Beat-to-beat continuous blood pressure estimation with optimal feature set of PPG and ECG signals using deep recurrent neural networks

Hanjie Chen¹, Liangyi Lyu¹, Zezhen Zeng², Yanwei Jin¹, Yuanting Zhang¹*¹

¹Hong Kong Center for Cerebro-Cardiovascular Health Engineering (COCHE), Hong Kong 999077, China.
²Research Center for applied Mathematics and Machine Intelligence, Zhejiang Lab, Hangzhou 310000, Zhejiang, China.
³Department of Electronic Engineering, the Chinese University of Hong Kong, Hong Kong 999077, China.

*These authors contributed equally to this work and should be considered co-first authors.

Correspondence to: Prof. Yuanting Zhang, Department of Electronic Engineering, the Chinese University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong 999077, China. E-mail: ytzhang@cuhk.edu.hk


Received: 15 May 2023  First Decision: 7 Sep 2023  Revised: 16 Sep 2023  Accepted: 12 Oct 2023  Published: 20 Oct 2023

Abstract

Aim: Continuous blood pressure (BP) monitoring can provide invaluable information for cardiovascular disease (CVD) diagnosis. The purpose of this study is to develop a deep recurrent neural network (RNN) model with an optimal feature set of photoplethysmogram (PPG) and electrocardiogram (ECG) signals for continuous BP estimation.

Methods: This paper presents a novel deep recurrent neural network (RNN), which consists of 2-layered bidirectional Long Short-term Memory (Bi-LSTM) and 6-layered LSTM networks. It is used to estimate BP based on the optimal feature set of PPG and ECG signals. In this work, the optimal feature set is determined using five different feature selection methods.

Results: The proposed method is evaluated based on 660 subjects from the University of California Irvine (UCI) machine learning repository. The RNN model with optimal feature set achieved root mean square error (RMSE) of 3.223 and 1.781 mmHg for systolic BP (SBP) and diastolic BP (DBP), respectively. It also showed mean absolute error (MAE) of 2.514 and 1.383 mmHg for SBP and DBP, respectively. Regarding the British Hypertension Society (BHS) standard, the results attained grade A for the estimation of SBP and DBP.
Conclusion: The experimental results suggest that the proposed deep RNN model with an optimal feature set can improve the performance of BP prediction. Thus, it is possible to further apply our proposed method to develop a wearable device for real-time BP monitoring.

Keywords: Continuous blood pressure, photoplethysmogram, electrocardiogram, feature selection, recurrent neural network

INTRODUCTION

According to the World Health Organization (WHO), hypertension has remained one of the leading causes of morbidity and mortality worldwide\(^1\)\(^2\). Hypertension, as one of the major causes of cardiovascular disease, is often asymptomatic in the early stage and thus has a low diagnosis rate\(^3\). Therefore, continuous and stable monitoring of blood pressure is essential for preventing cardiovascular diseases\(^4\). Blood pressure (BP) refers to the force exerted on the walls of blood vessels by the flow of blood caused by the heartbeat and serves as a critical indicator of cardiovascular health. An electrocardiogram (ECG) is a noninvasive diagnostic procedure that captures and records the electrical activity of the heart. By analyzing the ECG signals, it becomes possible to observe the electrical activity of the heart, which is closely linked to changes in blood pressure\(^5\). Photoplethysmographic volumetric tracing (PPG) is an optical technique that captures volumetric tracings, allowing for the detection of changes in blood volume within the small blood vessels within the tissues. The formation of PPG signals is closely correlated with blood pressure and can reflect a wealth of information about cardiovascular function. The combination of ECG and PPG computations can yield various physiological indicators, such as Pulse Arrival Time (PAT), Pulse Transmission Time (PTT), and Pulse Wave Velocity (PWV), which are highly correlated with blood pressure. Therefore, the combination of ECG and PPG signals plays a crucial role in predicting and managing blood pressure for healthcare purposes\(^6\).

