Computers and Electrical Engineering 118 (2024) 109319

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

An improved deep regression model with state space reconstruction for continuous blood pressure estimation ☆

Liangyi Lyu [a], Lei Lu [b,d], Hanjie Chen [a,*], David A. Clifton [b], Yuanting Zhang [c], Tapabrata Chakraborti [e]

[a] Hong Kong Centre for Cerebro-Cardiovascular Health Engineering (COCHE), Hong Kong
[b] Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, United Kingdom
[c] Department of Electronic Engineering at the Chinese University of Hong Kong
[d] School of Life Course and Population Sciences, King's College London, United Kingdom
[e] The Alan Turing Institute and University College London, UK

ARTICLE INFO

Keywords:
Continuous blood pressure estimation
Recurrent neural network
Feature selection
State space reconstruction

ABSTRACT

Continuous blood pressure (BP) monitoring is crucial for diagnosing and preventing cardiovascular disease (CVD). However, existing approaches for continuous cuffless BP monitoring using photoplethysmogram (PPG) and electrocardiogram (ECG) signals suffer from instability and susceptibility to various factors. This poses a challenge in accurately estimating BP levels, hindering effective disease diagnosis and prevention. Therefore, there is a need for an improved method that can provide reliable and accurate continuous BP estimation from PPG and ECG signals, overcoming the limitations of existing approaches. In this study, we proposed a deep regression model with state space reconstruction (SSR) for continuous BP estimation. A feature voting system with a variety of feature selection algorithms is introduced to select the optimal feature set of PPG and ECG signals. The SSR technique is applied to feature data to reveal useful hidden information. The proposed method is evaluated based on 660 subjects from a well-known benchmark dataset and a multi-day BP dataset. Random forest and the proposed deep regression model are tested to show the advantages of using SSR on feature data. The results show a promising performance of the improved deep regression model. On the benchmark dataset, the root mean square error (RMSE) and mean absolute error (MAE) of the improved deep regression model are 3.613 and 2.765 mmHg respectively for systolic BP (SBP), 1.978 and 1.543 mmHg for diastolic BP (DBP). The results achieved high accuracy for estimating SBP and DBP according to the British Hypertension Society (BHS) standard. On the multi-day BP dataset, the proposed model achieved RMSE of 5.387, 3.338 and 3.611 mmHg for SBP and achieved MAE of 4.115, 2.553, and 2.927 mmHg for DBP. Additionally, the robustness of our proposed model is validated by adding random noise into the PPG signals. The results demonstrate that the proposed deep regression model with SSR can improve the performance of BP estimation. It is possible to apply our proposed method further to develop a wearable device for real-time BP monitoring.

☆ This research is supported by InnoHK and the Hong Kong Centre for Cerebro-Cardiovascular Health Engineering (COCHE).
* Corresponding author.
E-mail addresses: hjchen@hkcoche.org (H. Chen), ytzhang@cuhk.edu.hk (Y. Zhang), tchakraborty@turing.ac.uk, t.chakraborty@ucl.ac.uk (T. Chakraborti).

https://doi.org/10.1016/j.compeleceng.2024.109319
Received 14 December 2023; Received in revised form 14 May 2024; Accepted 15 May 2024
Available online 28 May 2024

1. Introduction

Hypertension or high blood pressure (BP) as one of the major causes of cardiovascular diseases, which is often asymptomatic in the early stage and thus the diagnosis rate is low [1]. Continuous BP monitoring can help improve the awareness and early prevention of hypertension [2]. For hypertensive patients who need to take BP frequently, inflatable cuffs are uncomfortable and inconvenient to measure non-invasive BP. Therefore, non-invasive and continuous methods for BP estimation attracted extensive attention in recent years.

There are two main methods used for BP measurement: the invasive method and the noninvasive method. The invasive method is considered the gold standard in BP measurement [3]. Although the invasive method can be accurate in obtaining BP, it can be used only in clinical environments. Thus, it is inconvenient for daily BP monitoring. For the noninvasive cuffless BP estimation, there are two main aspects of blood pressure estimation: Pulse Transit Time based (PTT-based) method and waveform-based analysis [4]. Several studies have found that Pulse Wave Velocity (PWV) and Photoplethysmogram Intensity Ratio (PIR) are two of the important parameters for estimating BP [5]. In some cases, PWV is usually expressed accordingly by measuring Pulse Transit Time (PTT) or Pulse Arrival Time (PAT) [6]. Photoplethysmography (PPG) as a low-cost measurement method, utilizes an infrared light to measure the volumetric variations of blood circulation [7]. The wristband-type PPG monitoring devices are used as the most popular device because they are convenient to wear by users, inexpensive, and highly portable. Besides, different features including PPT and PAT can be extracted from PPG signals, which are used for BP estimation. BP is correlated with PTT based on the Moens–Korteweg (M–K) equation [8] and Hughes equation [9]. PTT is the time required for an arterial pulse to propagate from the heart to a peripheral site, and it can be calculated as the time interval between the peak R wave of the electrocardiogram (ECG) and a characteristic point of the PPG [8]. Based on the M–K equation and Hughes equation, a mathematical relationship between BP and PTT can be established, which is shown in Eq. (1).

$$P = \frac{1}{\gamma}\left[2\ln\left(\frac{D}{PTT}\right) + \ln\left(\frac{2\rho d}{E_0 h}\right)\right] \tag{1}$$

PIR is the difference between PPG peak and valley in one cardiac cycle, it reflects the change of arterial diameter in one cardiac cycle. From the previous study, the change in arterial diameter has an important effect on peripheral resistance and blood volume, which are two main factors that affect BP [8]. The relationship between PIR and BP is shown in Eq. (2) and Eq. (3).

$$DBP = DBP_0 \cdot \frac{PIR_0}{PIR} \tag{2}$$

$$SBP = DBP_0 \cdot \frac{PIR_0}{PIR} + PP_0 \cdot \left(\frac{PTT_0}{PTT}\right)^2 \tag{3}$$

where the DBP_0 and SBP_0 refer to the initial calibration constant values of the measurement for diastolic BP (DBP) and systolic BP (SBP) respectively.

Many studies used PTT and PIR as the input feature in mathematical or regression models for BP estimation [3,8,10]. However, the M–K equation involves two assumptions and the Hughes equation has no theoretical basis [11]. Therefore, relying exclusively on PTT to capture the inherent fluctuation of blood pressure is insufficient, as it can compromise the prediction accuracy and reliability of the model. Recent studies have tried to extract more features and use more complex models for BP estimation.

The limited accuracy observed in current cuffless blood pressure (BP) measurement methods can be primarily attributed to two sources of uncertainty: epistemic uncertainty, which pertains to the inherent limitations and assumptions of the employed models, and aleatoric uncertainty, which arises from the inherent variability and unpredictability present in the collected BP data. These uncertainties significantly impact the accuracy and reliability of BP estimation [12]. BP estimation is studied in two steps, feature extraction and model construction. Current BP estimation models can be divided into two categories based on the temporal dependencies: point-to-point BP estimation model and sequence-to-point BP estimation model [13]. For continuous cuffless BP estimation, a point-to-point model will be employed. Point-to-point models are commonly used in the field of BP estimation, where individual data points from the input signal are translated to estimate corresponding BP values. A model based on Gaussian process regression (GPR), which is a nonparametric kernel-based probabilistic model was proposed in [2]. In [14], several regression methods based on 20 extracted features were applied to estimate BP, including linear regression, ridge regression, support vector machines (SVM), AdaBoost, and random forest (RF). Given the inherent temporal nature of BP data, the prediction of BP can be considered as a study of predicting time-series data. Consequently, there is a growing interest in the research community towards exploring time-series related models for non-invasive continuous blood pressure estimation. This emerging focus on time-series models reflects the recognition of the importance of capturing the dynamic patterns and temporal dependencies present in BP data. By incorporating advanced time-series modeling techniques, researchers aim to enhance the accuracy and effectiveness of BP prediction methods. The focus of current research in non-invasive continuous blood pressure estimation has shifted towards feedback neural networks. These neural networks exhibit interconnected hidden layers, where the inputs to the hidden layers consist of both the outputs from the input layer and the outputs from the hidden layers at the previous time step. This unique architecture of feedback neural networks provides a strong foundation for effectively processing time-series data. An LSTM model that helps to learn the time dependency in BP sequences was proposed in [15].

All machine learning and deep learning methods mentioned above are based on feature extraction. There are mainly two ways for feature extraction: the first one is manually extracting features from ECG and PPG signals [15], and the second one is using deep neural networks to extract features automatically [16]. Both approaches have their advantages and disadvantages. Manually

extracting features helps us determine exactly which features are used for the models. However, it is difficult to determine whether all extracted features are useful for BP estimation. Automatically extracting features can learn information that cannot be extracted manually, while since the features are extracted by DNN, it is difficult to interpret them physically [17]. The accuracy of manually extracting features has a great impact on the accuracy of BP estimation models. More features mean that more information will be learned by the models thus it is possible to improve the model accuracy. If the ECG and PPG information is tried to be extracted as much as possible, then many features will be input into the model. In this case, the complexity of the model is increased, leading to a high computational cost. On the other hand, a smaller number of features may not be able to provide sufficient information for BP estimation.

To overcome the issue of how many features should be used for model fitting, feature selection is presented to find the optimal feature set of PPG and ECG signals for BP estimation. In this paper, we proposed a novel feature voting system consisting of 8 feature selection algorithms to select the optimal feature set. After sorting the feature importance in each feature selection algorithm, 8 results were combined and got the final feature set, which is defined as the optimal feature set. Subsequently, a deep recurrent neural network (RNN), which contains 2-layered bidirectional long short-term memory (Bi-LSTM) and 6-layered LSTM networks was built to estimate BP based on the optimal feature set.

Additionally, one-dimensional time series data may hide some information in high-dimensional space, which may cause randomness [18]. In this case, state or phase space reconstruction has been widely used to preprocess the raw data to extract more hidden information. In this study, the selected feature set through state space reconstruction (SSR) was used as the input. The results of the models with the original feature set and SSR feature set will be compared. The results show that the hidden information extracted through SSR can improve the prediction performance of models. When we employ SSR techniques to describe high-dimensional information, two components are necessary: the time delay and embedding dimension. Details on how to obtain these two components are described in Section 2.5. In addition, to compare the effect of SSR on the results more comprehensively, a random forest model was also developed to compare the prediction performance on the regression model.

The contributions of this article are summarized as follows:

(1) A new structured feature voting system is developed to identify the most informative feature set from ECG and PPG signals.
(2) The SSR technique is proposed to extract hidden information from ECG and PPG signals, enhancing the accuracy and reliability of BP estimation.
(3) Robustness analysis with noise injection is assessed to the proposed model by introducing random noise into the PPG signals.
(4) Extensive comparative studies with twelve deep learning methods using well-known datasets are presented to show the effectiveness of the proposed method.

The remainder of this paper is organized as follows. Section 2 presents the data and data preprocessing block, the feature extraction and selection methods, the SSR technique, as well as the structure of the RNN model. Section 3 shows the results of the random forest regression model and our proposed model. Section 4 presents the model comparison and the robustness analysis. Section 5 presents the limitations of the proposed work. Finally, a conclusion and future work are provided in Section 6.

