learn representation because of its better performance compared to other variants [25]. The fundamental idea behind the LSTM cell is the multiple nonlinear gating units, which controls the flow of the information into as well as out of the LSTM cells, and a memory unit that retains learned representation over a period of time.

The architecture of an LSTM cell is shown in Fig. 4. The input to the LSTM cell at a time stamp t are the data (x_t), cell state (c_{t-1}), and hidden state (h_{t-1}). c_{t-1} and h_{t-1} are the output of the LSTM cell in the previous time stamp. There are three gating units in an LSTM cell, i.e., forget gate (f_t), input gate (i_t), and output gate (o_t) to control the flow of information. All the gates take x_t and h_{t-1} as input and use sigmoid function as an activation function. The forget gate erases the information present in c_{t-1} , which mayn't be useful further. The forget gate (f_t) is expressed as;

$$f_t = \text{sigm}(W_{fx}x_t + W_{fh}h_{t-1} + b) \quad (1)$$

The representation learned from the current time stamp is given by \tilde{c}_t , which is computed using eq (2). Then, a point wise multiplication is done only to retain the necessary information. The input gate (i_t) and the cell state (c_t) are computed by using eq (3) and eq (4), respectively.

$$\tilde{c}_t = \text{tanh}(W_{gx}x_t + W_{gh}h_{t-1} + b) \quad (2)$$

$$i_t = \text{sigm}(W_{ix}x_t + W_{ih}h_{t-1} + b) \quad (3)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (4)$$

The output of the LSTM cell at the current time stamp is the hidden state (h_t), which is used in decision making. The hidden state is obtained by removing certain information that may not contribute to the decision at time stamp t . This is done by using the output gate (o_t). The mathematical expression to obtain o_t and h_t are shown in eq (5) and eq (6).

$$o_t = \text{sigm}(W_{ox}x_t + W_{oh}h_{t-1} + b) \quad (5)$$

$$h_t = o_t \odot c_t \quad (6)$$

The output taken from the last time stamp of a recurrent LSTM cell connection is called a vanilla LSTM architecture. Although the vanilla LSTM performs good, it does not learn the long term dependencies [24], [28]. This makes it perform poorer, where the temporal information is present over a long period of time. This has motivated us to design a hierarchical LSTM model, which can account for long term dependencies.

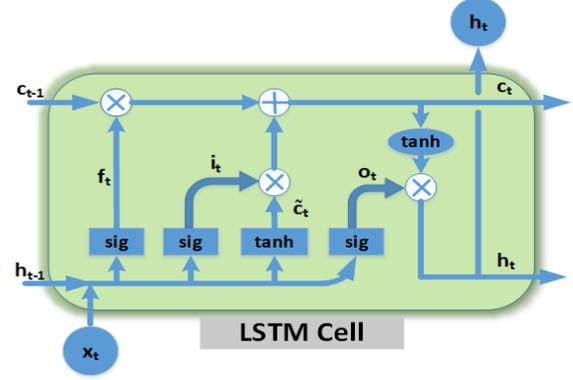


Fig. 4. The architecture of an LSTM cell

1) *Hierarchical LSTM*: One of the major advantages of the deep neural network is its ability to learn the representation at different levels of abstraction [29]. The deep learning architectures like CNN learn the hierarchical representation in the spatial domain. Unlike CNN, the vanilla LSTM doesn't learn the hierarchical representation in the temporal dimension. To address this issue, we have designed a hierarchical LSTM model by stacking layers of LSTM. The proposed hierarchical model is designed to learn the hierarchical temporal representation, which in turn can address the long term dependency issue of the vanilla LSTM model. In this work, we have designed the hierarchical LSTM model with two layers of LSTM. Fig 5(a) shows the proposed hierarchical LSTM model.

The layer1 of the architecture, which is represented as L_s in Fig 5, models the segment level representations of the ECG signal. The input to L_s are the ECG segments of an ECG sequence, which are represented by $\{x_1^s, x_2^s, x_3^s, \dots, x_{T^s}^s\}$. Here the length of the input ECG sequence is T^s , which in turn is the length of time steps in L_s . At each time step the LSTM cell gives an output, which is given as;

$$O_{t^s}^s = L_s(h_{t^s-1}^s, x_{t^s}^s) \quad (7)$$

The layer2 of the architecture is represented as L_B . This layer is introduced to learn the intra-beat and inter-beat variation effectively. The input to L_B is the output of L_s taken in an interval of l . The range of time steps of L_B layer, t^B , depends on the interval l . The time steps, i.e., t^B , and input to L_B , i.e., $x_{t^B}^B$ are expressed in eq (8) and eq (9), respectively.

$$t^B \in \{0, 1, \dots, \left\lceil \frac{T^s}{l} \right\rceil + 1\} \quad (8)$$

$$x_{t^B}^B \in \{O_{1^s}^{L_s}, \dots, O_{t^B \times l^s}^{L_s}, \dots, O_{T^s}^{L_s}\} \quad (9)$$

The output of the L_B at each time step is given by;

$$O_{t^B}^{L_B} = L_B(h_{t^B-1}^{L_B}, x_{t^B}^B) \quad (10)$$

The output of L_B layer is given to the attention block. The attention block decides the weight that is to be given to the output vector of L_B at each time step.