There are two main methods used for blood pressure measurement: the invasive method and the noninvasive method. The invasive method is widely regarded as the gold standard for blood pressure (BP) measurement\(^7\). Compared to the noninvasive cuff-based method, cuffless approaches can easily obtain individual continuous blood pressure in different situations. There have been several recent methods for cuffless BP estimation, including PTT model\(^8\), PTT and photoplethysmogram intensity ratio (PIR) model\(^9\) and machine learning (ML) models, such as support vector regression\(^10\), random forest regression\(^4\) and Gaussian process regression (GPR)\(^1\). In addition, a model based on convolutional neural networks (CNN) for extracting feature vectors and using LSTM to capture temporal dependencies was proposed in\(^11\). A complete BP estimation model contains two parts: feature extraction and model fitting. However, with the increasing number of features, the computational cost and model complexity will be increased. Conversely, the smaller number of features may not be able to extract enough information, leading to a poor performance of the BP estimation model. It is hard to say the feature set is optimal if only a specific number of features were used. Hence, the development of a feature selection system becomes crucial to identify the optimal feature set for accurate blood pressure estimation.

In this paper, we introduce a novel beat-to-beat blood pressure (BP) monitoring technique that combines a feature selection system with a deep recurrent neural network (RNN). The RNN architecture comprises a 2-layered bidirectional long short-term memory (Bi-LSTM) network and a 6-layered LSTM network. A total number of 28 features were first extracted from all PPG and ECG beats. Then, a proposed feature voting system consisting of five feature selection methods was applied to select the optimal feature sets. Subsequently, the proposed RNN model was used to estimate DBP, SBP, and MBP. We evaluated the
performance of our proposed model using the University of California Irvine (UCI) dataset as the benchmark. Our proposed model shows a better performance than other previously formally published models for BP estimation.

The remainder of the paper is structured as follows: Section “METHODOLOGY” provides the feature extraction, feature selection methods, and the architecture of the RNN model. Section “RESULTS” shows the data and results of our proposed model. In Section “DISCUSSION”, we conclude the paper and outline potential avenues for future research.

**METHODOLOGY**

This section introduces the proposed three-stage process for high-performance blood pressure (BP) estimation using the RNN model with an optimal feature set of photoplethysmogram (PPG) and electrocardiogram (ECG) signals. The algorithm is designed to achieve accurate BP estimation and is outlined in Figure 1. The initial stage of our approach involves feature extraction, which encompasses preprocessing the PPG and ECG signals and extracting a total of 28 features. In the following step, we perform feature selection by applying five distinct feature selection methods. Our objective is to identify the optimal feature set that ensures accurate estimation of blood pressure (BP). Finally, the proposed deep RNN model is developed by randomly taking 594 subjects (90%) from the University of California Irvine (UCI) machine learning repository. The remaining 66 subjects (10%) are used for testing. We employed the 10-fold cross-validation method in our study, dividing the dataset of 660 subjects into 10 subsets, each containing 66 subjects. In each iteration of the cross-validation process, 1 subset is designated as the testing set, while the remaining 9 subsets are used as the training set. After training 10 times, 10 different models are obtained and evaluated on their corresponding testing sets. Finally, the performance of each model is averaged as the final performance.

**Data and feature extraction**

For evaluating the proposed algorithm, we utilized raw data sourced from the MIMIC II dataset[12]. This dataset comprises a comprehensive collection of photoplethysmogram (PPG) and electrocardiogram (ECG) signals, along with corresponding blood pressure (BP) measurements, obtained from thousands of patients in intensive care units (ICUs). The inclusion of such diverse and extensive data from real-world ICU settings enhances the robustness and generalizability of our algorithm’s evaluation. This dataset contains PPG and ECG signals with corresponding BP from thousands of patients in intensive care units (ICU). However, ECG and PPG signals are instable, which can be affected by physiological activities, electromagnetic environment, and other factors. Signal preprocessing is necessary in order to extract features from ECG and PPG accurately. In this case, Kachuee et al. have made available preprocessed and clean data derived from the MIMIC II dataset[13]. Noises from all signals were removed using an averaging filter[13]. The dataset, consisting of 943 subjects and containing a total of 12,000 records, is available for access in the UCI machine learning repository. All PPG, ECG, and BP signals are sampled at 125 Hz and only the records longer than 8 minutes are chosen to develop our proposed RNN model. In our work, we used 660 subjects and each subject has one record. DBP and SBP were obtained from the minimum and maximum values in each heartbeat of each subject. Figure 2 illustrates the distribution of diastolic blood pressure (DBP) and systolic blood pressure (SBP). The dataset of 660 subjects was divided into four groups based on the office blood pressure threshold, as detailed in Table 1. Within our dataset, the standard deviation of SBP is 19.47 mmHg, while the standard deviation of DBP is 9.64 mmHg.