2. Methodology

In this section, we present the proposed deep regression model with the SSR technique for BP estimation, utilizing an optimized feature set derived from ECG and PPG signals. Our proposed technique is organized as a four-stage procedure, as depicted in Fig. 1. The initial stage entails the extraction of features, encompassing the pre-processing of both ECG and PPG signals, resulting in a total of 19 extracted features. Subsequently, the second step involves feature selection, where the proposed voting system is employed in conjunction with eight distinct feature selection algorithms. Through this process, the optimal feature set is ascertained, comprising a selection of five features. The next step is using the SSR technique on the 5 selected features. The time delay and embedding dimension are determined by the mutual information and false nearest neighbors (FNN) algorithm respectively. Finally, the proposed deep regression model is developed by taking 660 subjects from the University of California Irvine (UCI) machine learning repository randomly. To test and evaluate the performance of the proposed model, 10-fold cross-validation will be employed, ensuring robustness and reliability by partitioning the dataset into ten subsets and iterative using 90% of the data for training, and the remaining 10% data is used for testing.

2.1. Data source

MIMIC-II dataset. In this paper, data from the Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II) dataset [19] is applied to evaluating the SSR technique. This dataset contains thousands of ICU patients' PPG and ECG signals with corresponding BP. The sampling frequency is 125 Hz for ECG, PPG, and BP. Only records that are longer than 8 min are chosen for feature extraction since the performance of continuous BP estimation models should be validated on a long record. Besides, records with very low or very high BP values (e.g., DBP $\leq$ 50 and SBP $\leq$ 80, or, DBP $\geq$ 130 and SBP $\geq$ 180) are also removed, and the final database consists of 660 subjects.

Multi-day BP dataset. The dataset used in this study is a personal dataset consisting of recordings obtained from 12 healthy subjects, comprising 11 males and 1 female [15]. For each subject, the BP, ECG, and PPG data were collected during an 8-minute resting period on multiple days, namely the 1st day, 2nd day, 4th day, and 6 months after the initial recording.

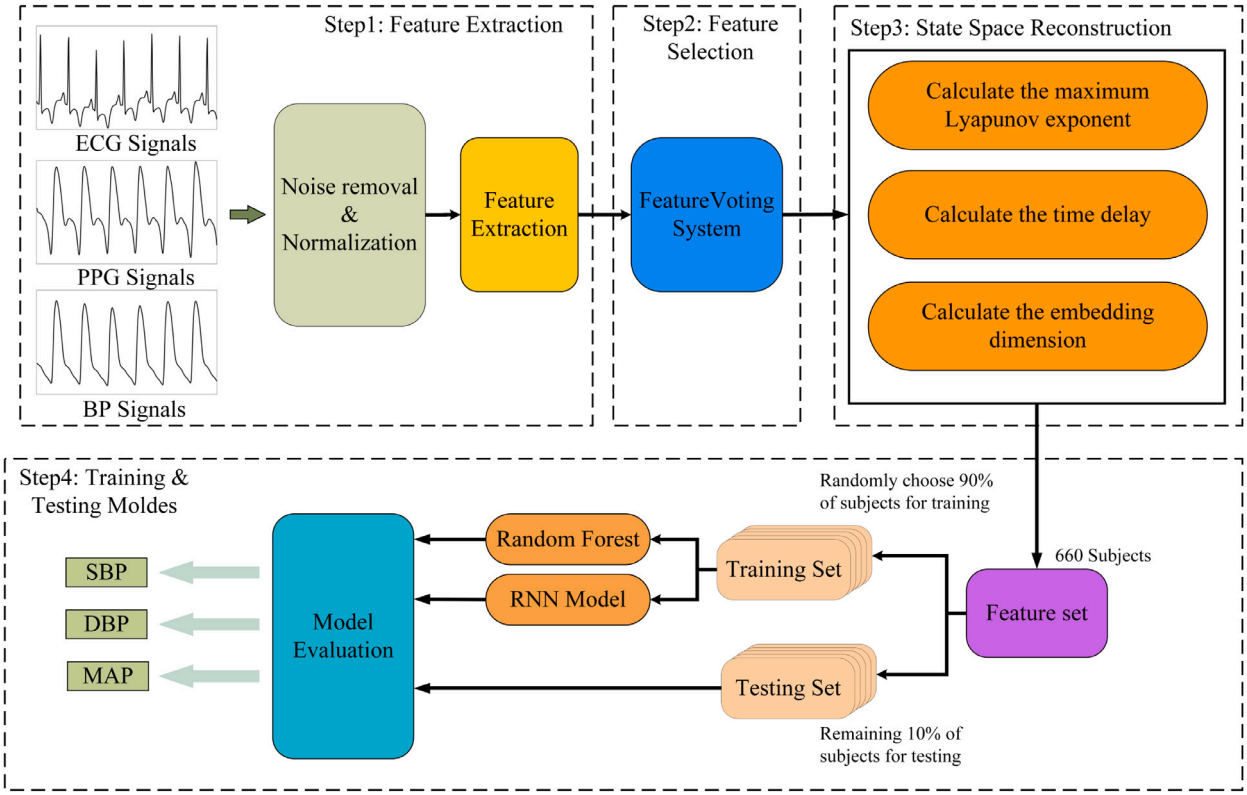

Fig. 1. Flowchart of the proposed continuous BP estimation algorithm.

Table 1
Office blood pressure thresholds.

Type (percentage)	SBP (mmHg)	DBP (mmHg)
Normal (36%)	<120	and <80
Prehypertension (35%)	120–139	or 80–90
Stage 1 Hypertension (21%)	140–159	or 90–99
Stage 2 Hypertension (8%)	≥160	or ≥100

Table 2
Statistical information for MIMIC II dataset.

	Mean	Maximum	Minimum	Standard deviation
DBP	66.43	129	50	9.64
SBP	128.65	185	80	19.47

The random forest (RF) and our proposed deep regression model are used to evaluate the performance of using the SSR technique. A total of 660 subjects obtained from the UCI dataset were used for evaluating the SSR technique and the proposed model. Besides, the Multi-day BP dataset is also used to evaluate the performance of the proposed model. For each subject, DBP and SBP were generated from the minimum and maximum values in each heartbeat. Fig. 2 shows the distribution of DBP and SBP from the MIMIC-II dataset. All 660 subjects are classified into 4 different types of groups based on the office blood pressure threshold, which is shown in Table 1. Besides, Table 2 shows the statistical information for SBP and DBP from the MIMIC-II dataset, and Table 3 shows the statistical information for SBP and DBP from the Multi-day BP dataset.

2.2. Data preprocessing

To mitigate the negative impact of noise and artifacts on the raw ECG and PPG signals, a preprocessing step is implemented to effectively filter and denoise the signals. ECG and PPG signals can be affected by various sources of noise, such as baseline wander,

Table 3
Statistical information for Multi-day BP dataset.

	Mean	Maximum	Minimum	Standard deviation
DBP	64.43	124	48	10.26
SBP	113.19	178	80	14.91

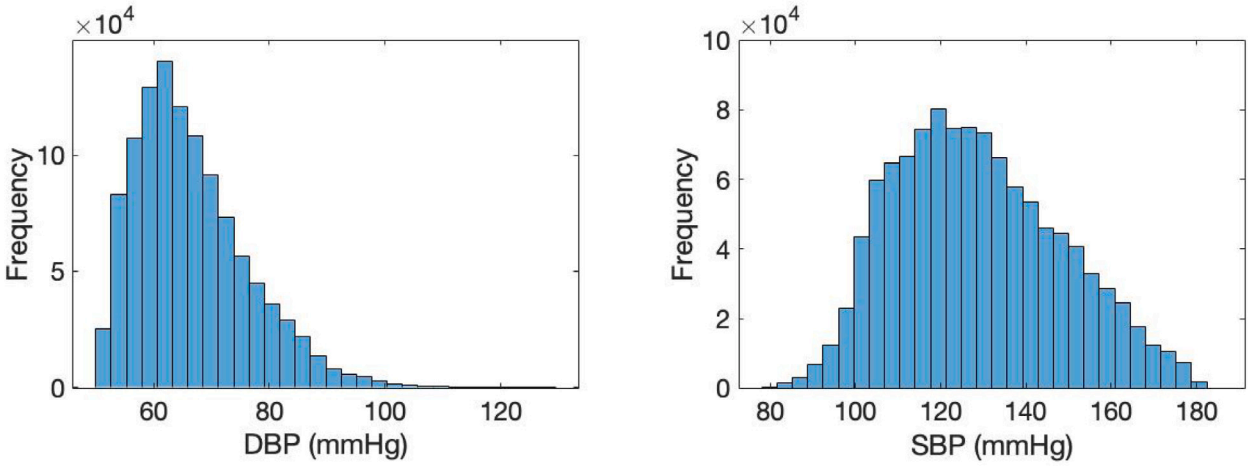

Fig. 2. Distribution of DBP and SBP.

powerline interference, and muscle artifacts [20]. Additionally, these signals may contain beats where the peak cannot be reliably detected due to irregular heartbeats or other factors. Here we filtered and denoise the ECG and PPG signals in MATLAB based on Kachuee et al. [21]. The specific process is as follows: (i) removing the baseline wander noise in ECG and PPG; (ii) using a bandpass filter to do wavelet transform; (iii) removing the beat that cannot detect peak by the peak detection algorithm. By using the filter methods and removing those beats, the preprocessing stage ensures that only high-quality beats with clearly detectable peaks are retained for further analysis. Meanwhile, data normalization is necessary to make sure the values of all samples for each record are in [0, 1]. Here Min–max normalization is used and the function is shown in Eq. (4).

$$x(t)_{scaled} = \frac{x(t) - x_{min}}{x_{max} - x_{min}} \tag{4}$$

2.3. Feature extraction

In our work, 5 types of features with a total of 19 features were extracted based on ECG and PPG signals. The extracted features are shown in Table 4. Some fiducial points of ECG and PPG signals need to be marked in order to extract those features. For example, R-peaks of ECG signals, pulse feet, systolic peaks, dicrotic notches, and diastolic peaks of PPG signals. The schematic diagram of fiducial points and some features are shown in Fig. 3.

2.4. 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. Which leads to overfitting of the model and increases the error rate of the learning algorithm [22]. To avoid the decrease in model prediction performance due to the irrelevance and redundancy of the high-dimensional input data, dimensionality reduction techniques are applied. For dimensionality reduction, one of the most popular techniques is feature selection [23]. In feature selection, relevant features are selected, and irrelevant and redundant features are discarded [24]. A good feature is defined as one that is relevant to the class but not redundant with any other relevant features [25]. In this case, feature selection aims to minimize the computational cost of the model and improve the performance of the model by using the most relevant features. In our work, 8 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), features for classification using minimum redundancy maximum relevance (FSCMRMR), recursive feature elimination (RFE), variance thresholding (VT), and relativity analysis (RA).

