Fair non-contact blood pressure estimation using imaging photoplethysmography

HONGLI FANG, JIPING XIONG,* AND LINYING HE

College of Physics and Electronic Information Engineering, Zhejiang Normal University, Jinhua 321000, China
*xjping@zjnu.cn

Abstract: Hypertension is typically manifested as a latent symptom that requires detection through specialized equipment. This poses an inconvenience for individuals who need to undergo long-term blood pressure monitoring in their daily lives. Therefore, there is a need for a portable, non-contact method for estimating blood pressure. However, current non-contact blood pressure estimation methods often rely on relatively narrow datasets, lacking a broad range of blood pressure distributions. Additionally, their applicability is confined to controlled experimental environments. This study proposes a non-contact blood pressure estimation method suitable for various life scenarios, encompassing multiple age groups, diverse ethnicities, and individuals with different skin tones. The aim is to enhance the practicality and accuracy of existing non-contact blood pressure estimation methods. The research extracts the imaging photoplethysmogram (IPPG) signal from facial videos and processes the signal through four layers of filtering operations to obtain an IPPG signal reflecting pulse wave variations. A CNN+BiLSTM+GRU network structure is constructed to improve the accuracy of current non-contact blood pressure estimation methods. In comparison to existing approaches, the mean absolute error (MAE) for systolic blood pressure (SBP) and diastolic blood pressure (DBP) is reduced by 13.6% and 16.4%, respectively.

1. Introduction

One of the most significant risk factors for cardiovascular disease is hypertension, which frequently goes unnoticed due to the absence of apparent symptoms. Effective blood pressure surveillance over an extended period of time is crucial for preserving cardiovascular health and averting hypertension-related complications, including kidney and heart disease. Prompt identification of hypertension facilitates timely intervention and mitigates the likelihood of severe complications resulting from hypertension [1,2]. However, contact methods, such as wrist or cuff blood pressure meters, are the most prevalent for measuring blood pressure [3]. These methods typically require the presence of a healthcare professional. Long-term daily blood pressure surveillance is unfeasible with this method of blood pressure measurement. As a result, a portable non-contact blood pressure detection method is immediately required.

Several research studies have established a correlation between photoplethysmography (PPG) signals and blood pressure [4]. The pulse wave data present in PPG signals can provide insights into the extent of vasoconstriction and dilation of blood vessels, and these variations are strongly associated with systolic and diastolic blood pressures (SBP and DBP, respectively). The aforementioned variations exhibit a strong correlation with the SBP and DBP of individuals' blood. By employing deep learning techniques to analyze PPG signals, it is possible to accurately and efficiently forecast the SBP and DBP readings of patients' blood pressure. However, the collection of PPG signals is not convenient, requiring contact-based acquisition through photonic sensors on areas like fingers, toes, or the nose. This poses an inconvenience for individuals with skin sensitivity or burn patients. In contrast, Imaging Photoplethysmography (IPPG) utilizes cameras or similar imaging devices to capture changes in skin color on the body's surface, extracting pulse wave signals [5]. Utilizing IPPG signals for blood pressure monitoring

significantly enhances the convenience of blood pressure detection, enabling a non-contact approach for long-term daily monitoring. Additionally, it provides a means of blood pressure measurement for remote diagnostics by healthcare professionals.

Based on the current limited research on non-contact blood pressure estimation and the ongoing need to enhance its accuracy, a non-contact blood pressure estimation method is presented in this article. This method utilizes face videos to extract IPPG signals, which are then subjected to a four-layer filtering procedure in order to enhance their accuracy. By effectively detecting the subject's blood pressure value from the IPPG signals, the CNN+LSTM+GRU network developed in this study improves the accuracy of current non-contact blood pressure detection for individuals with different skin tones in a variety of real-world scenarios, particularly for DBP.