In order to confine the values of all samples within the range of (0, 1) for each record, it is essential to perform data normalization[14]. The equation indicated as (1) presents the expression used for normalizing
Table 1. Office blood pressure threshold

<table>
<thead>
<tr>
<th>Type (percentage)</th>
<th>SBP (mmHg)</th>
<th>DBP (mmHg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal (36%)</td>
<td>120</td>
<td>&lt; 80</td>
</tr>
<tr>
<td>Prehypertension (35%)</td>
<td>120-139</td>
<td>80-89</td>
</tr>
<tr>
<td>Stage 1 hypertension (21%)</td>
<td>140-159</td>
<td>90-99</td>
</tr>
<tr>
<td>Stage 2 hypertension (8%)</td>
<td>≥ 160</td>
<td>≥ 100</td>
</tr>
</tbody>
</table>

In our work, six types of features with a total number of 28 were extracted from every ECG and PPG beat, as shown in Table 2. In order to extract these features, some fiducial points of each ECG and PPG beat are located, such as R-peaks of ECG beats, pulse foot, systolic peaks, dicrotic notches, and diastolic peaks of PPG beats. Figure 3 displays the schematic diagram illustrating the fiducial points and various features. The

\[
\tilde{x}(t) = \frac{x(t) - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}}
\]

(1)
### Table 2. Feature definitions

<table>
<thead>
<tr>
<th>Feature types</th>
<th>Features</th>
<th>Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time indices</td>
<td>$PTT_{pp}$</td>
<td>Pulse transit time: from R-peak of ECG to systolic peak of PPG</td>
</tr>
<tr>
<td></td>
<td>$PTT_{pf}$</td>
<td>Pulse transit time: from R-peak of ECG to pulse foot of PPG</td>
</tr>
<tr>
<td></td>
<td>$t_{sf}$</td>
<td>Time span from systolic peak to pulse foot</td>
</tr>
<tr>
<td></td>
<td>$t_{nf}$</td>
<td>Time span from dicrotic notch to pulse foot</td>
</tr>
<tr>
<td></td>
<td>$t_{ds}$</td>
<td>Time span from diastolic peak to systolic peak</td>
</tr>
<tr>
<td></td>
<td>$t_{dn}$</td>
<td>Time span from the next pulse foot to dicrotic notch</td>
</tr>
<tr>
<td></td>
<td>$t_{nf}$</td>
<td>Time span from the next pulse foot to pulse foot</td>
</tr>
<tr>
<td>Slope indices</td>
<td>$S_{nf}$</td>
<td>Slope from dicrotic notch to pulse foot</td>
</tr>
<tr>
<td></td>
<td>$S_{sf}$</td>
<td>Slope from systolic peak to pulse foot</td>
</tr>
<tr>
<td></td>
<td>$S_{sn}$</td>
<td>Slope from systolic peak to diastolic peak</td>
</tr>
<tr>
<td>Area indices</td>
<td>$A_{sf}$</td>
<td>PPG area from systolic peak to pulse foot</td>
</tr>
<tr>
<td></td>
<td>$A_{ns}$</td>
<td>PPG area from dicrotic notch to systolic peak</td>
</tr>
<tr>
<td></td>
<td>$A_{dn}$</td>
<td>PPG area from diastolic peak to dicrotic notch</td>
</tr>
<tr>
<td></td>
<td>$A_{fs}$</td>
<td>PPG area from the next pulse foot to systolic peak</td>
</tr>
<tr>
<td>Intensity indices</td>
<td>$RI_{sf}$</td>
<td>PPG ratio from systolic peak to diastolic peak</td>
</tr>
<tr>
<td></td>
<td>$RI_{nd}$</td>
<td>PPG ratio from systolic peak to pulse foot</td>
</tr>
<tr>
<td></td>
<td>$RI_{sn}$</td>
<td>PPG ratio from systolic peak to dicrotic notch</td>
</tr>
<tr>
<td></td>
<td>$RI_{dn}$</td>
<td>PPG ratio from diastolic peak to dicrotic notch</td>
</tr>
<tr>
<td>Phase space indices</td>
<td>$B_{al}$</td>
<td>All number of black boxes from phase space diagram of PPG</td>
</tr>
<tr>
<td></td>
<td>$B_{1}/B_{al}$</td>
<td>Degree of complexity in area 1 of phase space diagram</td>
</tr>
<tr>
<td></td>
<td>$B_{2}/B_{al}$</td>
<td>Degree of complexity in area 2 of phase space diagram</td>
</tr>
<tr>
<td></td>
<td>$B_{3}/B_{al}$</td>
<td>Degree of complexity in area 3 of phase space diagram</td>
</tr>
<tr>
<td></td>
<td>$B_{4}/B_{al}$</td>
<td>Degree of complexity in area 4 of phase space diagram</td>
</tr>
<tr>
<td></td>
<td>$B_{5}/B_{al}$</td>
<td>Degree of complexity in area 5 of phase space diagram</td>
</tr>
<tr>
<td></td>
<td>$B_{6}/B_{al}$</td>
<td>Degree of complexity in area 6 of phase space diagram</td>
</tr>
<tr>
<td></td>
<td>$B_{7}/B_{al}$</td>
<td>Degree of complexity in area 7 of phase space diagram</td>
</tr>
<tr>
<td></td>
<td>$B_{8}/B_{al}$</td>
<td>Degree of complexity in area 8 of phase space diagram</td>
</tr>
<tr>
<td>Other</td>
<td>$HR$</td>
<td>Heart rate (R-R interval of ECG)</td>
</tr>
</tbody>
</table>