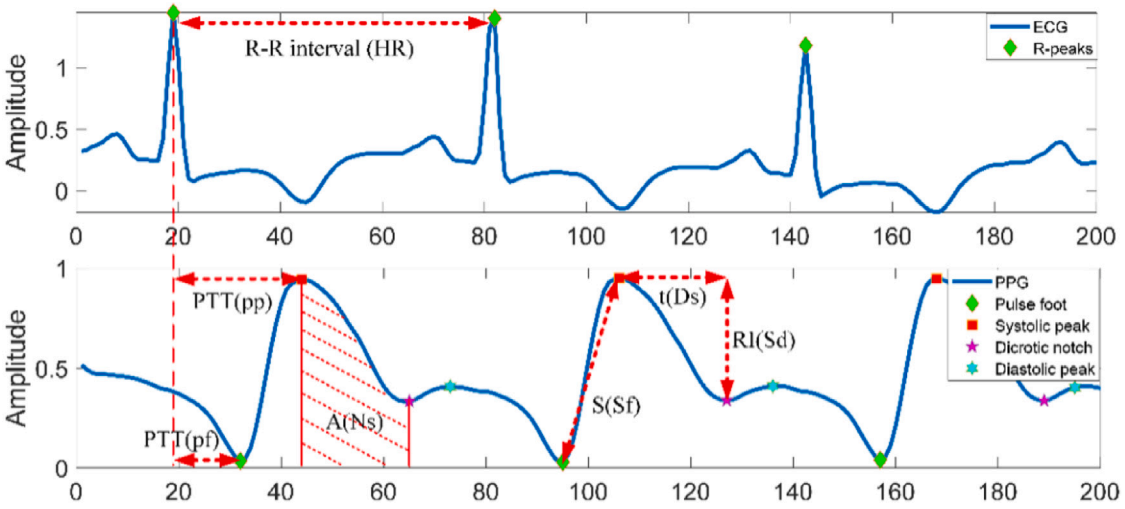

Fig. 3. Features of ECG and PPG signals.

Table 4
Feature definitions.

Feature types	Features	Definitions
Time indices	PTT_{pp}	from R-peak of ECG to systolic peak of PPG
	PTT_{pf}	from R-peak of ECG to pulse foot of PPG
	t_{Sf}	Time span from systolic peak to pulse foot
	t_{Nf}	Time span from dicrotic notch to pulse foot
	t_{Ds}	Time span from diastolic peak to systolic peak
	t_{Fn}	Time span from the next pulse foot to dicrotic notch
	t_{Ff}	Time span from the next pulse foot to pulse foot
Slope indices	S_{Nf}	Slope from dicrotic notch to pulse foot
	S_{Sf}	Slope from systolic peak to pulse foot
	S_{Sd}	Slope from systolic peak to diastolic peak
Area indices	A_{Sf}	Area under the PPG curve from systolic peak to pulse foot
	A_{Ns}	Area under the PPG curve from dicrotic notch to systolic peak
	A_{Dn}	Area under the PPG curve from diastolic peak to dicrotic notch
	A_{Fs}	Area under the PPG curve from the next pulse foot to systolic peak
Intensity indices	RI_{Sd}	Ratio of systolic peak intensity to diastolic peak intensity
	RI_{Sf}	Ratio of systolic peak intensity to pulse foot intensity
	RI_{Sn}	Ratio of systolic peak intensity to dicrotic notch intensity
	RI_{Dn}	Ratio of diastolic peak intensity to dicrotic notch intensity
Other	HR	Heart rate (R–R interval of ECG)

2.4.1. Relieff

The original Relief is a feature selection algorithm based on binary classification problems. The Relief algorithm is an instance-based learning algorithm [26]. Nearest-hit and Nearest-miss of an instance X is calculated and selected by Euclid distance [27]. One limitation of the original Relief algorithm is only two-class data problems could be used. For multi-class problems, the improved method ReliefF will be used, which combines Relief and K-nearest Neighbor [13]. ReliefF algorithm finding k Nearest-hit (H) from the same class of X and k Nearest-miss (M) from each different class. The weight of each feature is updated by Eq. (5).

$$w_i = w_i - \frac{\sum_{k=1}^{K} \text{diff}\left(x^i - H_k^i\right)}{(m^* K)} + \sum_{c \neq c1} \frac{p(c)}{1 - p(c1)} \frac{\sum_{k=1}^{K} \text{diff}\left(x^i - H_{c,k}^i\right)}{(m^* K)} \tag{5}$$

where w_i is the weight of ith feature. $p(c)$ is the prior probability of c and $c1$ is the class of X. k is the number of H and M.

2.4.2. CFS

Correlation-based feature selection (CFS) is one of the feature selection methods in the filter category. This approach sorts and selects feature values by looking at the values that are correlated to the class and have a relation with other features. One advantage of this algorithm is that it requires less computational complexity compared with other approaches [28]. CFS is usually implemented in conjunction with Best First Search (BFS) [29]. Here, correlation coefficient probability (p-value) will be used to select a significant feature set from the hold set of features. The selected set of features is highly correlated with the class but uncorrelated to each other.

2.4.3. MI

Mutual information is one of the commonly used statistical methods [22]. It 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 [30]. A higher MI score indicates a stronger relationship between the feature and the output value, suggesting that the feature is more informative and potentially more relevant for prediction or analysis purposes.

2.4.4. NCA

NCA is a non-parametric and embedded feature selection method that belongs to supervised learning [31]. It is a weighted feature selection technique to select the best subset of the total features by maximizing the objective function, which is the prediction accuracy of classification [32]. It is an advantage of NCA to provide information regarding significant features as well as to rank those features [33].

2.4.5. FSCMRMR

The principle of this method is to find the set of features in the original feature set that has the Max-Relevance to the final output but has Min-Redundancy to each other.

2.4.6. RFE

Recursive Feature Elimination (RFE) is used to select features by recursively considering smaller and smaller sets of features. First, for the initial feature set, calculate the importance of each feature. Then the least important feature will be removed from the current feature set. Repeat the process recursively on the pruned feature set until the desired number of features is finally reached.

2.4.7. Variance thresholding

The variance thresholding algorithm is a very simple way for feature selection. This approach reduces the complexity of the model by eliminating features with very low variance, i.e., features that do not have much useful information. Usually, we need to normalize the features to make sure that all the variances are in the same range. One shortage of this method is that it does not consider the correlation between features and outputs.

2.4.8. Relativity analysis

As a method for analyzing the relationship between features, Pearson correlation analysis will be employed [13]. The Pearson correlation is one of the most common approaches to measuring the statistical relationship or the linear correlation between two variables [34]. Therefore, the Pearson correlation can be used to evaluate the correlation between two features. The formula of the Pearson correlation coefficient is calculated in Eq. (6).

$$\rho_{X,Y} = \frac{E\left[\left(X - \mu_X\right)\left(Y - \mu_Y\right)\right]}{\sigma_X \sigma_Y} \tag{6}$$

where $E\left[\left(X - \mu_X\right)\left(Y - \mu_Y\right)\right]$ is the covariance of X, Y, σ_X and σ_Y are the standard deviation of X, Y, respectively. The absolute value of $\rho_{X,Y}$ is 1 indicates a perfect linear relationship. A coefficient close to 0 indicates that the variables are not linearly related. In our work, the relativity analysis was performed on the total 19 features extracted, and high correlation features will be removed based on the Pearson correlation coefficient.

2.4.9. Feature voting system

A voting system is established to overcome the limitations of the feature selection algorithms which can be seen in Fig. 4. First, for each specific approach, the importance of each feature was given, which is called the weight of features. Then, features are voted and the important features are determined from each specific algorithm. In this study, the top 5 important features of each method were assigned one vote. The final voting score was generated by summing up the vote through 7 feature selection methods. Besides, features with high correlation will be removed based on the Pearson correlation coefficient, the threshold is ± 0.8 here. Finally, the top 5 important features were selected as the optimal feature set and used as the input to the BP estimation model. The procedure of the feature voting system is described in Algorithm 1.

2.5. State space reconstruction

For one-dimensional time series data, especially biomedical signals like ECG and PPG. They are sometimes chaotic, which leads to high-dimensional information being hidden. Due to the random and intermittent characteristics of chaotic time series data, the analysis and prediction may be difficult [18]. According to Taken's theorem [35], the high-dimensional hidden information of one-dimensional time series can be exposed by the time delay τ and embedding dimension d, which are two key parameters in state space reconstruction. Thus the SSR technique was used on feature data to reveal more information. Results were compared with the original feature data to check the improvement of using SSR.

Algorithm 1: Feature voting system

input : $D = \{F_i, BP_i\}_{i=1}^{m}$, ($F_i \in \mathbb{F}^{\mathbb{T}}$, $BP_i \in BP_c$), BP_c = classified reference BP, θ = No. of selecting features, T = No. of features, K = No. of feature selection algorithms

output: Optimal feature set F_{opt}

/* Initialization */

Set $\theta \leftarrow 5$, $F_{opt} \leftarrow \emptyset$, $ListF \leftarrow \emptyset$, $List_v \leftarrow \emptyset$;

/* Calculate the feature weight using each feature selection algorithm */

$fsm \leftarrow$ {ReliefF, CFS, MI, NCA, FSCMRMR, RFE, VT};

for $k \leftarrow 1$ **to** K **do**

  Caculate $W_{fsmk} \leftarrow fsm_k$;

  Sort W_{fsmk};

  $List_F^{fsmk} \leftarrow$ first θ features;

end

/* Calculate voting score for each feature */

for $t \leftarrow 1$ **to** T **do**

  Calculate $V_{score_t} \leftarrow \text{argmax}_{f_t \in F} \sum_{k=1}^{K} ListF_t^{fsm_k}$;

end

Sort the $List_v$ according to V_{score};

/* Remove features with high correlation */

for F_t *in* $List_v$ **do**

  Calculate ρ_{F_i,F_j} based on Equation (8);

  if $\rho_{F_i,F_j} < -0.8$ *or* $\rho_{F_i,F_j} > 0.8$ **then**

    Remove F_t with lower V_{score_t} $List_v \leftarrow List_v(\text{remove}(F_t))$

  end

end

$F_{opt} \leftarrow$ first θ features in $List_v$;

return F_{opt}

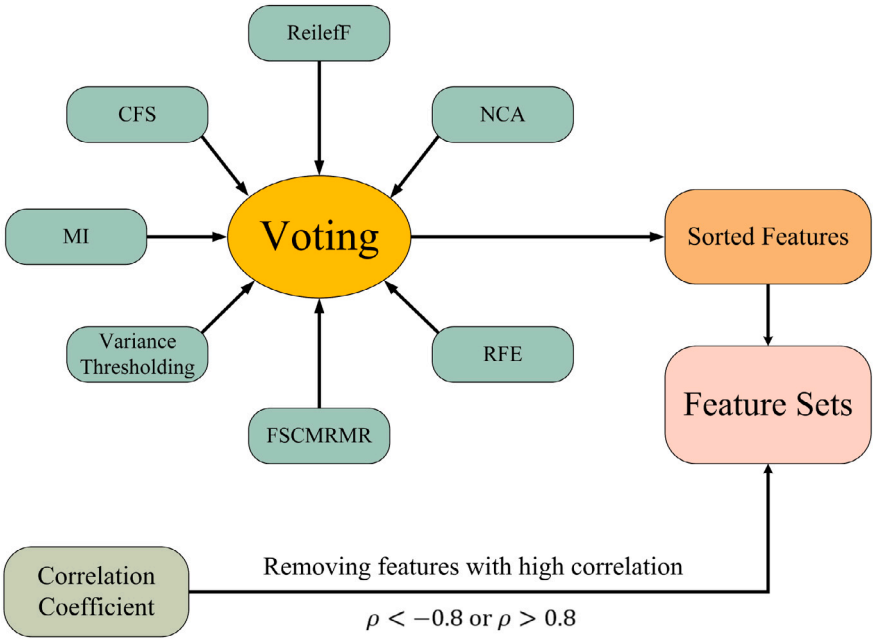

Fig. 4. Flowchart of feature voting system.