2) *Attention Model*: The attention mechanism has gained popularity due to its successful application in the field of natural language processing (NLP), computer vision (CV), and automatic speech recognition (ASR). The attention mechanism

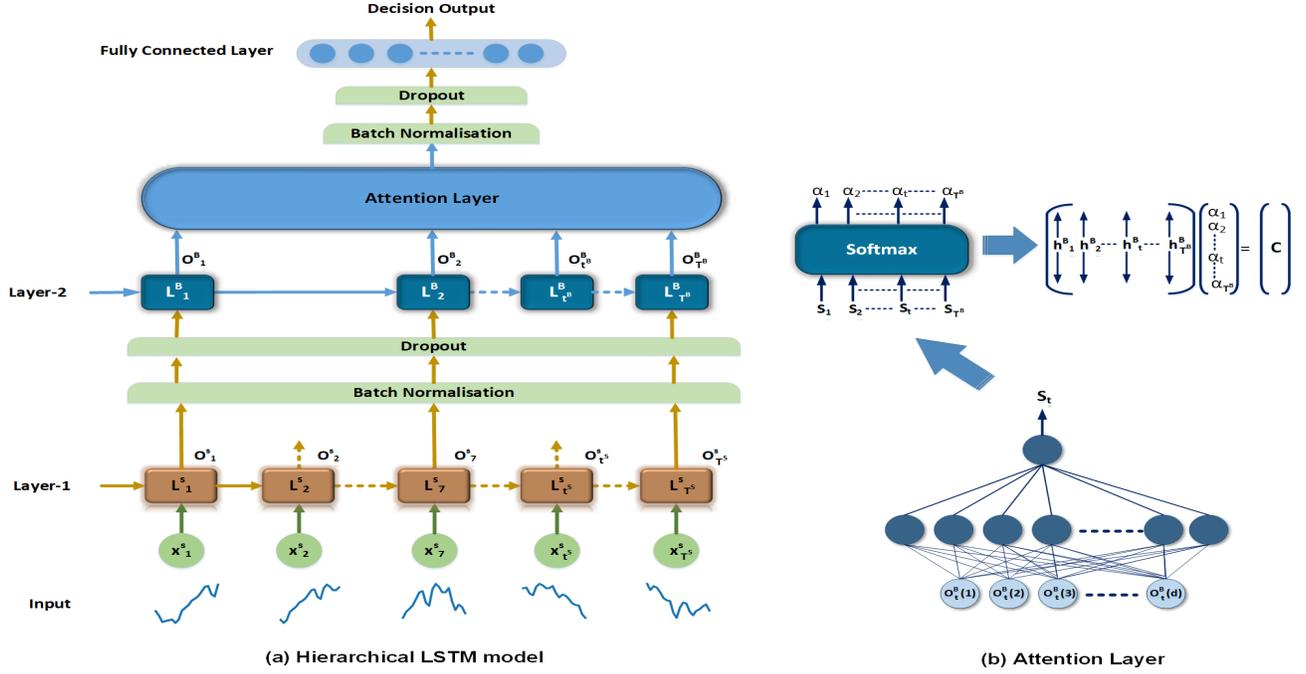


Fig. 5. (a) Architecture of the proposed attention based hierarchical LSTM model. (b) Attention module used in this framework

allows the model to focus on necessary portions of the learned representation. This allows the RNN based models to learn better representation by using the information from all the time stamps, which otherwise uses only the last time stamp's output. Various attention mechanisms have been proposed in the literature addressing the location of the attention and weights given to the area of attention [28], [30]–[32]. In the ECG-based biometric system, different ECG complexes of the ECG signal have different levels of contribution towards the identification problem [11]. The attention mechanism in this work allows the model to give the necessary attention needed to a particular portion of the ECG signal. Fig 5 (b) shows the attention mechanism used in this work. The inputs to the attention block are the outputs of L_B layer, i.e., $\{O_0^{L_B}, O_1^{L_B}, \dots, O_{T^B}^{L_B}\}$. Each output of L_B layer is passed through a fully connected layer to obtain a score value as in eq (12). Then all the score values are fed to the softmax function to obtain the attention weights. Finally, the context vector (C) is obtained by a weighted summation of the outputs of L_B layer. This is expressed in eq (11) and eq (13), respectively.

$$\alpha_t = \text{softmax}(\text{score}(O_t^{L_B})) = \frac{\exp(\text{score}(O_t^{L_B}))}{\sum_{t=1}^{T(B)} \exp(\text{score}(O_t^{L_B}))} \quad (11)$$

$$\text{score}(O_t^{L_B}) = W_A O_t^{L_B} + b_A \quad (12)$$

$$C = \sum_{t=1}^{T(B)} c_t = \sum_{t=1}^{T(B)} \alpha_t O_t^{L_B} \quad (13)$$

The context vector obtained from the attention block is a representation vector of a person's ECG signal. This vector is further used either for person identification or verification, depending on the mode of operation.