In recent research endeavors, notable progress has been achieved in non-contact blood pressure detection. For instance, W. Xing et al. [6] conducted an investigation into the influence of pulse wave signal extraction on blood pressure prediction. They integrated traditional Chinese medicine methodologies with distinct Regions of Interest (ROI) on the face. Their findings indicated that the left and right cheeks, along with the chin region, are the three most crucial facial ROI for accurate blood pressure prediction, yielding a final prediction accuracy of 90%.

In a similar vein, Yuheng Chen et al. [7] proposed a methodology involving the conversion of facial videos from RGB format to an enhanced YUV format, which separates color from luminance and enhances blood pressure prediction accuracy. They developed a ResNet18+BiLSTM model, achieving Mean Absolute Errors (MAE) of 12.35 and 9.54 for SBP and DBP, respectively, on the MMSE-HR dataset [8]. Recognizing the challenges associated with the extraction of IPPG signals due to factors such as subject head movements and variations in ambient light, some researchers have adopted dedicated IPPG acquisition devices. Yunjie Li et al. [9], for example, utilized a self-developed IPPG signal acquisition device to extract signals from the forehead region. They collected blood pressure and IPPG data from 403 subjects aged 17 to 21 years. The constructed deep-learning models demonstrated MAE ± STD values of 8.36 ± 6.22 for SBP and 5.69 ± 3.97 for DBP. Moreover, K. Iuchi et al. [10] employed a facial holder to stabilize facial videos taken at a camera frame rate of 160 fps. They proposed deep learning Convolutional Neural Network (CNN) architectures based on ResNet [10] and CBAM [11]. Through their approach, the SBP and DBP errors were minimized, achieving MAE (mmHg) values of 6.7 and 5.4, respectively. Notably, IPPG signals can also be obtained from regions beyond the face. Bin Lin et al. [12] introduced a method involving side-view videos, selecting the cheek and neck regions for Photoplethysmogram (PPG) signal extraction. They utilized a compact two-stage CNN model for blood pressure prediction.

Due to the current prevalence of non-contact blood pressure detection studies being conducted on relatively singular datasets, where the age and blood pressure distribution of participants are not sufficiently diverse, and the data collection environments are not representative of real-world settings, there is a lack of a blood pressure detection method that leans more towards practical everyday use. Therefore, the dataset chosen for this study encompasses three distinct populations with different races and skin tones. It covers a wide range of ages and blood pressure distributions, including data from various indoor and outdoor scenarios. A four-layer filter is devised to tailor IPPG signals, facilitating the extraction of pulse wave signals that characterize physiological signal variations.

Subsequently, a CNN+BiLSTM+GRU model is constructed to train and test the dataset, aiming to enhance the accuracy of non-contact blood pressure prediction in practical real-world scenarios. This approach represents a departure from the limitations associated with single datasets, contributing to a more comprehensive understanding of non-contact blood pressure detection across diverse populations and realistic environments.