Suppose $X = [X_1, X_2, \ldots, X_k]$ is the optimal feature set containing k features and $X_i = [x_1, x_2, \ldots, x_n]$ is the time series data of ith feature. In order to employ the SSR technique, the time delay τ and embedding dimension d must be determined. As a result,

the trajectory matrix is created as shown below.

$$\mathbf{Z} = \begin{bmatrix} x_1 & x_{1+\tau} & \cdots & x_{1+(d-1)\tau} \\ x_2 & x_{2+\tau} & \cdots & x_{2+(d-1)\tau} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n-(d-1)\tau} & x_{n-(d-1)\tau} & \cdots & x_n \end{bmatrix} \tag{7}$$

where n is the length of the input feature data. Rows of $Z = \left[Z_1, \ldots, Z_{n-(d-1)\tau}\right]$ were used to display the trajectory's points in the order in the state space. From the trajectory matrix, it is obvious to find that the key of the SSR technique is to determine the suitable time delay τ and embedding dimension d. According to previous studies, the MI method [36] and the FNN algorithm [37] are employed to find the suitable time delay and embedding dimension, respectively. The SSR technique can only be used when the time series data is chaotic, the maximum Lyapunov exponent (λ_{max}) is used in this study to determine whether the data is chaotic. The data is available for reconstruction only when λ_{max} is larger than 0 [18].

2.5.1. Mutual information algorithm

MI is a statistical measurement that shows the relationship between two variables, it quantifies the amount of information transmitted in one random variable about another [38]. This method had been used in Section 2.2.3 for feature selection. The following provides the details of the algorithm. Let $H(X)$ be the average amount of information conveyed by X, which is called the information entropy. The following equation shows the function of $H(X)$.

$$H(X) = -\sum_{i=1}^{n} P(x_i) \log P(x_i) \tag{8}$$

where X is a condition set, n is the length of X, and $P(x_i)$ is the probability of condition x_i. The uncertainty of X given a measurement of Y is defined as $H(X \mid Y)$, which is the conditional information entropy. Eq. (10) shows the function of $H(X \mid Y)$.

$$H(X \mid Y) = -\sum_{i=1}^{n} \sum_{j=1}^{m} P(y_j) P(x_i \mid y_j) \log P(x_i \mid y_j) \tag{9}$$

The mutual information of X, Y ($I(X,Y)$) is the amount that a measurement of Y reduces the uncertainty of X, which is shown as follows:

$$I(X,Y) = H(X) - H(X \mid Y) = H(X) + H(Y) - H(X,Y) \tag{10}$$

where $H(X,Y)$ is the joint information of X and Y and the function is shown as follows:

$$H(X,Y) = -\sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log P(x_i, y_j) \tag{11}$$

2.5.2. False nearest neighbors algorithm

The embedding dimension is another important factor in SSR, it represents how much hidden information can be exhibited in the one-dimensional space. If the choice of embedding dimension is not proper, it is not possible to effectively extract the patterns in the time series [18]. In this paper, the FNN algorithm is used to determine the embedding dimension for the selected feature data, which was proposed by Kennel [37]. The FNN points are defined as two adjacent points that are a large distance apart in high-dimensional space. The idea of the FNN algorithm is that in the progress of transforming from dimension d to dimension $d+1$, points on the orbit x_n can be differentiated by which of them are "true" neighbors and which of them are 'false' neighbors. In a too small embedding dimension, the attractor is hard to fold, some points appear as neighbors but actually are far from each other in a high dimension. For a d embedding dimension, let $Z_n^{(r)}$ be the rth nearest neighbor of Z_n. Then the Euclidean distance between the point and is neighbor $Z_n^{(r)}$ can be expressed as follows:

$$R_d(n,r) = \left\| \sum_{k=0}^{d-1} \left[x_{n+k\tau} - x_{n+k\tau}^{(r)} \right] \right\| \tag{12}$$

When d is increased by 1, the updated distance can be given as:

$$R_{d+1}^2(n,r) = R_d^2(n,r) + \left[x_{n+d\tau} - x_{n+d\tau}^{(r)} \right]^2 \tag{13}$$

two points are regarded as false nearest neighbor points when $R_d(n,r)$ is much less than $R_{d+1}(n,r)$. The criterion is defined as follows:

$$R_{diff}(n,r) = \frac{\left\| x_{n+d\tau} - x_{n+d\tau}^{(r)} \right\|}{R_d(n,r)} > \theta \tag{14}$$

where θ is threshold in the range of $[10, 50]$ [39]. When $R_{diff}(n,r)$ is larger than θ, the two points are considered as false nearest neighbors.

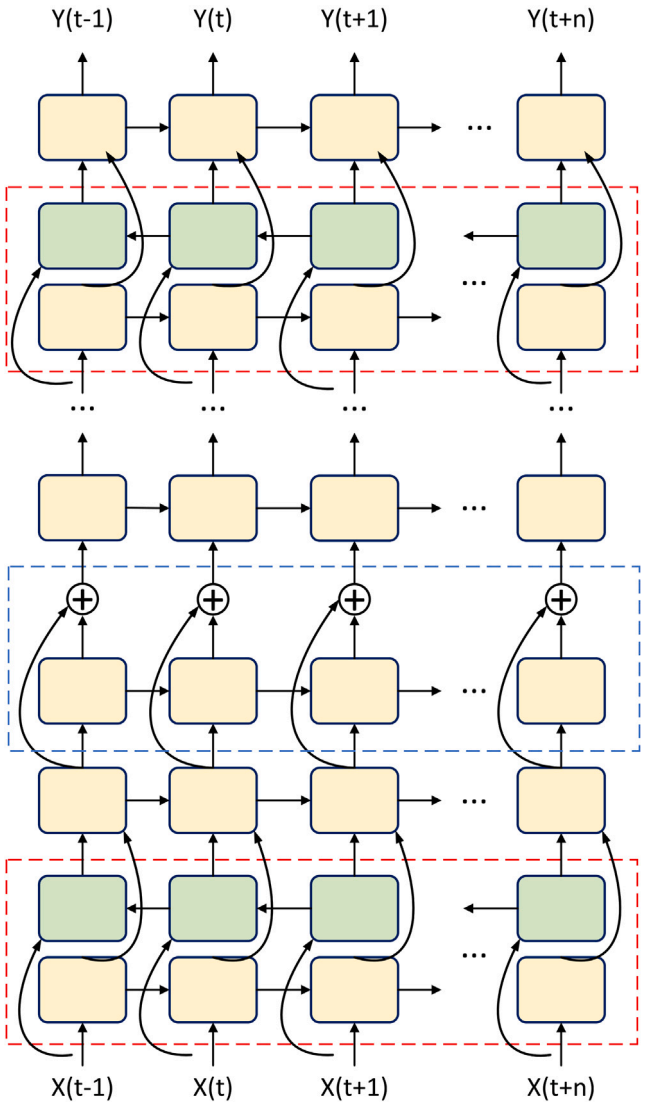

Fig. 5. The structure of the deep RNN model. Each rectangular represents an LSTM cell. The red dashed box is a Bi-LSTM layer consisting of a forward (yellow) and a backward (green) LSTM layer. The blue dashed box is the LSTM layer with residual connections.

2.5.3. Maximum Lyapunov exponent

The Maximum Lyapunov exponent (λ_{max}) is applied in this study to identify the chaotic feature in the selected feature set. Grebogi *etal*. [40] found that the data can be considered as chaotic data when λ_{max} is larger than zero. The exponential divergence between two neighboring trajectories in phase space is measured based on the maximum Lyapunov exponent [41] and the algorithm used in this paper was proposed by Rosenstein [42].

2.6. Deep recurrent neural network

After applying the SSR technique on the selected features, a deep RNN model is built for BP estimation for both feature sets. The overview of the proposed deep RNN model is shown in Fig. 5. The deep RNN model consists of a Bi-LSTM layer at the bottom, 5 LSTM layers with residual connections, followed by a Bi-LSTM and an LSTM at the top two layers.

2.6.1. The structure of LSTM layer

LSTM is the improved model of RNN which solves the problem of gradient vanishing by adding a memory cell state. The LSTM layer consists of three types of "gates" named forget gate, input gate, and output gate. Based on the cell state, the input gate determines what information can be stored in the cell state, the output gate determines what information can be output from the

cell state, and the forget gate decides what information will be removed from the cell state [43]. The mathematical expressions of the LSTM cell are shown as follows:

$$f_t = \sigma\left(W_f \cdot \left[h_{t-1}, x_t\right] + b_f\right) \tag{15}$$

$$i_t = \sigma\left(W_i \cdot \left[h_{t-1}, x_t\right] + b_i\right) \tag{16}$$

$$\tilde{C}_t = \tanh\left(W_C \cdot \left[h_{t-1}, x_t\right] + b_C\right) \tag{17}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{18}$$

$$O_t = \sigma\left(W_o \cdot \left[h_{t-1}, x_t\right] + b_o\right) \tag{19}$$

$$h_t = O_t \cdot \tanh\left(C_t\right) \tag{20}$$

where f_t, i_t, C_t, O_t, and h_t represent the forget gate, the input gate, the cell state of LSTM, the output gate, and the output of the cell, respectively. The subscript t and $t-1$ represent the time stamp. W and b are weights and bias respectively. The '$\cdot$' denotes the vectors' pointwise multiplication. $\sigma(\cdot)$ and $\tanh(\cdot)$ are activation functions that help map the vectors in the range of (0, 1) and (−1, 1). For the forget gate, $f_t = 1$ means all information in the previous time will be kept while the forget gate removes all information when $f_t = 0$ [17].

2.6.2. The structure of bi-LSTM layer

Conventional LSTM can only use past and present information $x_1, \ldots, x_{t-1}, x_t$. In order to overcome this shortcoming, the bidirectional RNN (BRNN) layer was added to the model, which was proposed by Schuster and Paliwal [44]. For BRNN, h_t not only captures the past information but also captures the information from the future by processing the data in both forward and backward directions with two separate hidden layers. As shown in the bottom of Fig. 5, each BI-LSTM contains a forward layer and a backward layer. The forward layer is the same as in the conventional LSTM, while for the backward layer, the hidden state and output are iterated from T to 1. The following shows the mathematical expressions of the LSTM cell.

$$h_{t_f} = \mathcal{H}(W_{h_f} \bullet [h_{t-1_f}, x_t\ ] + b_f) \tag{21}$$

$$h_{t_b} = \mathcal{H}(W_{h_b} \bullet [h_{t-1_b}, x_t\ ] + b_b) \tag{22}$$

$$h_t = W_{h_f} h_{t_f} + W_{h_b} h_{t_b} + b_h \tag{23}$$

3. Experiments and results

The experiment results are presented in the following parts. First, the results of feature selection algorithms are presented, as well as the SSR results for each specific feature in the selected feature set. Second, the ablation study is conducted to show the importance of different parts in the proposed model. Then, the results of RF and the proposed deep regression model are calculated and compared. Finally, the robustness analysis and the model comparison with BHS standard and other works are presented.