D. Identification Mode

In the identification mode, the context vector is used to identify the individual from a pool of registered individuals. For the identification, the context vector is fed to a fully connected layer with a softmax activation function and an output dimension equal to the number of registered individuals. The fully connected layer gives the posterior probability output ($\hat{P}(S_k/C)$) of the context vector (C) belonging to a subject S_k .

$$\hat{P}(S_k/C) = \text{softmax}(WC + b) \quad (14)$$

The subject corresponding to the maximum posterior probability is identified as the candidate subject. The parameters of the model are trained by using the multiclass cross entropy loss L . The cross entropy loss L is computed over all the training samples as;

$$L = - \sum_{n=1}^N \sum_{k=1}^K P(S_k/C) \log(\hat{P}(S_k/C)) \quad (15)$$

where N is the number of samples, and K is the number of subjects enrolled. $P(S_k/C)$ stands for the true posterior probability of the training sample belonging to subject S_k .

E. Verification Mode

In the verification mode, the context vector is used to verify the authenticity of the claimed identity. In this case, the model is first trained in the identification mode with the ECG data taken from a pool of subjects. Then the trained model is used to enrol a different set of subjects. The enrolment is done by finding out the ECG identity vector ($iECG$) for each subject in the enrolment set. The identity vector is obtained by taking the average of the normalised context vectors, C_{n_e} , for all

the templates corresponding to a subject. The expression to compute $iECG$ is given in eq (16).

$$iECG = -\frac{1}{N_e} \sum_{n_e=1}^{N_e} \frac{C_{n_e}}{\|C_{n_e}\|} \quad (16)$$

During verification, the cosine similarity value between the claimant $iECG$ and the claimed $iECG$ is computed. Finally, the authenticity is verified by comparing it with a pre-set threshold value.

IV. EXPERIMENTAL SETUP

The proposed model is evaluated both for person identification and verification purpose. For the evaluation, we have used four publicly available databases. Out of the four databases, two are recorded in the on-the-person setting, i.e., PTB database [33], ECG-ID database [34], and two are recorded in the off-the-person setting, i.e., CYBHi database [35], UofTDB database [36]. We have evaluated our method for two scenarios; intra-session and inter-session. In the case of person identification in an intra-session scenario, both the training and the testing samples are collected from the same recording. However, unlike the works proposed in the literature, we have extracted the training and testing samples from different locations of the recording. This is done by extracting the training samples first, and then after a time duration, the testing samples are collected. This process minimises the biasing in the testing data, which otherwise would have resulted in false improvement in the performance. For verification in the intra-session scenario, the recordings are first randomly divided into two sets without any overlap of subjects; training set and enrolment set. The training set recordings are used to train the model. The data from the enrolment set are used to obtain the enrolment data and the testing data for each subject. This process of extracting the enrolment and the testing data is similar to the training and the testing data extraction process in the identification mode.

In inter-session scenario, the training and the testing samples are collected from recordings of different sessions for identification. In case of verification in an inter-session scenario, both the recordings of a subject belonging to the training set are used for training the model. The enrolment data and the testing data are extracted from the recordings of different sessions. We have briefly described the datasets used in this work below.

1) *PTB database*: PTB ECG database is a publicly available database recorded using both the conventional 12-lead ECG recording system and the frank lead recording system. The database contains records from 290 subjects. In this work, we have used the ECG signal of limb lead II, which is generally in the direction of the heart's electrical position.

2) *ECG-ID database*: The ECG-ID database contains data recorded from 90 subjects. Each subject in the database, except subject number 74, has at least two recordings taken in multiple different sessions. The recordings contain limb lead I ECG signals recorded for a duration of 20 seconds. In the case of the intra-session scenario, we have extracted half of the total training and testing data from one recording. While,

in the case of the inter-session scenario, all the training data are extracted from one recording and the testing data from the other recording.

3) *CYBHi database*: The CYBHi database is a publicly available ECG database recorded in the off-the-person recording setup. The database contains ECG recordings taken from a subject in two different sessions. Depending on the duration between the two different sessions, the data have been divided into two categories; short-term and long-term. The short-term recordings are taken from an individual within two days, and the long-term recordings are taken from an individual with an interval of three months. The long-term data have been collected from 63 subjects, and the short-term data have recordings from 65 subjects. In this work, we have used long-term data. While extracting the ECG sequences, some ECG sequences that are completely distorted have been replaced by better ECG sequences.