2. Blood pressure testing procedure

Figure 1 depicts the blood pressure detection procedure used in this investigation. First, the nose and the portions of the left and right cheeks are chosen as ROIs for the purpose of obtaining IPPG signals using the Dlib-81 library (A Detection Model Containing 81 Keypoints of a Face) [13]. Next, the signal is taken out of the RGB spectrum's green channel. The original IPPG signal is then subjected to a four-layer filtering procedure in order to improve the pulse wave signal's accuracy. Next, a 250-length sequence cropping is applied to the filtered IPPG signal, eliminating any sequences that are less than 250 frame lengths. The CNN+BiLSTM+GRU network model receives the cropped IPPG signal as input, allowing for independent training of the SBP and DBP values. In the end, the SBP and DBP values of the measured subject are produced by the network model.
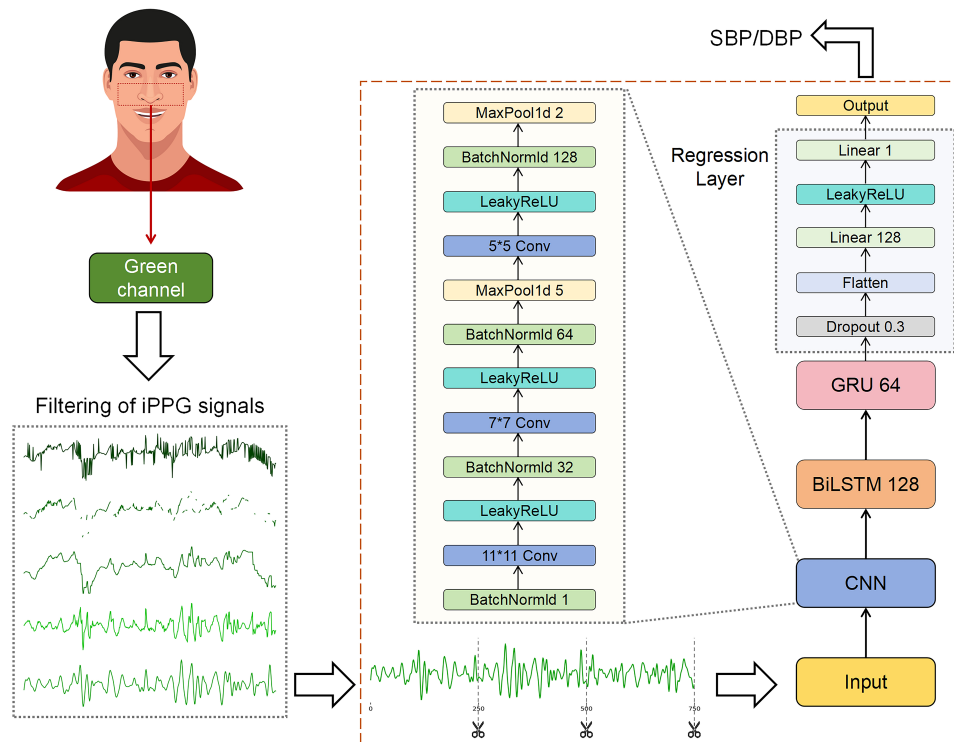
Fig. 1. Blood pressure detection process and CNN+BiLSTM+GRU network model.

3. Sources of datasets

We have chosen the publicly available VV-Small dataset [14] as a component of our dataset. This dataset encompasses facial videos and blood pressure data from 100 participants aged 8 to 91 in diverse environments, including both indoor and outdoor settings. Due to the dataset's lack of samples from individuals of Asian descent and insufficient representation of young participants, our study incorporates an additional private dataset (CN-BP), comprising 43 Asian students aged 19 to 22, to augment the dataset. Simultaneously, we conducted training and testing of the constructed CNN+BiLSTM+GRU model on the publicly available MIMIC-II dataset, validating the model's applicability to Photoplethysmography (PPG) signals.