3.1. Feature selection

The optimal feature set for BP estimation is selected based on the proposed feature voting system introduced in Section 2.4. The Pearson correlation coefficient was calculated on the extracted 19 features. Setting $\rho_{X,Y} < -0.8$ or $\rho_{X,Y} > 0.8$ as a strong correlation, the features with high correlation among 19 features are shown in Table 5. Table 5 shows that PTT_{pp} and PTT_{pf} are highly correlated, t_{Ff}, t_{Fn} and HR are highly correlated, A_{Sf}, A_{Ns} and A_{Fs} are highly correlated. Based on considerations of computational complexity and accuracy, we have selected a set of five features for our experiment after experimenting with different numbers of features. Based on the algorithm shown in Section 2.4, for each algorithm, we sorted the feature score and selected the top 5 features. Results are listed in Table 6. Once a feature is selected, it is considered to receive one vote. The first five features are selected by counting the total number of votes for the selected feature. After applying the feature voting system, a total of 5 features are selected, which are PTT_{pp}, S_{Nf}, A_{Ns}, RI_{Sf}, and HR. The principle of exclusive high correlation features is to remove the feature with a lower vote, which may help improve the robustness of the whole system.

After selecting the feasible features, the SSR technique is applied to the chosen features to obtain more information. Before using SSR, it is necessary to determine whether the data is chaotic. Regarding the methods mentioned in Section 2.5, the three essential parameters λ_{max}, τ, and d for each feature were calculated. Table 7 shows the results of SSR operation on feature data. The maximum Lyapunov exponent of PTT_{pp} is less than zero, while λ_{max} for the other four features exceed zero, indicating that features S_{Nf}, A_{Ns}, RI_{Sf}, and HR are chaotic. Therefore, the SSR process can be used in those features.

After applying the proposed feature selection methods and the SSR technique, the original feature set and the reconstructed feature set are shown in Table 8.

Table 5
Features with high correlation and corresponding Pearson correlation coefficients.

	PTT_{pp}	PTT_{pf}	t_{Ff}	HR	A_{Sf}	A_{Ns}
PTT_{pf}	0.99	1				
t_{Fn}			0.94	−0.87		
t_{Ff}			1	−0.94		
A_{Ns}					0.89	1
A_{Fs}					0.86	0.94

Table 6
Selected features based on different algorithms.

Feature selection algorithm	Top 5 selected features
ReliefF	RI_{Sf}, PTT_{pf}, PTT_{pp}, t_{Ds}, A_{Ns}
CFS	RI_{Sd}, RI_{Sf}, A_{Fs}, S_{Nf}, PTT_{pp}
MI	A_{Ns}, RI_{Sf}, S_{Nf}, PTT_{pp}, PTT_{pf}
NCA	HR, S_{Nf}, t_{Nf}, t_{Fn}, PTT_{pp}
FSCMRMR	S_{Nf}, t_{Fn}, t_{Fn}, t_{Ff}, HR
RFE	RI_{Sn}, S_{Sf}, PTT_{pp}, HR, RI_{Sf}
Variance Thresholding	PTT_{pf}, PTT_{pp}, t_{Ds}, S_{Sd}, A_{Ns}

Table 7
Results for the maximum Lyapunov exponents, the time delay, and embedding dimension of selected features.

Features	λ_{max}	τ	d
PTT_{pp}	−0.6064	–	–
S_{Nf}	0.1423	4	3
A_{Ns}	0.0341	3	3
RI_{Sf}	0.3214	2	3
HR	0.4403	2	2

Table 8
Selected features with and without SSR.

Feature set	Features
Original (F_1)	PTT_{pp}, S_{Nf}, A_{Ns}, RI_{Sf}, HR
With SSR (F_2)	PTT_{pp}, S_{Nf}, $S_{Nf(\tau-3)}$, $S_{Nf(\tau-6)}$, A_{Ns}, $A_{Ns(\tau-3)}$, $A_{Ns(\tau-6)}$, RI_{Sf}, $RI_{Sf(\tau-3)}$, $RI_{Sf(\tau-6)}$, HR, $HR_{(\tau-3)}$, $HR_{(\tau-6)}$

3.2. Model performance

3.2.1. Evaluation criteria

In our work, two common evaluation metrics were used to evaluate the performance of the proposed method for BP estimation. One is performance metrics and the other one is relative diagrams.

Performance metrics are an important part of model analysis, which can tell us how well the model performed. For the BP estimation study, the outputs of the model are continuous, so regression metrics that calculate the sort of distance between predicted and ground truth are needed. Here 3 metrics are used to evaluate the model performance: the mean absolute error (MAE), the root mean square error (RMSE), and the correlation coefficient of the prediction (R). The formulas are shown as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|BP_i - \widehat{BP_i}\right| \tag{24}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(BP_i - \widehat{BP_i})^2} \tag{25}$$

$$R = \frac{\sum_{i=1}^{n}(BP_i - \overline{BP})(\widehat{BP_i} - \overline{\widehat{BP_i}})}{\sqrt{\sum_{i=1}^{n}(BP_i - \overline{BP})^2 \sum_{i=1}^{n}(\widehat{BP_i} - \overline{\widehat{BP_i}})^2}} \tag{26}$$

where $\overline{BP} = \frac{1}{n}\sum_{i=1}^{n} BP_i$, $\overline{\widehat{BP_i}} = \frac{1}{n}\sum_{i=1}^{n} \widehat{BP_i}$, BP_i is the reference BP value, and $\widehat{BP_i}$ is the predicted BP value. Smaller MAE and RMSE imply a higher accuracy of the model. A R^2 near 1 indicates that the model can predict the relationship well while R^2 equals zero explains 0% of the relationship between the reference value and predicted value.

Diagrams plotted in the results make it easier to observe the performance. Three different types of plots are analyzed in our study. First, the comparison plot is generated to show the prediction ability of the model. Then, the Bland–Altman graphs and correlation plots are produced to show the agreement between the ground truth BP and the predicted BP respectively.

Table 9
Ablation study on different feature sets based on MIMIC-II dataset.

Algorithm	Feature set	RMSE			MAE		
		SBP (Mean ± Std)	DBP (Mean ± Std)	MAP (Mean ± Std)	SBP (Mean ± Std)	DBP (Mean ± Std)	MAP (Mean ± Std)
RNN	Selected Features (F_1)	3.761 ± 2.00	1.987 ± 1.18	2.384 ± 1.21	2.939 ± 1.66	1.535 ± 1.02	1.872 ± 1.06
RNN	Total Features	4.078 ± 2.08	2.077 ± 1.06	2.433 ± 1.32	3.075 ± 1.82	1.648 ± 1.13	2.036 ± 1.23
RNN	SSR Features (F_2)	3.613 ± 1.49	1.978 ± 0.96	2.346 ± 0.97	2.765 ± 1.24	1.543 ± 0.84	1.795 ± 0.84

Table 10
Performance comparison on two feature sets using RF and proposed deep regression model based on MIMIC-II dataset.

Algorithm	Feature set	RMSE			MAE		
		SBP (Mean ± Std)	DBP (Mean ± Std)	MAP (Mean ± Std)	SBP (Mean ± Std)	DBP (Mean ± Std)	MAP (Mean ± Std)
RF	F_1	10.419 ± 5.12	5.722 ± 3.2	6.417 ± 3.45	8.11 ± 4.45	4.436 ± 2.78	4.976 ± 2.96
RF	F_2	10.303 ± 5.04	5.651 ± 3.16	6.342 ± 3.36	7.943 ± 4.3	4.325 ± 2.72	4.863 ± 2.83
RNN	F_1	3.761 ± 2.00	1.987 ± 1.18	2.384 ± 1.21	2.939 ± 1.66	1.535 ± 1.02	1.872 ± 1.06
RNN	F_2	3.613 ± 1.49	1.978 ± 0.96	2.346 ± 0.97	2.765 ± 1.24	1.543 ± 0.84	1.795 ± 0.84

3.2.2. Ablation study

Ablation studies are commonly used to explain the importance of features in modeling by systematically removing or modifying them to observe their impact on model performance [45]. To investigate the effectiveness of our proposed model, we conducted an ablation study on different input feature sets to demonstrate that (a) the feature voting system part in our proposed model improves the performance in BP estimation results, and (b) the SSR technique can improve the accuracy of the model prediction. The ablation study was conducted by using different input feature sets, results of SBP, DBP, and mean arterial pressure (MAP) are shown in Table 9. By comparing the model's performance using all features versus selected features, it was observed that the feature voting system enhances the predictive capability of the model. Similarly, when comparing the results using the SSR features and selected features, it was found that the SSR technique improves the model's prediction ability.

3.2.3. Experimental results

First, we tested the model performance by using the MIMIC-II dataset from the UCI database. Here we used random forest and our proposed deep regression model to evaluate the model performance by using both the original selected feature set (F_1) and the feature set with SSR (F_2). The results of the RF and the proposed deep RNN model trained with F_1 and F_2 are calculated indicated by the mean (Mean) and the standard deviation (Std) in Table 10. For the proposed deep regression model, the number of Bi-LSTM layer is 2 and the LSTM layer is 6. The hidden size of each cell is 128. For the training and test beat-to-beat feature sequence, we set each sequence length to 16 and the batch size to 32. The optimizer is set as Adam, with an initial learning rate of 0.001 and the learning rate is decreased to 0.0001, using 100 epochs for training. The source code of the proposed deep RNN model, as well as the random forest model, the feature extraction and selection part are released at https://github.com/LeoLyu111/An-Improved-DRM-with-SSR-for-Continuous-BP-Estimation.

It is easy to observe that our proposed model is much better than the random forest. For both of those two models, RMSE and MAE calculated based on F_2 are slightly lower than those got based on F_1. For RF model, from F_1 to F_2, the RMSE is decreased by 0.116 and 0.071 for SBP and DBP, the MAE is decreased by 0.167 and 0.111 for SBP and DBP, respectively. For the deep regression model, the RMSE is reduced by 0.148 and 0.09 for SBP and DBP from F_1 to F_2, the MAE for SBP decreased by 0.174 from F_1 to F_2 while the MAE for DBP got a small value based on F_1. The results indicate that the state space reconstruction technique has a great impact on improving the performance of BP prediction. The comparison plot of the ground truth BP and prediction BP of a representative subject from the MIMIC-II dataset by using SSR are plotted in Fig. 6. In general, both SBP and DBP are performed in close agreement.

To select the suitable batch size for the proposed deep RNN model, we calculated the RMSE and MAE of SBP and DBP by randomly selecting 10 subjects from the dataset evaluate models with batch size of 32 and 64, by applying the Wilcoxon test [46]. For RMSE, the p-value is 0.105 and 0.064 for SBP and DBP. For MAE, the p-value is 0.002 and 0.006 for SBP and DBP. Results show that modifying the value of batch size does not significantly influence the model performance on RMSE, while it has a significant difference on MAE. Based on the results, we selected the batch size of 32.