4) *UofTDB database*: The UofTDB database is the largest publicly available ECG database for the biometric application. The database contains ECG recordings of 1019 subjects recorded in the off-the-person recording setup. The data were collected for six different sessions in five different body postures. The distribution of the number of subjects in different sessions and different body postures are given in Table I. The session-I data of 1019 subjects taken in sitting posture is used in the intra-session analysis. For inter-session analysis, recordings of 82 subjects are used who have more than one recordings in sitting posture. The session-I data constitute one set of data, while another set of data consists of one recording per subject selected from the rest of the sessions. We have replaced some of the ECG sequences that are completely distorted during recording with better ECG sequences from that recording.

TABLE I
STATISTICAL DISTRIBUTION OF NUMBER OF SUBJECTS IN DIFFERENT SESSIONS

Session	Sit	Stand	Exercise	Supine	Tripod
S1	1012	0	0	0	0
S2	72	72	0	0	0
S3	76	5	71	0	0
S4	63	0	0	0	0
S5	0	0	0	63	63
S6	65	65	0	0	0

A. Performance Measure

The performance of the algorithm is evaluated using two parameters equal error rate (EER) and accuracy (Acc). EER is the trade-off point between false acceptance rate (FAR) and false rejection rate (FRR), where both FAR and FRR becomes equal. FAR stands for the rate of false acceptance of an imposter claim, and FRR stands for the rate of false rejection of a legitimate claim. In this case, all the enrolled subjects, except the legitimate one, have been used as an imposter. The accuracy is used in the case of person identification. It is obtained by evaluating the number of testing templates correctly identified against a total number of testing templates.

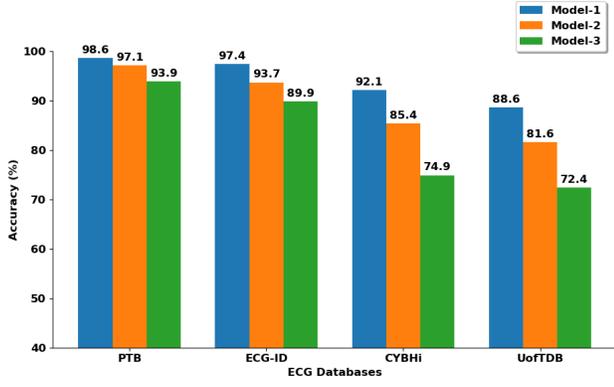


Fig. 6. Comparison of identification accuracy for three LSTM based architectures.

B. Network Parameters

The parameters of the proposed attention based hierarchical LSTM model was trained in an end-to-end manner using a stochastic gradient descent based optimisation algorithm. The initial learning rate of the model is set at 0.01. The learning rate is varied up to 0.001 using a cosine learning rate scheduler. The layer-1 LSTM has a hidden dimension of 150, and the layer-2 LSTM has a hidden dimension of 100. The dropout layers had a dropout ratio of 0.4. The deep learning model was implemented using the PyTorch programming library. The rest of the implementations have been done using the python programming language. All the experiments are conducted in a Titan P-100 GPU server facility.

V. RESULTS AND DISCUSSION

The person identification accuracy of the proposed model (Model-1) is obtained for all four datasets, and the results are compared with the simple vanilla LSTM model (Model-3) and the hierarchical LSTM model without attention mechanism (Model-2). The vanilla LSTM model is an LSTM architecture with only one layer, where only the final output vector is considered for identification purposes. The performance comparison is shown in Fig. 6. We have considered a cluster interval, l , of 7 in the case of the hierarchical LSTM model. It can be observed that the simple vanilla LSTM performs very poorly for the person identification application. From Fig. 6 it can be observed that the hierarchical LSTM performs significantly better than the vanilla LSTM model. This shows that the hierarchical LSTM model can capture the temporal variation of the ECG signal in different abstractions. Thus, the hierarchical LSTM model solves the long term dependency problem of the vanilla LSTM model, and it models the intra-beat variation and inter-beat variation better. Among the three models, the proposed attention based hierarchical LSTM model performs the best. The proposed model's attention mechanism helps give more weight to the segments, which are important for the identification problem. Fig. 7 shows the attention weight given to ECG sequences taken from two different persons. Fig. 7(a1) and Fig. 7(b1) are the ECG sequences of two different persons and their respective figures with normalised attention

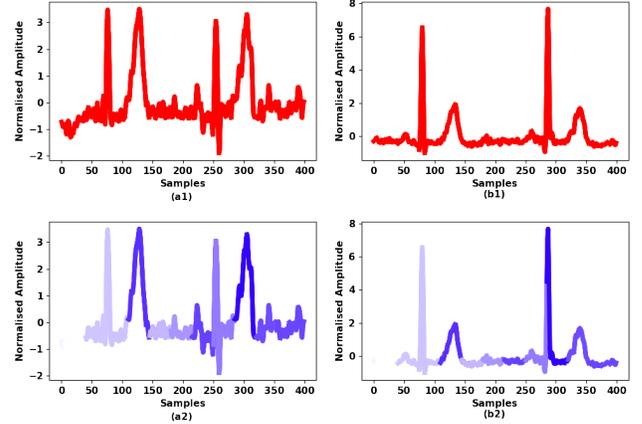


Fig. 7. (a1) and (b1) shows the ECG sequence of two different persons. (a2) and (b2) shows the attention weight given by the model to different portions of the ECG sequence in (a1) and (b1), respectively. The portions with more colour saturation correspond to more attention weights.