R-peaks of the ECG beats are extracted using the hierarchical clustering method\cite{15} and all fiducial points in PPG beats are extracted based on the wavelet transform method\cite{16}. In terms of phase space features of PPG signals, phase space reconstruction (PSR) and box-counting techniques are used\cite{14}. **Figure 4** provides an example illustrating how the phase-space diagram is divided into $N \times N$ pixels, where $N$ represents an integer value.
Figure 3. Some features of ECG and PPG signals.

As an illustration, the phase-space diagram of a single PPG beat is plotted using a time delay of 20 milliseconds and $N$ is set to 20. The phase-space diagram consists of pixels, with certain pixels being considered as black boxes ($B_{all}$) if they are traversed by one or more trajectories, and the remaining pixels are designated as white boxes ($W_{all}$). The count of black boxes ($B_{all}$) among all the pixels provides an indication of the spread of these trajectories. Subsequently, the phase-space diagram is divided into nine areas. The complexity or fractal dimension of a specific area in the PSR diagram is defined as $d_i = B_i/B_{all}$, $i = 1, 2, ..., 9$. The index $i$ corresponds to a specific area within the diagram. Here, we only measured the degree of complexity in eight areas because there are no trajectories in area 9.

**Feature selection**

The efficiency of the model will be affected when a large number of features are used, because some features might be related to others, leading to redundant computational costs. Therefore, feature selection is used to construct the model by selecting the most relevant features among all features. It can reduce the dimensionality of input data of the model\textsuperscript{[17]}. The objective in this case is to minimize the computational cost of the model while enhancing its performance through the incorporation of the most relevant features. In our work, five feature selection methods will be used, which are named ReliefF feature selection (ReliefF), correlation-based feature selection (CFS), mutual information method (MI), neighborhood component analysis (NCA), and relativity analysis (RA).
ReliefF feature selection
The original Relief is a feature selection algorithm based on binary classification problems\textsuperscript{[18]}. For multi-class problems, the improved method ReliefF will be used, which combines Relief and K-nearest Neighbor\textsuperscript{[19]}. ReliefF algorithm finds the $k$ nearest hits (H) from the same class as $X$, as well as the $k$ nearest misses (M) from each distinct class.