Besides, the Bland–Altman graphs (BA graph) of the estimation results of our proposed model with SSR are shown in Fig. 7. The BA graph is widely used in biomedical research to compare the agreement between two measurements [47]. BA graph uses the mean difference between BP and the average of BP to show the agreement between the prediction BP and reference BP [14]. The x-axis of the BA graph represents the average value of the reference BP and estimated BP and the difference between the reference BP and estimated BP is shown in the y-axis of the BA graph. The bias (Mean) and the limits of agreement ($Mean \pm 1.96 \times Std$) of our work are shown as dashed lines in Fig. 7. The mean difference of DBP and SBP are −0.363 and −0.221, respectively. It is shown that for both DBP and SBP, most of the points are within the limits of agreement, indicating that the estimated BP and the reference BP are in close agreement. In addition, Fig. 8 presents the correlation-based comparison of the reference BP and the predicted BP. The ground truth is shown in the x-axis and the predicted BP is shown in the y-axis. The red dashed line represents the relationship that the ground truth and the predicted BP are equal while the green line represents the linear regression line between the ground truth BP and the predicted BP. The correlation coefficient for DBP and SBP on test data is 0.706 and 0.718 respectively. Fig. 8 shows the

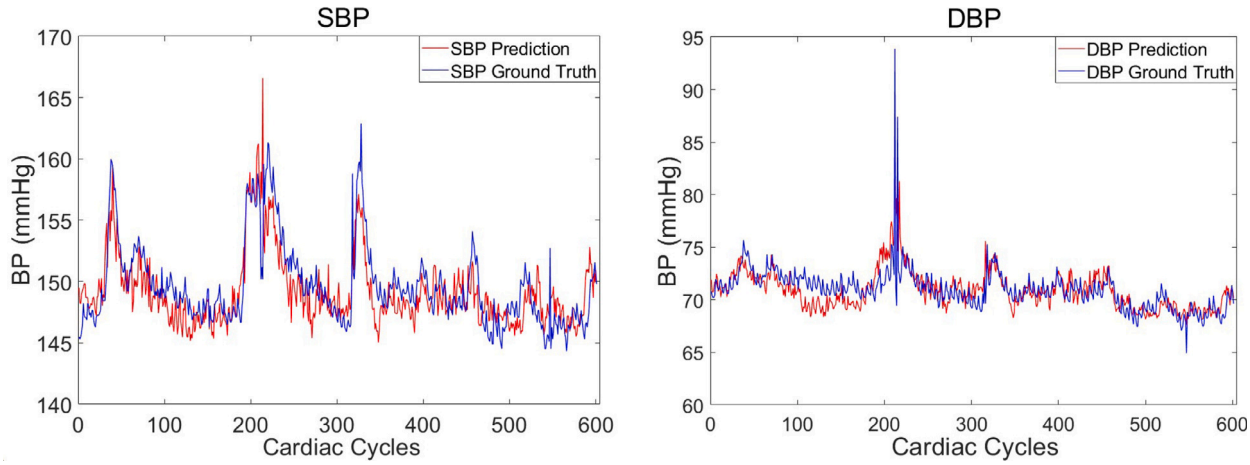

Fig. 6. Comparison of the ground truth BP and predicted BP of a representative subject from MIMIC-II with SSR.

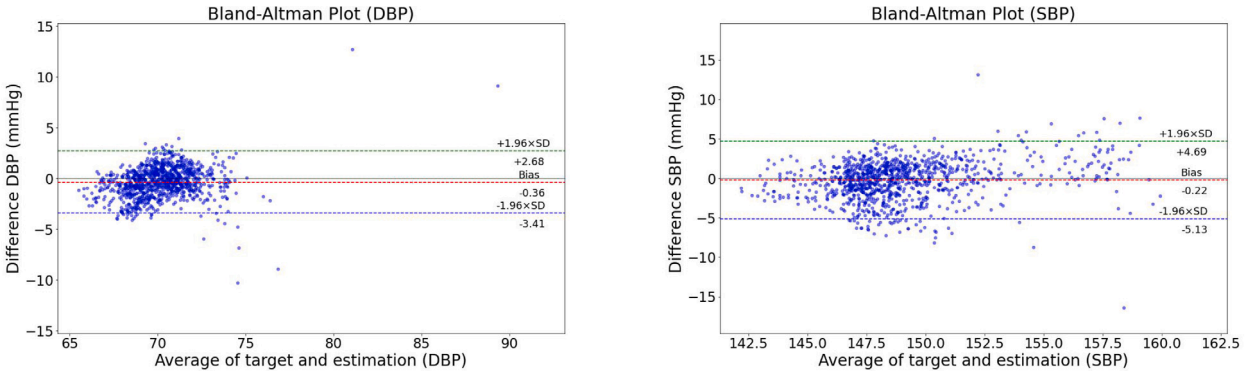

Fig. 7. Bland–Altman plots of the DBP (left) and SBP (right) predictions by the proposed deep RNN model with F_2.

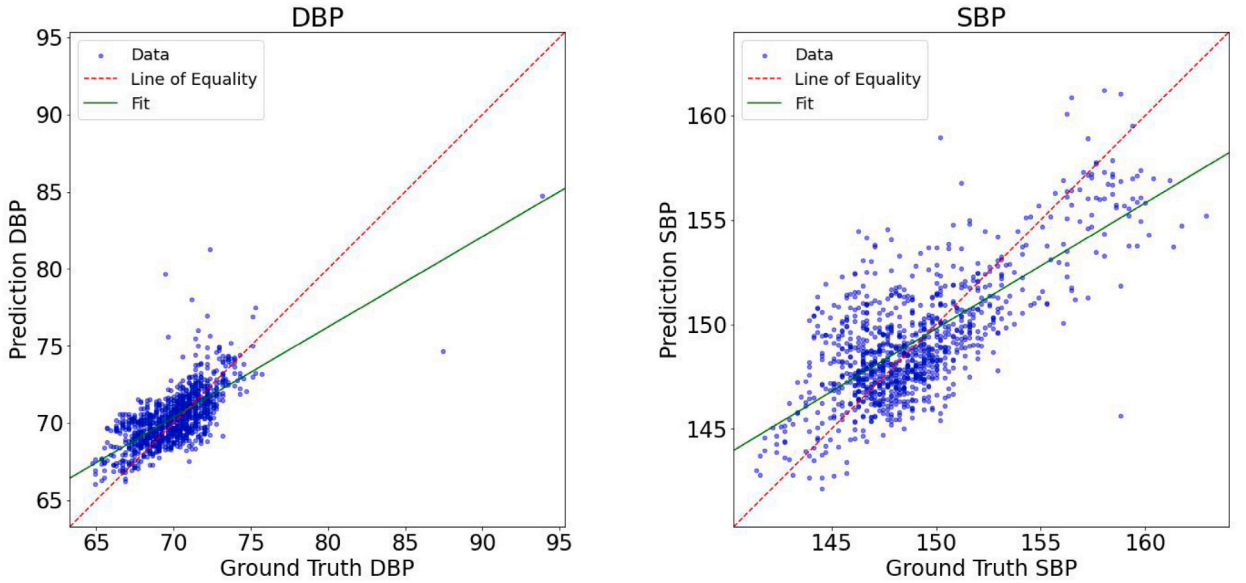

Fig. 8. Bland–Altman plots of the DBP (left) and SBP (right) predictions by the proposed deep regression model with F_2.

Table 11
Performance comparison on two feature sets using RF and proposed deep regression model based on Multi-day BP dataset.

Algorithm	Feature set	RMSE			MAE		
		SBP (Mean ± Std)	DBP (Mean ± Std)	MAP (Mean ± Std)	SBP (Mean ± Std)	DBP (Mean ± Std)	MAP (Mean ± Std)
RNN	F_1	5.819 ± 1.35	3.585 ± 0.79	3.783 ± 0.92	4.350 ± 1.21	2.760 ± 0.64	3.061 ± 0.69
RNN	F_2	5.387 ± 0.83	3.338 ± 0.60	3.611 ± 0.84	4.115 ± 0.75	2.553 ± 0.49	2.927 ± 0.61

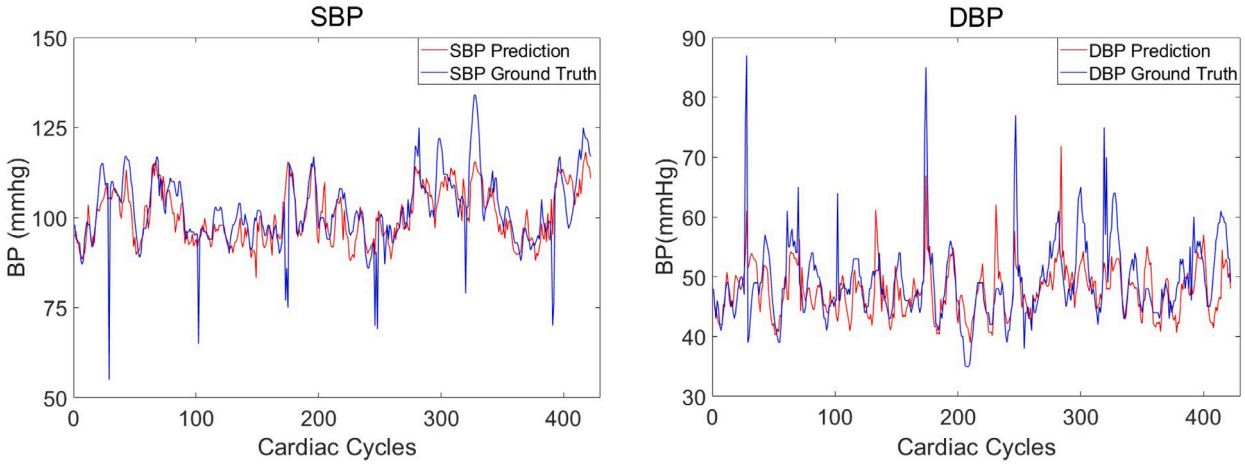

Fig. 9. Comparison of the ground truth BP and predicted BP of a representative subject from the multi-day BP with SSR.

matching degree between the model outputs and the reference BP. A correlation coefficient close to 1 indicates that the predicted value and the reference value are in close agreement.

To offer a more robust understanding of the proposed methodology's performance, we evaluated the proposed method on the multi-day BP dataset. Here we used the 1st day, 2nd day, and the 4th day's data for training and tested the BP value for the data collected 6 months after the first day. The efficacy of the proposed method is evaluated using two different feature sets, F_1 and F_2. Results are shown in Table 11 and the comparison plot of the ground truth BP and prediction BP of a representative subject in the multi-day BP dataset by using SSR are plotted in Fig. 9. On the multi-day BP dataset, the proposed model achieved RMSE of 5.387, 3.338, and 3.611 mmHg for SBP and achieved MAE of 4.115, 2.553, and 2.927 mmHg for DBP. The results indicate that the proposed method, which incorporates the SSR technique leads to a noticeable improvement in model performance. Overall, the results from the MIMIC-II dataset and the Multi-day BP dataset strongly support the efficacy of the proposed method in leveraging the SSR technique for improved predictive capabilities.

4. Discussion

4.1. Model comparison

4.1.1. Comparison with BHS standard

The British Hypertension Society (BHS) is used here to evaluate the BP estimation performance of the proposed model with SSR and the results are presented in Table 12. Based on the percentage of cumulative errors MAE less than the 3 threshold values, i.e., 5, 10, and 15 mmHg respectively, BHS grades the BP estimation into Grade A, B, and C [48]. Based on the BHS standard, the proposed deep regression model with SSR technique achieved Grade A for DBP, SBP, and MAP estimations.

4.1.2. Comparison with other methods

In order to better illustrate the performance of our proposed method in BP estimation, the comparative analysis of the results of our proposed method with the contemporary works is presented in Table 13. Most of the works compared in the table are based on the UCI dataset. Our proposed method gave an average RMSE of 1.978, 2.346, and 3.613 for DBP, MAP, and SBP, and gave an average MAE of 1.543, 1.795, and 2.765 for DBP, MAP, and SBP, respectively. El-Hajj et al. [49] proposed a 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. This study mentioned that the manual checks were performed to ensure that only good-quality segments are included for analysis. In this way, the work only included the segments with good quality and the error due to feature extraction can be reduced. Thus, it may cause an overly optimistic performance for BP estimation. The studies of [4,50–53] were extracted features based on CNN, which is different from manually extracted features. Our proposed model yields slightly inferior MAE results compared to the two-scale LRCN model introduced by Yan et al. [51]. However, it is important to note that the standard errors for SBP and DBP in the two-scale LRCN

Table 12
Comparison with the BHS standard.