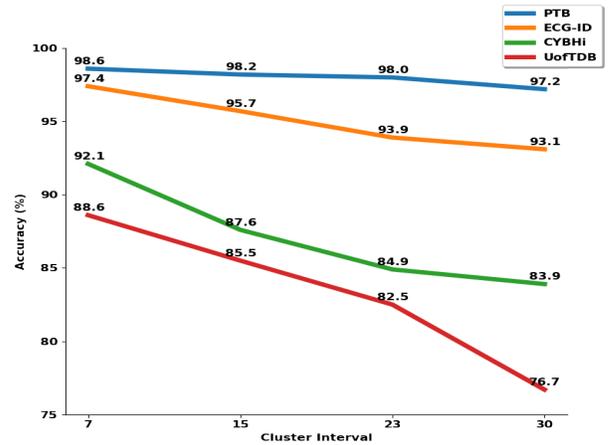


Fig. 8. Variation of person identification accuracy of the model for different cluster lengths.

weights are given in Fig. 7(a2) and Fig 7(b2). The portions with more colour saturation have been given more attention weight compared to the portions with less colour saturation. From Fig. 7(a2), it can be observed that more weight has been given to the T wave. In the case of ECG sequence in Fig. 7(b2), more weight has been given to both the QRS complex and T wave. Thus, the attention mechanism learns to give more weight to the portions of the ECG signal that have more biometric information corresponding to each subject.

The identification accuracy of the proposed model for different cluster interval (l) has been calculated, and the results are shown in Fig. 8. From Fig. 8, it can be observed that the best performance is obtained for a cluster interval of 7 and the identification accuracy decreases as we increase the cluster length. This may be because smaller cluster lengths give better temporal resolution, and thus, the model can learn the intra-beat and inter-beat variations better for smaller cluster length.

The EER values of the proposed model operated in the identification mode are obtained for different databases. The results obtained are tabulated in Table II. We have obtained

a significantly low EER of 0.89% and 1.7% for the intra-session analysis of PTB and ECG-ID database, respectively. Similar performance has also been observed in the case of the CYBHi and UofTDB databases with a minimum EER of 3.98% and 2.19%, respectively. The results of inter-session analysis have also been tabulated in Table II. A low EER of 3.16% has been obtained for the ECG-ID database, which is of the on-the-person recording category. The inter-session analysis of the CYBHi dataset has been done in two ways. Once the model is trained using (S1) data and then evaluated using (S2) data, which is represented by (S1-S2). The next time the model is trained using (S2) data and evaluated using (S1) data, which is represented by (S2-S1). We have obtained an EER of 13.43% and 12.05% for the S1-S2 session and S2-S1 session analysis, respectively. In the case inter-session analysis using the UofTDB database, we have obtained an EER of 11.36%.

TABLE II
EER VALUE OBTAINED BY THE MODEL IN IDENTIFICATION MODE

EER (in %)	Database	PTB	ECG-ID	CYBHi	UofTDB
	Intra-session	0.89	1.70	3.98	2.19
Inter-session	-	3.16	13.43	11.36	

The performance of the model in the verification mode has been shown in Table III. The evaluation is done by using different percentages of data as training data and the rest as enrolment data. The lowest EER is obtained when more data are used for training. However, it can be observed that the model gives comparable results with less training data also. This shows that the performance of the model is robust. The model has also been evaluated in the verification mode using the inter-session data. Fig. 9 shows the graph between FRR and FAR of the results obtained for the inter-session analysis using UofTDB and ECG-ID database. An EER of 4.25% and 10.55% are obtained for inter-session analysis of the ECG-ID database and UofTDB databases, respectively.

TABLE III
EER VALUE OBTAINED BY THE MODEL IN VERIFICATION MODE

Database	Share of data used for training (in %)		
	40	60	80
PTB	1.89	1.39	1.33
ECG-ID	4.81	4.2	3.13
UofTDB	1.94	2.10	2.06

A. Comparison

A comparison of the proposed model with the state-of-the-art works has been made in this section. Various aspects of the biometric system have been taken into account for comparison. We have considered recent deep learning based methods, i.e., [10], [15], [20], [21], [37], [38], for comparison in both on-the-person category and off-the-person category. A comparison of the literature works on the on-the-person ECG database has been made in Table IV.