Correlation-based feature selection
Correlation is a statistical method utilized to assess the strength of the relationship between a set of features\textsuperscript{[20]}. A feature is considered valuable if it contributes significantly to the class prediction without duplicating the information provided by other features\textsuperscript{[21]}. Here, correlation coefficient probability ($P$-value) will be used to select a significant feature set from the hold set of features. The selected feature set will exhibit a high correlation with the class while demonstrating little to no correlation among the features themselves.

Mutual information method
Mutual information is a measure of how informative a random variable is about another variable. This method is very useful in feature selection because it can help quantify the correlation between a feature and the output value\textsuperscript{[22]}. A high MI score shows a high dependency between the feature and the output value.

Neighborhood component analysis
Neighborhood component analysis (NCA) is a non-parametric and embedded feature selection method that belongs to supervised learning\textsuperscript{[23]}. Its goal is to maximize the prediction accuracy of classification algorithms. The advantage of NCA is that it provides information on significant features while giving the ranking of the features\textsuperscript{[24]}.

Relativity analysis
To analyze the relationship between features, Pearson correlation analysis will be utilized\textsuperscript{[20]}. Using the Pearson correlation coefficient $\rho_{X,Y}$ to represent the relevance of the features, which is calculated using the equation provided below:

$$
\rho_{X,Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X\sigma_Y}
$$

where $E[(X - \mu_X)(Y - \mu_Y)]$ is the covariance of $X$, $Y$ and $\sigma_X, \sigma_Y$ are the standard deviation of $X$ and $Y$. A high coefficient shows that the two features are highly related. In our work, we conducted a feature analysis on a total of 28 extracted features, and we will remove highly correlated features based on the Pearson correlation coefficient.

Vote system
To address the limitations of feature selection algorithms, a voting system is implemented, as illustrated in Figure 5. First, for each specific algorithm, the importance of each feature was given, which is called the weight of features and features were sorted based on the weights at the same time. Then, for each algorithm, the top $N$ features were assigned one vote and the final voting score sorted to select features. Here, $N$ represents the number of features that are selected to build the model. Finally, features with high correlation will be removed based on the Pearson correlation coefficient.
Deep recurrent neural network

The proposed deep recurrent neural network (RNN) model, as depicted in Figure 6, provides an overview of its architecture. It consists of a Bi-LSTM at the bottom layer, which is followed by a stack of five LSTM layers with residual connections. Furthermore, at the top two layers, there is a Bi-LSTM layer and a LSTM layer.

In general, RNNs with deeper architecture tend to outperform shallow RNNs by a significant margin\cite{25,26}. According to the experimental findings reported in \cite{26}, it was observed that the utilization of multiple layers of RNN led to a notable enhancement in expressive capacity. Nevertheless, employing a fully deep network can result in an escalation of computational complexity. Moreover, deep RNNs can be challenging to train due to issues such as gradient exploding and vanishing. To overcome these problems while preserving the benefits of recurrent connections, specific strategies are employed. An independently RNN (Ind-RNN) model was proposed\cite{26}. This model effectively addresses the problems of gradient vanishing and exploding by regulating the gradient backpropagation through time.

An independently RNN (Ind-RNN) model was proposed\cite{26}. It can regulate the gradient backpropagation through time to effectively address the gradient vanishing and exploding problems. The degradation problem was effectively addressed in\cite{27} through the introduction of a deep residual learning framework. The experimental results indicated that their proposed deep residual nets achieved significant accuracy improvements as a result of the substantial increase in depth. Inspired by the deep residual learning framework in\cite{27}, we introduce a residual connection block designed to enhance the propagation of gradients within deep RNNs, specifically LSTM-based models. This approach facilitates more efficient gradient propagation in deep RNN architectures. An example of residual connections is given in Figure 7. The residual connections are implemented between \(a\)th LSTM layer and \(a\)th layer (\(a = 1, 2, 3, \ldots, n\)). In this case, \(x_t^a\) and \(h_t^a\) are the input and hidden state of the \(a\)th LSTM layer at \(t\)th moment, respectively. \(\Phi\) represents the LSTM function. The input of the LSTM layer at \(t\)th moment is then calculated by adding the input and hidden state of the \(a\)th LSTM layer (\(x_t^a\) and \(h_t^a\)). The integration of residual connections within the LSTM block can be accomplished by utilizing the following equations:
Finally, the proposed deep RNN model can be developed by sequentially stacking multiple LSTM blocks of this type. Once the hidden state of the top layer \((a = n)\) of LSTM is calculated, the final output \(y_t\) at the \(t\)th moment can be computed by equation (19):