3.1. VV-Small dataset

VV-Small is a dataset publicly released by Toye Pieter-Jan in 2023 [14]. For each participant, facial videos were recorded using the Basler acA1920-40uc camera model. Each participant contributed a 30-second video, recorded at a camera resolution of 1920 x 1200 and a frame rate of 30 fps. The blood pressure values of each participant were measured using the Omron M7 cuff blood pressure monitor. The dataset was collected in diverse real-life settings, including shopping malls, libraries, and communities. It encompasses a varied demographic, with participants featuring different skin tones, including 100 individuals of European and African descent, spanning ages from 8 to 91 years. The data utilized in our study consists of blood pressure values and facial videos from the VV-Small dataset. The facial videos are crucial for extracting IPPG signals. Due to instances of severe facial video exposure, faces moving out of the recording area, and insufficient lighting in some videos, a curation process resulted in the retention of facial videos and blood pressure data from 85 participants.
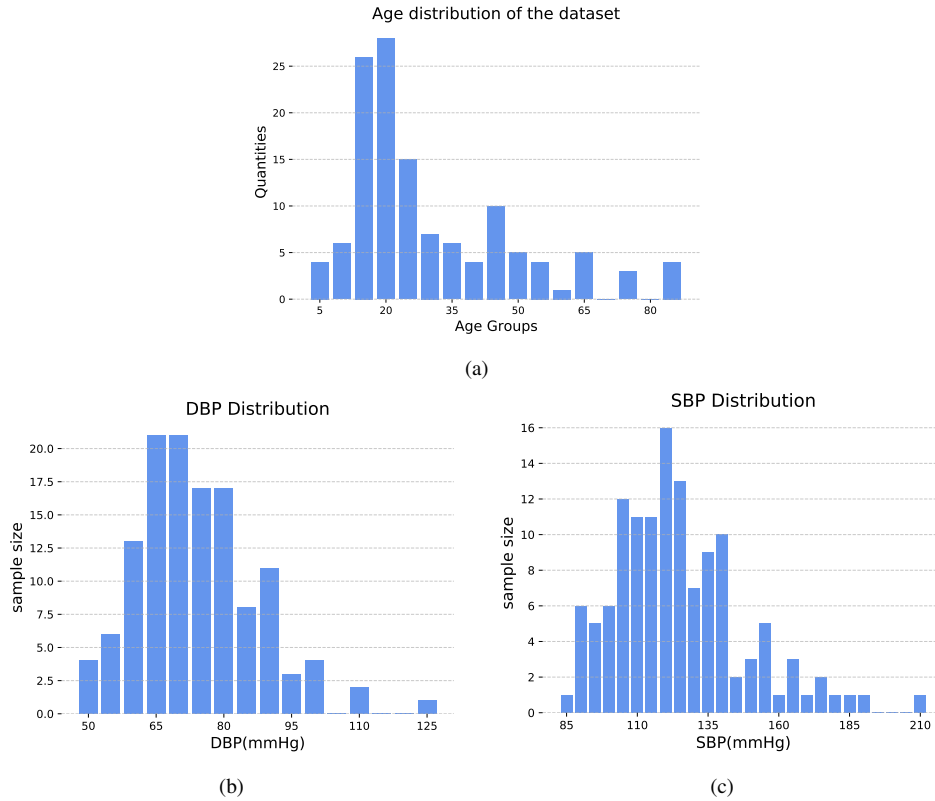
3.2. CN-BP dataset

To augment the dataset with additional samples and include data from individuals of Asian descent, we recorded facial videos and collected blood pressure data from 43 Zhejiang Normal University students, thus establishing the CN-BP dataset. The measurement process is illustrated in Fig. 2. The video recording employed a camera with a resolution of 1920x1080 and a frame rate of 30 fps. Each student's facial video was recorded for a duration of 30 seconds, and simultaneous blood pressure measurements were taken using the Yuyue YE670A cuff blood pressure monitor during the recording process.
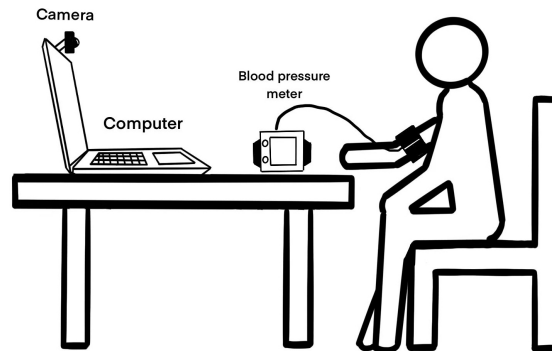
Fig. 2. Schematic diagram of blood pressure detection in the CN-BP dataset.