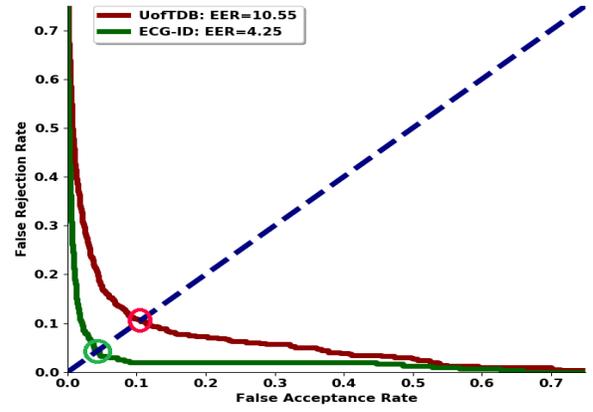


Fig. 9. DET curve obtained from the inter-session analysis of ECG-ID and UofTDB database in the verification mode.

TABLE IV
COMPARISON WITH THE STATE-OF-THE-ART WORKS USING ON-THE-PERSON ECG DATA. FPD STANDS FOR FIDUCIAL POINT DETECTION.

Database	Works	Session type	No. of Subjects	FPD	Amount of training data	Acc.
PTB	[20] (2019)	intra	52	Yes	40 s	100%
	[12] (2013)	intra	100	Yes	100 templates of 2-4 s	99.48%
	[15] (2020)	intra	290	No	96 s on average	94.9%
	[39] (2019)	intra	285	Yes	15 ECG beats	97.19%
	Proposed	intra	290	No	12.5s	98.6%
	Proposed	intra	100	No	12.5s	100%
	Proposed	intra	52	No	12.5s	100%
ECG-ID	[37] (2018)	intra	50	Yes	36 s on average	96.63%
	[40] (2015)	intra	90	Yes	-	83.88%
	Proposed	intra	90	No	12.5s	97.4%
	Proposed	intra	90	No	12.5s	92.32%

From Table IV, it can be observed that the proposed method performs best in the on-the-person category. Unlike most of the methods in the literature, we have evaluated our model for all the subjects present in the database. For better comparison with the models with less enrolment in the case of the PTB database, we have evaluated our model for 52 healthy subjects and 100 subjects taken randomly. Our model achieves 100% accuracy for both the cases outperforming the existing models. From the table, it can be observed that the proposed model requires the least amount of data for training. The non-fiducial point based CNN architecture proposed in [15] requires an average of 96 seconds of training data per subject, which is significantly more than the 12.5 seconds of training data used in this work. The LSTM based method proposed in [21] is implemented in our experimental setup, and the results are compared in Table IV.

Table V represents the comparison of the proposed model with the works in literature for the off-the-person ECG database. From Table V, it can be observed that the proposed model performs better than the methods proposed in the literature except in two cases, i.e., inter-session (S1-S2) analysis of [10] and inter-session analysis of UofTDB [3].

TABLE V
COMPARISON WITH THE STATE-OF-THE-ART WORKS USING
OFF-THE-PERSON ECG DATA. FPD STANDS FOR FIDUCIAL POINT
DETECTION.

Database	Works	Session type	No. of Subjects	FPD	Amount of training data	EER
CYBHi	[10] (2018)	intra	63	Yes	50% of data	1.33%
	[10] (2018)	inter (S1-S2)	63	Yes	50% of data	12.78%
	[10] (2018)	inter (S2-S1)	63	Yes	50% of data	13.93%
	[38] (2019)	intra	63	Yes	90% of data	4.47%
	Proposed	intra	63	No	12.5s	3.98%
	Proposed	inter (S1-S2)	61	No	12.5s	13.43%
	Proposed	inter (S2-S1)	61	No	12.5s	12.05%
UofTDB	[10] (2019)	inter	82	Yes	50% of data	14.27%
	[3] (2016)	intra	1012	Yes	80% of data	7.89%
	[3] (2016)	inter	82	Yes	80% of data	10.10%
	Proposed	intra	1019	No	12.5s	2.19%
	Proposed	inter	82	No	12.5s	11.36%

However, the proposed model performs better than [10] in the (S2-S1) session analysis. Similarly, our method performs significantly better than the work in [3] for the intra-session analysis of UofTDB data. The literature methods benefit from the cosine similarity based outlier removal method employed in the preprocessing stage. Thus, the works in literature have used similar templates of a subject for the training and testing.