		Cumulative error percentage		
		5 mmHg	10 mmHg	15 mmHg
Our results (MIMIC-II)	SBP	85.1%	98.5%	99.7%
	MAP	95.7%	99.5%	100%
	DBP	96.9%	99.6%	100%
Our results (Multi-day BP)	SBP	75.6%	89.8%	96.5%
	MAP	83.7%	92.7%	97.9%
	DBP	89.9%	95.3%	99.3%
BHS	Grade A	60%	85%	95%
	Grade B	50%	75%	90%
	Grade C	40%	65%	85%

Table 13
Performance comparison with related works for BP estimation.

Authors	Method	Dataset	RMSE			MAE		
			SBP	DBP	MAP	SBP	DBP	MAP
El-Hajj et al. [49]	Bi-GRU+GRU+attention	MIMIC II	–	–	–	2.58	1.26	–
Thambiraj et al. [14]	RF	MIMIC II	13.83	6.80	4.03	9.54	5.48	3.27
Ganti et al. [54]	PTT based	personal	4.75	2.72	2.99	4.03	2.24	2.5
Yan et al. [6]	SVM	personal	5.00	3.69	4.14	–	–	–
Zhang et al. [13]	MPGA-BPN	MIMIC II	8.23	8.92	–	6.51	6.11	–
Baker et al. [50]	CNN+LSTM	MIMIC III	–	–	–	4.53	3.37	3.36
Yen et al. [51]	two-scale LRCN	MIMIC II	–	–	–	2.24	1.40	–
Huang et al. [52]	MLPLSTM-BP	MIMIC II	5.10	3.13	–	3.52	2.13	–
Sharifi et al. [10]	MARS+SSR	MIMIC II	–	–	–	7.83	4.86	3.63
Malayeri et al. [53]	Concat_CNN	MIMIC II	5.26	2.61	–	3.05	1.58	–
Cheng et al. [4]	ABP-net	MIMIC II	–	–	–	3.27	1.90	1.49
Chowdhury et al. [55]	GPR	PPG-BP Dataset [56]	6.74	3.59	–	3.02	1.74	–
Our proposed method	RNN+SSR	MIMIC II	**3.61**	**1.98**	**2.35**	**2.77**	**1.54**	**1.80**

model are relatively higher at 3.59 and 2.56, respectively. In contrast, our proposed model demonstrates greater stability, surpassing the performance of the model presented in [51]. Zhang et al. [13] represented a fewer parameters BP estimation model based on RA-ReliefF feature selection and MPGA-BPN model. The idea of feature selection in this study is impressive while the performance of the model is poor. Overall, the performance of our model is comparable and better than most of the recent existing methods. Therefore, it is possible that our proposed method can be used in real-time BP monitoring applications.

4.2. Robustness analysis

The robustness of a model is measured by how much its performance changes when using new data versus training data. PPG is collected by utilizing an infrared light to measure the volumetric variations of blood circulation, it can be easily affected by sensing, biological, and cardiovascular factors [57]. Involuntary or voluntary movements can cause changes in inner tissues, such as muscle movement and dilation [7]. The receiving light will be modified due to these movements, which limits the accuracy of the PPG data. Similarly, ECG signals are susceptible to influence and thus there may exist noise when collected. ECG noise can stem from various sources, including poor electrode contact, motion artifacts, electromyography (EMG) interference, and baseline wander [58]. One way to test the robustness of the model is adding random noise on the original data [59]. Thus in this section, we test the robustness of our model by adding a 3 dB noise to the original PPG signals, as well as the original ECG signals. Then extract and select the same feature set using the new ECG and new PPG signals. Table 14 shows the results for both random forest and our proposed model with the feature set F_1 and F_2. Based on the results shown in Table 14, adding a 3 dB random noise on PPG or ECG or adding a 3 dB noise on both of them makes the prediction performance worse for both the random forest and our proposed deep regression model. For random forest on both F_1 and F_2, the increase of RMSE and MAE for SBP is around 4 mmHg and 5 mmHg respectively for adding 3 dB noise on PPG, and the increase of RMSE and MAE for DBP is about 2 mmHg. For our proposed model, regardless of whether we add noise to the PPG or ECG signal, or both the PPG and ECG signals, the increase of RMSE and MAE are less than 0.5 mmHg for SBP, DBP, and MAP. Results show that our proposed model is robust even if there is random noise in the input signals. Besides, the obtained results indicate that the RMSE and MAE values for PPG signals with noise were consistently higher compared to those of the ECG signals. This suggests that the PPG signal is more susceptible to the introduced noise, leading to a larger deviation between the noisy and original signals. These findings emphasize the importance of considering the robustness of PPG-based systems when dealing with noise interference, as it can significantly impact the accuracy and reliability of the measurements.

Table 14
Performance comparison of models by adding random noise.

Random Forest							
	Feature set	RMSE			MAE		
		SBP ± Std	DBP ± Std	MAP ± Std	SBP ± Std	DBP ± Std	MAP ± Std
Without random noise	F_1	10.419 ± 5.12	5.722 ± 3.20	6.417 ± 3.45	8.110 ± 4.45	4.436 ± 2.78	4.976 ± 2.96
	F_2	10.303 ± 5.04	5.651 ± 3.16	6.342 ± 3.36	7.943 ± 4.30	4.325 ± 2.72	4.863 ± 2.83
With 3 dB PPG noise	F_1	14.819 ± 8.24	7.557 ± 3.87	8.642 ± 4.48	12.694 ± 8.27	6.448 ± 3.72	7.389 ± 4.42
	F_2	14.425 ± 8.22	7.541 ± 3.87	8.601 ± 4.46	12.61 ± 8.23	6.418 ± 3.73	7.339 ± 4.40
With 3 dB ECG noise	F_1	11.789 ± 6.04	6.254 ± 3.97	7.104 ± 4.05	9.192 ± 5.42	4.951 ± 3.52	5.588 ± 3.59
	F_2	11.578 ± 5.94	6.067 ± 3.82	6.918 ± 3.90	8.946 ± 5.19	4.764 ± 3.33	5.406 ± 3.35
With 3 dB PPG and ECG noise	F_1	15.281 ± 7.37	8.243 ± 4.25	9.066 ± 4.11	11.623 ± 7.41	6.953 ± 3.95	7.502 ± 3.96
	F_2	14.836 ± 7.41	8.032 ± 4.20	8.982 ± 4.13	11.595 ± 7.36	6.893 ± 3.92	7.476 ± 3.94
RNN							
	Feature set	RMSE			MAE		
		SBP ± Std	DBP ± Std	MAP ± Std	SBP ± Std	DBP ± Std	MAP ± Std
Without random noise	F_1	3.761 ± 2.00	1.987 ± 1.18	2.384 ± 1.21	2.939 ± 1.66	1.535 ± 1.02	1.872 ± 1.06
	F_2	3.613 ± 1.49	1.978 ± 0.96	2.346 ± 0.97	2.765 ± 1.24	1.543 ± 0.84	1.795 ± 0.84
With 3 dB PPG noise	F_1	3.998 ± 1.83	2.104 ± 1.01	2.544 ± 1.09	2.989 ± 1.40	1.641 ± 0.87	1.981 ± 0.92
	F_2	3.936 ± 1.51	2.073 ± 0.88	2.501 ± 0.91	3.031 ± 1.19	1.589 ± 0.72	1.929 ± 0.75
With 3 dB ECG noise	F_1	3.875 ± 1.94	2.043 ± 1.04	2.439 ± 1.10	3.012 ± 1.54	1.595 ± 0.90	1.914 ± 0.94
	F_2	3.821 ± 1.83	2.023 ± 0.87	2.411 ± 1.00	3.010 ± 1.32	1.578 ± 0.82	1.899 ± 0.84
With 3 dB PPG and ECG noise	F_1	4.142 ± 2.02	2.181 ± 1.12	2.641 ± 1.23	3.189 ± 1.57	1.687 ± 0.97	2.046 ± 1.05
	F_2	4.104 ± 1.87	2.140 ± 1.09	2.610 ± 1.05	3.156 ± 1.40	1.664 ± 0.88	2.022 ± 0.92

5. Limitations

The proposed work in the paper has several limitations that should be considered. Firstly, the evaluation of the proposed method is limited to a benchmark dataset with 660 subjects. To enhance the reliability and applicability of the approach, future research could focus on validating the method on larger and more diverse datasets to assess its performance in different scenarios. Since external validation is crucial for assessing the generalizability and reliability of the proposed model beyond the specific dataset used in the study. Secondly, here we briefly mention the validation of the proposed model's robustness by introducing random noise to the PPG signals, the extent to which the model can handle various levels and types of noise is not thoroughly discussed. Investigating the model's noise robustness under different conditions is essential to determine its reliability in practical scenarios.

6. Conclusion and future work

In this paper, we proposed a BP monitoring approach for continuous and noninvasive BP estimation by combining the feature selection system and the state space reconstruction technique with the proposed deep regression model. Primarily, the proposed method consists of signal preprocessing, feature extraction, feature selection, state space reconstruction, and the deep RNN model. A feature voting system is proposed to obtain the feature set by combining the voting of eight feature selection algorithms. Then the stats space reconstruction technique is applied to generate more hidden information from the features. The time delay and embedding dimension are calculated by the MI algorithm and false nearest neighbors algorithm, respectively. Then, the RF model and the proposed deep regression model are developed for estimating BP. Using the SSR technique improved the performance of BP prediction for both of the two models. The combination of SSR and the proposed deep RNN model provides the best results for continuous BP estimation. The results of our proposed method are validated using the BHS standard. The estimation of DBP, MAP, and SBP achieved Grade A according to the BHS standard. In addition, we test the robustness of our model by adding 3 dB random noise on PPG signals and the result shows that our model is robust. However, the results may not reflect the overall performance of our proposed method because it was trained and tested on 660 subjects. In the future, we first plan to test our proposed method on more datasets. Then, based on the advantages and disadvantages of two feature extraction ways, i.e., manually extracting features and using deep neural networks to extract features automatically, we try to combine those two feature extraction ways together to improve the model performance by using more useful features.