\[
y_t = \sigma(W^n \cdot [h^n_t, x^n_t] + b^n)
\]

In our work, the depth of the teacher model was set as 8. The experimental results demonstrated that when the depth of the deep RNN model exceeds 8, it can lead to the issue of overfitting. The reason for this phenomenon might be the limited training samples of VA classification. In addition, the sequence length of each training sample is initially set to 32, but it can be flexibly adjusted to higher as needed when employing a deeper RNN model. The hidden state size in the deep RNN model is 128 at each layer.

RESULT

Feature selection

The proposed feature voting system was introduced in Section “METHODOLOGY” in order to select the optimal feature sets for BP estimation. The Pearson correlation coefficient was calculated based on 28 features, setting \(\rho_{x,y} < -0.8\) or \(\rho_{x,y} > 0.8\) as strong correlation. Table 3 shows the features with high correlation among 28 features. Finally, 11 different feature sets with the number of features from 7 to 17 are determined in our work, which are shown in Table 4.

Model performance

This section presents the outcomes of our proposed deep RNN model tested on 660 subjects based on different feature sets using 10-fold validation. All features are derived from ECG and PPG signals and each feature set is determined using a vote system of five feature selection methods.

To facilitate comparisons between our proposed model and those found in the literature, we evaluate the performance of blood pressure (BP) estimation using the following equations, which measure estimation accuracy through root mean square error (RMSE) and mean absolute error (MAE).

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{7}
\]

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{8}
\]

where \(y_i\) is the ground truth of BP and \(\hat{y}_i\) is the BP estimated by the proposed model. Table 5 displays the performance of the proposed model based on different feature sets. MAE and RMSE of SBP, DBP, and MAP are the estimation accuracy metrics. For BP estimating, the best result of the overall performance is achieved using a 14-feature set. The values of RMSE for predicted SBP and DBP are 3.223 and 1.781, respectively, while the MAE values for SBP and DBP estimation are 2.514 and 1.383, respectively. The statistical hypothesis testing is applied on all feature sets via a feature voting system, and a significance level of 0.05
Table 3. Strongly correlated features and corresponding Pearson correlation coefficients

<table>
<thead>
<tr>
<th></th>
<th>PTT$_{pp}$</th>
<th>PTT$_{pf}$</th>
<th>t$_{ft}$</th>
<th>HR</th>
<th>A$_{ns}$</th>
<th>B$_{all}$</th>
<th>B$_{1/2all}$</th>
<th>B$_{4/2all}$</th>
<th>B$_{5/2all}$</th>
<th>B$_{6/2all}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTT$_{pp}$</td>
<td>0.99</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t$_{fn}$</td>
<td>0.94</td>
<td>-0.87</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t$_{ff}$</td>
<td>1</td>
<td>-0.94</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A$_{ns}$</td>
<td>0.89</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A$_{fs}$</td>
<td>0.86</td>
<td>0.94</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B$_{all}$</td>
<td>-0.92</td>
<td>-0.88</td>
<td>-0.87</td>
<td>-0.86</td>
<td>-0.87</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B$_{1/2all}$</td>
<td>1</td>
<td>0.98</td>
<td>0.96</td>
<td>0.81</td>
<td>0.93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B$_{4/2all}$</td>
<td>1</td>
<td>0.94</td>
<td>0.91</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B$_{5/2all}$</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B$_{6/2all}$</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.87</td>
</tr>
</tbody>
</table>

was used. The results are shown in Table 5, where * indicates that the 14-feature set is significantly better than its competitor. Figure 8 shows the performance of DBP and SBP estimation using a 14-feature set. In general, both DBP and SBP are performed in close agreement. We also compared the results with and without containing PSR features. The results show that for both DBP, SBP and MBP, both RMSE and MAE for 14 features with PSR are smaller than those of 14 features without PSR.