From the comparative analysis, it can be understood that the proposed biometric system has several important advantages over the existing methods. The attention based hierarchical LSTM model requires a very low duration of enrolment data, i.e., 12.5 seconds of ECG recording. Similarly, the model requires only 2 seconds of data during verification. The proposed framework follows a non-fiducial point based approach, which makes it more suitable for application in a practical scenario. The LSTM based framework learns the representation of the temporal variation of the ECG signal in two different levels of abstraction. This might allow the model to learn the differences between different cardiac signals, which arise due to the physiological differences of cardiovascular systems [41]. The physiological differences between different cardiovascular systems are reflected in the systolic and diastolic phase of the ECG signal. The proposed model is designed to exploit these differences optimally with the help of the attention mechanism. The attention mechanism could give more weight to portions of the ECG signal important for identification purpose, which leads to better performance of the model.

VI. CONCLUSION

In this paper, we have proposed a novel LSTM based framework for the ECG based biometric system. The proposed LSTM model's hierarchical architecture could learn the temporal representation of the ECG signal in different hierarchies, addressing the long-term dependency issue. A smaller cluster

interval in the hierarchical structure has been found to be better in learning the hierarchical representation of the temporal variation of the ECG signal. The attention mechanism used in this model could learn the important ECG complexes for person identification, which boosts the performance of the model. The proposed model performs superior both in the on-the-person and off-the-person ECG data. However, the performance of the model may improve if the low quality ECG signal is discarded beforehand. In our future work, we would like to develop a method for the assessment of the quality of the ECG signal for biometric application.

REFERENCES

- [1] J. S. Arteaga-Falconi, H. Al Osman, and A. El Saddik, "Ecg authentication for mobile devices," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 3, pp. 591–600, 2015.
- [2] S. Pirbhulal, H. Zhang, W. Wu, S. C. Mukhopadhyay, and Y.-T. Zhang, "Heartbeats based biometric random binary sequences generation to secure wireless body sensor networks," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 12, pp. 2751–2759, 2018.
- [3] W. Louis, M. Komeili, and D. Hatzinakos, "Continuous authentication using one-dimensional multi-resolution local binary patterns (1dmrlbp) in ecg biometrics," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 12, pp. 2818–2832, 2016.
- [4] E. M. Antman and J. Loscalzo, "Precision medicine in cardiology," *Nature Reviews Cardiology*, vol. 13, no. 10, p. 591, 2016.
- [5] T. Ince, S. Kiranyaz, and M. Gabbouj, "A generic and robust system for automated patient-specific classification of ecg signals," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 5, pp. 1415–1426, 2009.
- [6] A. K. Jain, A. Ross, S. Prabhakar *et al.*, "An introduction to biometric recognition," *IEEE Transactions on circuits and systems for video technology*, vol. 14, no. 1, 2004.
- [7] M. Merone, P. Soda, M. Sansone, and C. Sansone, "Ecg databases for biometric systems: A systematic review," *Expert Systems with Applications*, vol. 67, pp. 189–202, 2017.
- [8] J. R. Pinto, J. S. Cardoso, and A. Lourenço, "Evolution, current challenges, and future possibilities in ecg biometrics," *IEEE Access*, vol. 6, pp. 34 746–34 776, 2018.
- [9] I. Odinaka, P.-H. Lai, A. D. Kaplan, J. A. O'Sullivan, E. J. Sirevaag, and J. W. Rohrbach, "Ecg biometric recognition: A comparative analysis," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 6, pp. 1812–1824, 2012.
- [10] E. J. da Silva Luz, G. J. Moreira, L. S. Oliveira, W. R. Schwartz, and D. Menotti, "Learning deep off-the-person heart biometrics representations," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 5, pp. 1258–1270, 2017.
- [11] A. D. Chan, M. M. Hamdy, A. Badre, and V. Badee, "Wavelet distance measure for person identification using electrocardiograms," *IEEE transactions on instrumentation and measurement*, vol. 57, no. 2, pp. 248–253, 2008.
- [12] J. Wang, M. She, S. Nahavandi, and A. Kouzani, "Human identification from ecg signals via sparse representation of local segments," *IEEE Signal Processing Letters*, vol. 20, no. 10, pp. 937–940, 2013.
- [13] Y. Wang, F. Agrafioti, D. Hatzinakos, and K. N. Plataniotis, "Analysis of human electrocardiogram for biometric recognition," *EURASIP journal on Advances in Signal Processing*, vol. 2008, no. 1, p. 148658, 2007.
- [14] Q. Zhang, D. Zhou, and X. Zeng, "Heartid: A multiresolution convolutional neural network for ecg-based biometric human identification in smart health applications," *IEEE Access*, vol. 5, pp. 11 805–11 816, 2017.
- [15] S. S. Abdeldayem and T. Bourlai, "A novel approach for ecg-based human identification using spectral correlation and deep learning," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2019.
- [16] J. Liu, L. Yin, C. He, B. Wen, X. Hong, and Y. Li, "A multiscale autoregressive model-based electrocardiogram identification method," *IEEE Access*, vol. 6, pp. 18 251–18 263, 2018.
- [17] J. M. Irvine, S. A. Israel, W. T. Scruggs, and W. J. Worek, "eigenpulse: Robust human identification from cardiovascular function," *Pattern Recognition*, vol. 41, no. 11, pp. 3427–3435, 2008.
- [18] Z. Zhao and L. Yang, "Ecg identification based on matching pursuit," in *2011 4th International conference on biomedical engineering and informatics (BMEI)*, vol. 2. IEEE, 2011, pp. 721–724.