CRediT authorship contribution statement

Liangyi Lyu: Conceptualization, Methodology, Software, Validation, Writing – original draft. **Lei Lu:** Conceptualization, Writing – review & editing. **Hanjie Chen:** Conceptualization, Software, Resources, Data curation, Writing – review & editing. **David A. Clifton:** Conceptualization. **Yuanting Zhang:** Supervision, Project administration. **Tapabrata Chakraborti:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

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

References

[1] Chow CK, Teo KK, Rangarajan S, Islam S, Gupta R, Avezum A, et al. Prevalence, awareness, treatment, and control of hypertension in rural and urban communities in high-, middle-, and low-income countries. JAMA 2013;310(9):959–68.
[2] Liu Q, Zheng Y, Zhang Y, Poon CCY. Beats-to-beats estimation of blood pressure during supine cycling exercise using a probabilistic nonparametric method. IEEE Access 2021;9:115655–63.
[3] Ding X-R, Zhao N, Yang G-Z, Pettigrew RI, Lo B, Miao F, et al. Continuous blood pressure measurement from invasive to unobtrusive: Celebration of 200th birth anniversary of Carl Ludwig. IEEE J Biomed Health Inform 2016;20(6):1455–65.
[4] Cheng J, Xu Y, Song R, Liu Y, Li C, Chen X. Prediction of arterial blood pressure waveforms from photoplethysmogram signals via fully convolutional neural networks. Comput Biol Med 2021;138:104877.
[5] Esmaelpoor J, Moradi MH, Kadkhodamohammadi A. A multistage deep neural network model for blood pressure estimation using photoplethysmogram signals. Comput Biol Med 2020;120:103719.
[6] Yan W-R, Peng R-C, Zhang Y-T, Ho D. Cuffless continuous blood pressure estimation from pulse morphology of photoplethysmograms. IEEE Access 2019;7:141970–7.
[7] Castaneda D, Esparza A, Ghamari M, Soltanpur C, Nazeran H. A review on wearable photoplethysmography sensors and their potential future applications in health care. Int J Biosens Bioelectron 2018;4(4):195.
[8] Ding X-R, Zhang Y-T, Liu J, Dai W-X, Tsang HK. Continuous cuffless blood pressure estimation using pulse transit time and photoplethysmogram intensity ratio. IEEE Trans Biomed Eng 2015;63(5):964–72.
[9] Yang L, Wang H, Yuan W, Li Y, Gao P, Tiwari N, et al. Wearable pressure sensors based on mxene/tissue papers for wireless human health monitoring. ACS Appl Mater Interfaces 2021;13(50):60531–43.
[10] Sharifi I, Goudarzi S, Khodabakhshi MB. A novel dynamical approach in continuous cuffless blood pressure estimation based on ECG and PPG signals. Artif Intell Med 2019;97:143–51.
[11] Ma Y, Choi J, Hourlier-Fargette A, Xue Y, Chung HU, Lee JY, et al. Relation between blood pressure and pulse wave velocity for human arteries. Proc Natl Acad Sci 2018;115(44):11144–9.
[12] Shen Z, Chakraborti T, Wang W, Yao S, Fu Z, Chen Y, et al. Uncertainty quantification of cuffless blood pressure estimation based on parameterized model evidential ensemble learning. Biomed Signal Process Control 2024;92:106104.
[13] Zhang L, Ji Z, Yang F, Chen G. Noninvasive continuous blood pressure estimation with fewer parameters based on RA-relieff feature selection and MPGA-BPN models. Biomed Signal Process Control 2023;84:104757.
[14] Thambiraj G, Gandhi U, Mangalanathan U, Jose VJM, Anand M. Investigation on the effect of womersley number, ECG and PPG features for cuffless blood pressure estimation using machine learning. Biomed Signal Process Control 2020;60:101942.
[15] Su P, Ding X-R, Zhang Y-T, Liu J, Miao F, Zhao N. Long-term blood pressure prediction with deep recurrent neural networks. In: 2018 IEEE eMBS international conference on biomedical & health informatics. BHI, IEEE; 2018, p. 323–8.
[16] Cheng J, Xu Y, Song R, Liu Y, Li C, Chen X. Prediction of arterial blood pressure waveforms from photoplethysmogram signals via fully convolutional neural networks. Comput Biol Med 2021;138:104877.
[17] Lu L, Tan Y, Oetomo D, Mareels I, Zhao E, An S. On model-guided neural networks for system identification. In: 2019 IEEE symposium series on computational intelligence. SSCI, IEEE; 2019, p. 610–6.
[18] Dong M, Tang C, Ji J, Lin Q, Wong K-C. Transmission trend of the COVID-19 pandemic predicted by dendritic neural regression. Appl Soft Comput 2021;111:107683.
[19] Saeed M, Villarroel M, Reisner AT, Clifford G, Lehman L-W, Moody G, et al. Multiparameter intelligent monitoring in intensive care II (MIMIC-II): a public-access intensive care unit database. Crit Care Med 2011;39(5):952.
[20] Barvik D, Cerny M, Penhaker M, Noury N. Noninvasive continuous blood pressure estimation from pulse transit time: A review of the calibration models. IEEE Rev Biomed Eng 2021;15:138–51.
[21] Kachuee M, Kiani MM, Mohammadzade H, Shabany M. Cuffless blood pressure estimation algorithms for continuous health-care monitoring. IEEE Trans Biomed Eng 2016;64(4):859–69.
[22] Venkatesh B, Anuradha J. A review of feature selection and its methods. Cybern Inf Technol 2019;19(1):3–26.
[23] Remeseiro B, Bolon-Canedo V. A review of feature selection methods in medical applications. Comput Biol Med 2019;112:103375.
[24] Guyon I, Gunn S, Nikravesh M, Zadeh LA. Feature extraction: foundations and applications, vol. 207, Springer; 2008.
[25] Kavsaoğlu AR, Polat K, Hariharan M. Non-invasive prediction of hemoglobin level using machine learning techniques with the PPG signal's characteristics features. Appl Soft Comput 2015;37:983–91.
[26] Sarker IH. Machine learning: Algorithms, real-world applications and research directions. SN Comput Sci 2021;2(3):160.
[27] Urbanowicz RJ, Meeker M, La Cava W, Olson RS, Moore JH. Relief-based feature selection: Introduction and review. J Biomed Inform 2018;85:189–203.
[28] Mohamad M, Selamat A, Krejcar O, Fujita H, Wu T. An analysis on new hybrid parameter selection model performance over big data set. Knowl-Based Syst 2020;192:105441.
[29] Nguyen BH, Xue B, Zhang M. A survey on swarm intelligence approaches to feature selection in data mining. Swarm Evol Comput 2020;54:100663.
[30] Solorio-Fernández S, Carrasco-Ochoa JA, Martínez-Trinidad JF. A review of unsupervised feature selection methods. Artif Intell Rev 2020;53(2):907–48.
[31] Ruehle F. Data science applications to string theory. Phys Rep 2020;839:1–117.
[32] Malan NS, Sharma S. Feature selection using regularized neighbourhood component analysis to enhance the classification performance of motor imagery signals. Comput Biol Med 2019;107:118–26.

[33] Raghu S, Sriraam N. Classification of focal and non-focal EEG signals using neighborhood component analysis and machine learning algorithms. Expert Syst Appl 2018;113:18–32.
[34] Makowski D, Ben-Shachar MS, Patil I, Lüdecke D. Methods and algorithms for correlation analysis in R. J Open Source Softw 2020;5(51):2306.
[35] Takens F. Detecting strange attractors in turbulence. In: Dynamical systems and turbulence, warwick 1980: proceedings of a symposium held at the university of warwick 1979/80. Springer; 2006, p. 366–81.
[36] Zou Y, Donner RV, Marwan N, Donges JF, Kurths J. Complex network approaches to nonlinear time series analysis. Phys Rep 2019;787:1–97.
[37] Kennel MB, Brown R, Abarbanel HD. Determining embedding dimension for phase-space reconstruction using a geometrical construction. Phys Rev A 1992;45(6):3403.
[38] Learned-Miller EG. Entropy and mutual information. Department of Computer Science, University of Massachusetts, Amherst; 2013, p. 4.
[39] Sangiorgio M, Dercole F. Robustness of LSTM neural networks for multi-step forecasting of chaotic time series. Chaos Solitons Fractals 2020;139:110045.
[40] Grebogi C, Ott E, Yorke JA. Crises, sudden changes in chaotic attractors, and transient chaos. Physica D 1983;7(1–3):181–200.
[41] Caesarendra W, Kosasih B, Tieu K, Moodie CA. An application of nonlinear feature extraction-A case study for low speed slewing bearing condition monitoring and prognosis. In: 2013 IEEE/aSME international conference on advanced intelligent mechatronics. IEEE; 2013, p. 1713–8.
[42] Rosenstein MT, Collins JJ, De Luca CJ. A practical method for calculating largest Lyapunov exponents from small data sets. Physica D 1993;65(1–2):117–34.
[43] Yu Y, Si X, Hu C, Zhang J. A review of recurrent neural networks: LSTM cells and network architectures. Neural Comput 2019;31(7):1235–70.
[44] Schuster M, Paliwal KK. Bidirectional recurrent neural networks. IEEE Trans Signal Process 1997;45(11):2673–81.
[45] Rao S, Li Y, Ramakrishnan R, Hassaine A, Canoy D, Cleland J, et al. An explainable transformer-based deep learning model for the prediction of incident heart failure. Ieee J Biomed Health Inform 2022;26(7):3362–72.
[46] Lu L, Zhu T, Ribeiro AH, Clifton L, Zhao E, Zhou J, et al. Decoding 2.3 million ECGs: Interpretable deep learning for advancing cardiovascular diagnosis and mortality risk stratification. Eur Heart J-Digital Health 2024;ztae014.
[47] Karun KM, Puranik A. BA. plot: An R function for bland-altman analysis. Clin Epidemiol Glob Health 2021;12:100831.
[48] O'Brien E, Petrie J, Littler W, de Swiet M, Padfield PL, O'Malley K, et al. The british hypertension society protocol for the evaluation of automated and semi-automated blood pressure measuring devices with special reference to ambulatory systems. J Hypertens 1990;8(7):607–19.
[49] El-Hajj C, Kyriacou PA. Deep learning models for cuffless blood pressure monitoring from PPG signals using attention mechanism. Biomed Signal Process Control 2021;65:102301.
[50] Baker S, Xiang W, Atkinson I. A computationally efficient CNN-LSTM neural network for estimation of blood pressure from features of electrocardiogram and photoplethysmogram waveforms. Knowl-Based Syst 2022;250:109151.
[51] Yen C-T, Chang S-N, Liao C-H. Estimation of beat-by-beat blood pressure and heart rate from ECG and PPG using a fine-tuned deep CNN model. IEEE Access 2022;10:85459–69.
[52] Huang B, Chen W, Lin C-L, Juang C-F, Wang J. MLP-BP: A novel framework for cuffless blood pressure measurement with PPG and ECG signals based on MLP-mixer neural networks. Biomed Signal Process Control 2022;73:103404.
[53] Malayeri AB, Khodabakhshi MB. Concatenated convolutional neural network model for cuffless blood pressure estimation using fuzzy recurrence properties of photoplethysmogram signals. Sci Rep 2022;12(1):6633.
[54] Ganti VG, Carek AM, Nevius BN, Heller JA, Etemadi M, Inan OT. Wearable cuff-less blood pressure estimation at home via pulse transit time. IEEE J Biomed Health Inform 2020;25(6):1926–37.
[55] Chowdhury MH, Shuzan MNI, Chowdhury ME, Mahbub ZB, Uddin MM, Khandakar A, et al. Estimating blood pressure from the photoplethysmogram signal and demographic features using machine learning techniques. Sensors 2020;20(11):3127.
[56] Liang Y, Chen Z, Liu G, Elgendi M. A new, short-recorded photoplethysmogram dataset for blood pressure monitoring in China. Sci Data 2018;5(1):1–7.
[57] Song SB, Nam JW, Kim JH. NAS-PPG: PPG-based heart rate estimation using neural architecture search. IEEE Sens J 2021;21(13):14941–9.
[58] Bashar SK, Ding E, Walkey AJ, McManus DD, Chon KH. Noise detection in electrocardiogram signals for intensive care unit patients. IEEE Access 2019;7:88357–68.
[59] Lu L, Tan Y, Oetomo D, Mareels I, Clifton DA. Weak monotonicity with trend analysis for unsupervised feature evaluation. IEEE Trans Cybern 2022.