The Bland-Altman graph (BA graph) is widely used in biomedical research to compare the agreement between two measurements\cite{28}. It shows the agreement between the prediction BP and the reference BP by using the mean difference BP and the average BP\cite{4}. In Figure 9, the Bland-Altman (BA) graph displays the average value of the reference BP and estimated BP on the x-axis, while the y-axis represents the difference between them. The bias (Mean) and the limits of agreement (Mean ± 1.96 × SD) of our work are shown as dashed line. The mean difference of DBP and SBP are -0.13 and -0.59, respectively. It is shown that for both DBP and SBP, the majority of data points fall within the limits of agreement, signifying a close agreement between the estimated BP and the reference BP. Additionally, Figure 10 illustrates correlation graphs for DBP and SBP. The x-axis represents the ground truth BP, while the y-axis represents the predicted BP. The red dashed line indicates when the ground truth and predicted BP are equal. The green line signifies the linear regression relationship between ground truth BP and predicted BP. On the test data, the correlation coefficients are 0.634 for DBP and 0.888 for SBP.

DISCUSSION

In this paper, we introduced a beat-to-beat blood pressure (BP) monitoring method that integrates feature selection techniques with deep recurrent neural networks. The method mainly includes signal preprocessing, feature extraction, and feature selection. The feature voting system is used to obtain the optimal feature sets with different numbers of features based on the main votes of five feature selection algorithms. In addition, the RNN model based on Bi-LSTM and LSTM was developed to estimate BP. In this case, the proposed model was validated using 66 subjects, which is 10% of the total dataset. The combination of
feature voting system and the deep RNN model achieved good results on both MAE and RMSE between BP ground truth and BP prediction. Table 6 offers a comparison of our proposed method with contemporary approaches based on the same dataset from UCI. Our proposed method gave an average RMSE of 1.781, 2.062 and 3.223 for DBP, MAP and SBP, and gave an average MAE of 1.383, 1.618 and 2.514 for DBP, MAP and SBP, respectively. Overall, our model’s performance is on par with or superior to other methods. Qiu et al. proposed a neural network with squeeze and excitation (SE) block for BP estimation with the MAE of 3.7 and 2.81 for SBP and DBP\cite{29}. El-Hajj et al. suggested the use of BI-GRU and attention mechanism to estimate BP based on 22 features and the MAE is 2.58 and 1.26 for SBP and DBP, respectively\cite{30}. Although this method outperformed ours in terms of DBP MAE, it considered only segments with good quality. It was mentioned that manual checks were conducted to include only high-quality segments for analysis in order to reduce errors during feature extraction. Thus, it may cause an overly optimistic performance for BP estimation. A simple PTT-based model was also introduced in\cite{19} to estimate BP. Zhang et al. represented a fewer parameters BP estimation model based on RA-ReliefF feature selection and the MPGAA-BPN model, but the performance is poorer than others\cite{31}. In addition, Liu et al. integrated several probabilistic non-parametric methods for BP estimation and Gaussian process regression (GPR) and gave the best results with an MAE of 4.8 and 3.4 for SBP and DBP, respectively\cite{31}. They also proposed a CNN-LSTM model for BP estimation using ECG and PPG segment series, resulting in an MAE of 3.97 and 2.1 for SBP and DBP, respectively\cite{31}. In this work, CNN is applied to extract features automatically, and the LSTM model is used to estimate BP. The results reflect that our voting system may have higher performance than CNN for feature selection. Collectively, the results demonstrate that our
Table 6. Performance comparison with related works for BP estimation