- [19] C. L. P. Lim, W. L. Woo, S. S. Dlay, and B. Gao, "Heart-rate-dependent heartwave biometric identification with thresholding-based gmm-hmm methodology," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 1, pp. 45–53, 2018.
- [20] R. D. Labati, E. Muñoz, V. Piuri, R. Sassi, and F. Scotti, "Deep-ecg: Convolutional neural networks for ecg biometric recognition," *Pattern Recognition Letters*, vol. 126, pp. 78–85, 2019.
- [21] R. Salloum and C.-C. J. Kuo, "Ecg-based biometrics using recurrent neural networks," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 2062–2066.
- [22] L. Sharma, S. Dandapat, and A. Mahanta, "Ecg signal denoising using higher order statistics in wavelet subbands," *Biomedical Signal Processing and Control*, vol. 5, no. 3, pp. 214–222, 2010.
- [23] Y. Sun, K. L. Chan, and S. M. Krishnan, "Ecg signal conditioning by morphological filtering," *Computers in biology and medicine*, vol. 32, no. 6, pp. 465–479, 2002.
- [24] D. Jyotishi and S. Dandapat, "An lstm based model for person identification using ecg signal," *IEEE Sensors Letters*, 2020.
- [25] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "Lstm: A search space odyssey," *IEEE transactions on neural networks and learning systems*, vol. 28, no. 10, pp. 2222–2232, 2016.
- [26] V. Campos, B. Jou, X. G. i Nieto, J. Torres, and S.-F. Chang, "Skip RNN: Learning to skip state updates in recurrent neural networks," in *International Conference on Learning Representations*, 2018. [Online]. Available: <https://openreview.net/forum?id=HkwVAXyCW>
- [27] A. G. A. P. Goyal, A. Sordani, M.-A. Côté, N. R. Ke, and Y. Bengio, "Z-forcing: Training stochastic recurrent networks," in *Advances in neural information processing systems*, 2017, pp. 6713–6723.
- [28] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.
- [29] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.
- [30] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio, "Show, attend and tell: Neural image caption generation with visual attention," in *International conference on machine learning*, 2015, pp. 2048–2057.
- [31] M.-T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," *arXiv preprint arXiv:1508.04025*, 2015.
- [32] J. Cheng, L. Dong, and M. Lapata, "Long short-term memory-networks for machine reading," *arXiv preprint arXiv:1601.06733*, 2016.
- [33] R. Bousseiljot, D. Kreiseler, and A. Schnabel, "Nutzung der ekg-signal-datenbank cardiodat der ptb über das internet," *Biomedizinische Technik/Biomedical Engineering*, vol. 40, no. s1, pp. 317–318, 1995.
- [34] T. S. Lugovaya, "Biometric human identification based on ecg," *PhysioNet*, 2005.
- [35] H. P. Da Silva, A. Lourenço, A. Fred, N. Raposo, and M. Aires-de Sousa, "Check your biosignals here: A new dataset for off-the-person ecg biometrics," *Computer methods and programs in biomedicine*, vol. 113, no. 2, pp. 503–514, 2014.
- [36] S. Wahabi, S. Pouryayevali, S. Hari, and D. Hatzinakos, "On evaluating ecg biometric systems: session-dependence and body posture," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 11, pp. 2002–2013, 2014.
- [37] Z. Zhao, Y. Zhang, Y. Deng, and X. Zhang, "Ecg authentication system design incorporating a convolutional neural network and generalized s-transformation," *Computers in biology and medicine*, vol. 102, pp. 168–179, 2018.
- [38] M. Hammad, S. Zhang, and K. Wang, "A novel two-dimensional ecg feature extraction and classification algorithm based on convolution neural network for human authentication," *Future Generation Computer Systems*, vol. 101, pp. 180–196, 2019.
- [39] S.-C. Wu, P.-T. Chen, A. L. Swindlehurst, and P.-L. Hung, "Cancelable biometric recognition with ecgs: subspace-based approaches," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 5, pp. 1323–1336, 2018.
- [40] M. N. Dar, M. U. Akram, A. Usman, and S. A. Khan, "Ecg biometric identification for general population using multiresolution analysis of dwt based features," in *2015 Second International Conference on Information Security and Cyber Forensics (InfoSec)*. IEEE, 2015, pp. 5–10.
- [41] R. Hoekema, G. J. Uijen, and A. Van Oosterom, "Geometrical aspects of the interindividual variability of multilead ecg recordings," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 5, pp. 551–559, 2001.