Merging the collected data from 43 Asian students with the dataset of 85 participants from VV-Small, we obtained a comprehensive dataset comprising blood pressure data and facial videos from 128 individuals representing three ethnicities (Asian, European, and African) across various age groups. The age distribution histogram of the combined CN-BP and VV-Small dataset is illustrated in Fig. 3(a), with a standard deviation of 19.13, encompassing individuals aged 8 to 91. The distribution of DBP and SBP is shown in Fig. 3(b) and (c), with mean values (mmHg) ± standard deviations (mmHg) of 76.21 ± 13.24 and 126.59 ± 23.02, respectively. Additionally, based on the Fitzpatrick scale (a classification method assessing human skin color into six categories), the merged dataset includes individuals representing six different skin color categories. Consequently, research based on this dataset is more versatile and applicable to a diverse range of skin tones. Considering these factors, we will conduct experiments using the merged dataset of 128 participants to objectively and practically evaluate the effectiveness of the proposed non-contact blood pressure detection method in this study.

Fig. 3. Distribution of the combined VV-Small+CN-BP dataset. (a) Age distribution in the merged dataset. (b) Distribution of Diastolic Blood Pressure (DBP) in the merged dataset. (c) Distribution of Systolic Blood Pressure (SBP) in the merged dataset

3.3. MIMIC-II dataset

MIMIC-II (Medical Information Mart for Intensive Care II) is a publicly available dataset collaboratively created by the Massachusetts Institute of Technology and Harvard Medical School [15]. It includes fingertip PPG signals and invasive blood pressure signals among various vital sign data from over 15,000 patients in Intensive Care Units (ICU). Currently, it is widely employed in research focused on predicting blood pressure using PPG signals. For our study, we selected 12,200 data entries from 80 patients, each comprising fingertip PPG signals along with SBP and DBP values. The PPG signal is sampled at a frequency of 125 Hz. The distribution of DBP and SBP is illustrated in Fig. 4, with DBP ranging from 50 to 118 and SBP ranging from 65 to 185. The mean values ± standard deviations are 63.71 ± 11.89 for DBP and 127.56 ± 19.90 for SBP. We utilize this dataset to validate the proposed non-contact blood pressure detection model and filtering method in our study, assessing their applicability to blood pressure detection from PPG signals.
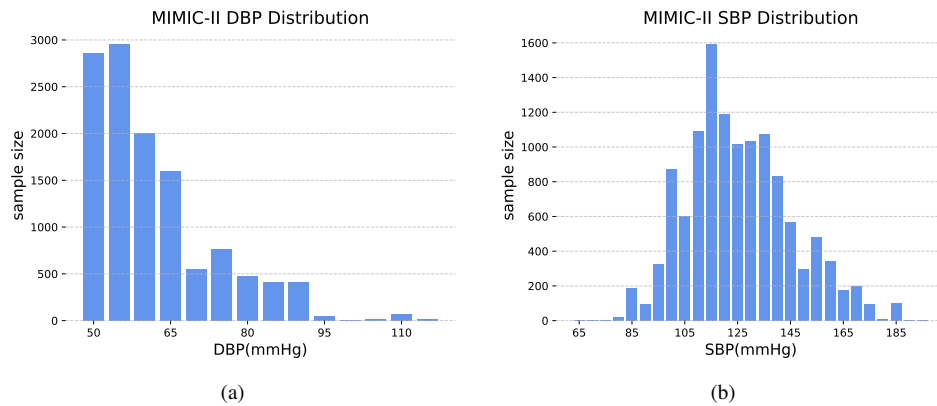
Fig. 4. Distribution of blood pressure in the MIMIC-II dataset. (a) Distribution of Diastolic Blood Pressure (DBP) in the MIMIC-II dataset. (b) Distribution of Systolic Blood Pressure (SBP) in the MIMIC-II dataset.