<table>
<thead>
<tr>
<th>Authors</th>
<th>Method</th>
<th>RMSE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>BP</td>
<td>DBP</td>
</tr>
<tr>
<td>Qiu, et al. (2021)</td>
<td>Composite neural network with SE block</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>El-hajj and Kyriacou P</td>
<td>Bi-GRU+GRU+attention</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ganti, et al. (2021)</td>
<td>PTT based</td>
<td>4.75</td>
<td>2.72</td>
</tr>
<tr>
<td>Zhang, et al. (2023)</td>
<td>MPG-A-BPN</td>
<td>8.23</td>
<td>8.92</td>
</tr>
<tr>
<td>Liu, et al. (2021)</td>
<td>GPR</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Esmaeipour, et al. (2020)</td>
<td>CNN+LSTM</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Our current work</td>
<td>Proposed model</td>
<td>3.223</td>
<td>1.781</td>
</tr>
</tbody>
</table>

The proposed deep RNN architecture. Each circle represents a LSTM cell. Each red dashed box is a Bi-LSTM layer, which consists of forward (green) and backward (blue) LSTM. The blue dashed box represents the LSTM layer with residual connections.

Proposed method is on par with or outperforms recent existing methods. Hence, our proposed method holds promise for real-time blood pressure monitoring.

Table 7 presents an evaluation of the proposed method by BHS standard. BHS grades the BP estimation into Grades A, B, and C based on the percentage of cumulative errors MAE less than the three threshold values, i.e., 5, 10, and 15 mmHg, respectively. Our proposed method achieved Grade A for DBP, MAP, and SBP estimation according to the BHS standard.
Table 7. Comparison with the BHS standard

<table>
<thead>
<tr>
<th></th>
<th>Cumulative error percentage</th>
<th>≤ 5 mmHg</th>
<th>≤ 10 mmHg</th>
<th>≤ 15 mmHg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our results</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SBP</td>
<td>89.3%</td>
<td>100%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>MAP</td>
<td>95.5%</td>
<td>100%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>DBP</td>
<td>96.9%</td>
<td>100%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>Grade A</td>
<td>60%</td>
<td>85%</td>
<td>95%</td>
<td></td>
</tr>
<tr>
<td>Grade B</td>
<td>50%</td>
<td>75%</td>
<td>90%</td>
<td></td>
</tr>
<tr>
<td>Grade C</td>
<td>40%</td>
<td>65%</td>
<td>85%</td>
<td></td>
</tr>
<tr>
<td>BHS[32]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SBP</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAP</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DBP</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 7. An example of LSTM with residual connections.

Figure 8. Comparison of the ground truth and prediction of a representative subject.
In conclusion, we introduced a deep RNN model with an optimal feature set of ECG and PPG signals to achieve continuous and noninvasive BP estimation. Primarily, the proposed method comprises signal preprocessing, feature extraction, feature selection, and RNN model stages. The results indicate that the performance of this method is comparable or higher than other related works. In addition, our proposed blood pressure estimation method was validated on the BHS standard. The estimation of DBP, MAP, and SBP achieved BHS standard Grade. However, the results may not provide a comprehensive view of our method’s overall performance since the testing was conducted on a limited dataset of 66 subjects. Our future work will involve testing our method on larger and more diverse datasets. Moreover, some deep learning models have the ability to select features automatically; the comparison between our feature voting system and such deep learning models should be further analyzed. Finally, it is possible to integrate our proposed method to develop a healthcare monitoring system to improve patients’ life quality.

DECLARATIONS

Authors’ contributions
Conceptualization, data collection, methodology, software, validation, writing - original draft: Chen H
Conceptualization, methodology, software, validation, writing - original draft: Lyu L
Methodology, software: Zeng Z
Review, statistical analysis: Jin Y
Conceptualization, supervision, project administration: Zhang Y
Availability of data and materials
All patient records and data were extracted from the University of California Irvine (UCI) machine learning repository. The data used in this work are uploaded supplementary material.

Financial support and sponsorship
This work was supported fully by InnoHK Project at Hong Kong Centre for Cerebro-Cardiovascular Health Engineering (COCHE).

Conflicts of interest
All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Copyright
© The Author(s) 2023.

REFERENCES
Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No. 03CH37439); 2003 Sep 17-21; Cancun, Mexico. DOI