4. IPPG signal extraction and pre-processing

This study utilizes facial videos for the extraction of IPPG signals, employing the Dlib-81 facial keypoint detection model [16] for facial ROI segmentation. Due to potential obstructions such as hair covering the forehead and the presence of facial hair on the chin, we opted to extract the IPPG signals from the nose and left/right cheek regions. These facial areas (nose, and cheeks) exhibit a rich network of capillaries, enhancing the effective extraction of IPPG signals. Given that IPPG signals are subtle pulse wave signals susceptible to distortions from environmental light variations and facial movements, we applied filtering processes to the signals to ensure the accuracy of the IPPG signal waveform.

4.1. IPPG signal extraction method

Firstly, for each facial video in the CN-BP and VV-Small datasets, a frame-by-frame image capture was conducted. The Dlib-81 facial keypoint detection model was employed for facial localization, selecting the nose and left and right cheek areas as ROIs. These ROIs were segmented from the facial images using the detected keypoints for subsequent signal extraction. Choosing the green channel in image processing holds an advantage, as green light closely aligns with the spectrum of hemoglobin [17]. This allows for the extraction of more effective pulse waves [18]. This choice is further supported by the stability of the response of the green channel across individuals with different skin tones [6], making it suitable for extracting IPPG signals from diverse populations. Ultimately, in the frame-by-frame images of facial videos, the average pixel brightness of the ROIs area in the green channel was taken as the IPPG signal strength at each moment. Each participant in the dataset corresponds to a 30-second facial video, yielding a 900-frame IPPG signal extraction for further analysis.

4.2. IPPG signal pre-processing

As the experimental dataset was collected in diverse and complex environments, and pulse wave signals are extremely subtle, various interfering factors such as environmental conditions and noise significantly impact the variations in IPPG signals. Consequently, preprocessing of the IPPG signals extracted from facial videos is essential. The preprocessing involves a four-step filtering method, aimed at obtaining IPPG signals that accurately reflect real changes in the human pulse wave. The original IPPG signal undergoes anomaly detection to remove irregularities. Subsequently, correction is applied using Kalman filtering to eliminate anomalies,

resulting in a signal devoid of irregularities. This refined signal then undergoes Empirical Mode Decomposition (EEMD) filtering, enhancing the accuracy of Intrinsic Mode Functions (IMF) extraction and reducing mode mixing phenomena. Finally, bandpass filtering is applied to eliminate Gaussian white noise components generated during the EEMD filtering process. The progression from the initial signal to the four-stage filtering process is illustrated in Fig. 5(a) and (b) for two sample instances.
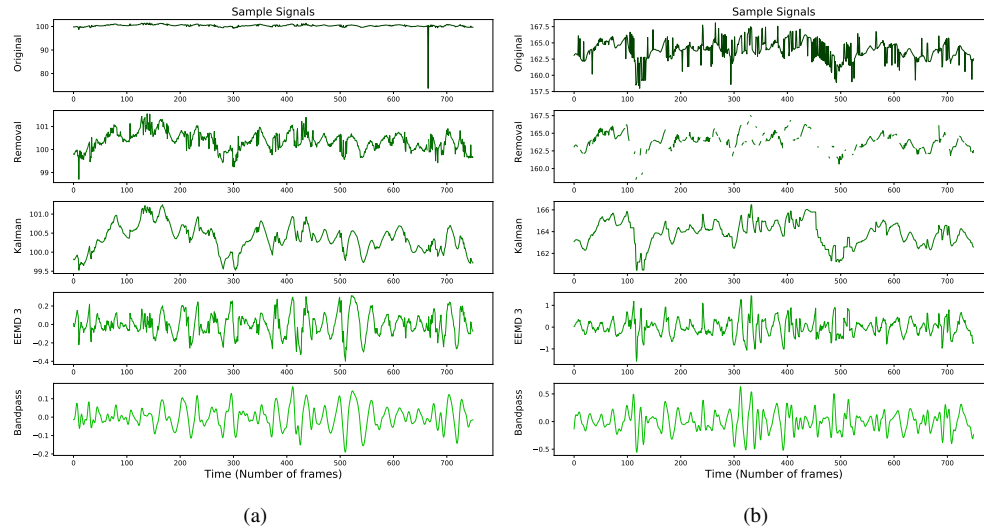
Fig. 5. The two samples (a) and (b) from the four-layer filtering process.

4.2.1. IPPG signal anomaly removal

Due to potential disturbances during the signal acquisition process, such as rapid facial movements and sudden changes in environmental light brightness, the obtained signals may exhibit abrupt variations at specific moments (frames), as illustrated in Fig. 5(a) and (b) by the Original signal. These irregularities significantly impact the accurate acquisition of pulse wave signals. Additionally, in the EEMD decomposition process outlined in section 5.2.3, the extraction of IMF relies on local extremal points. Eliminating abnormal points in the signal reduces the instability of extremal points during the EEMD decomposition, thereby enhancing the accuracy of IMF decomposition. To address this, we assess the amplitude difference between each pair of adjacent points in the signal sequence. If the amplitude difference exceeds a certain threshold (set at 1), we consider it an anomalous jump in signal amplitude for that frame. Importantly, we identify these jumps as anomalies only if they are not caused by the pulse wave signal. Signals with such anomalies are designated as exceptional, and we replace the values at these anomaly points with null values to eliminate signal irregularities. The processed signal is represented as the Removal signal in Fig. 5. In the Kalman filtering process, we set the initial state estimate to the mean of the IPPG signal. In cases where observational data is null, the Kalman filtering prediction process is sustained. The post-Kalman filtering IPPG signal, illustrated as the Kalman signal in Fig. 5, predicts abnormal points in the absence of data, ensuring the continuity of the IPPG signal and thus rectifying signal anomalies. Simultaneously, the application of Kalman filtering eliminates numerous white noise components from the signal, resulting in an IPPG signal devoid of anomalies and characterized by reduced noise levels.

4.2.2.   Kalman filter

We chose Kalman filtering to correct the anomalies in the signal because anomalous points are eliminated from the signal after anomaly detection and set to null values [19]. Kalman filtering reduces noise and estimates the state of the system. It anticipates the real state estimation value of the signal's anomalous points, which have been set to null values, and uses the estimate of the previous state to predict the present state. The Kalman filter's time update equation makes this operation possible. The predicted current state value is then corrected by the Kalman filter's state update equation in conjunction with the current state measurement value, resulting in a more precise final state estimation value. By using a prediction-correction procedure, Kalman filtering successfully removes noise that was introduced by outside sources when the IPPG signals were being collected. This creates the basis for accurately recovering pulse wave signals. The Kalman filter's state update equation and time update equation are defined as follows [20].

Time update equation:

$$\hat{X}_{k,k-1} = F\hat{X}_{k-1,k-1} \tag{1}$$

$$P_{k,k-1} = FP_{k-1,k-1}F^T + Q \tag{2}$$

State update equation:

$$K_k = P_{k,k-1}H^T \left(HP_{k,k-1}H^T + R\right)^{-1} \tag{3}$$

$$\hat{X}_{k,k} = \hat{X}_{k,k-1} + K_k \left(Z_k - H\hat{X}_{k,k-1}\right) \tag{4}$$

$$P_{k,k} = (I - K_k H)P_{k,k-1} \tag{5}$$

In the given equations, $\hat{X}_{k-1,k-1}$ represents the state estimate at time $k-1$, indicating the previous time's corrected result after prediction correction. $\hat{X}_{k,k}$ represents the state estimate at time k, the current time after correction. $\hat{X}_{k,k-1}$ denotes the predicted state estimate at the current time based on the prediction from time $k-1$, representing the uncorrected prediction for time k. F is the state transition matrix, which transforms the corrected state estimate at time $k-1$ to the state estimate at time k. $P_{k-1,k-1}$ and $P_{k,k}$ respectively represent the covariance matrices for the corrected state estimates at time $k-1$ and k. The covariance matrix reflects the uncertainty in the state estimate, continuously updating with each time increment. $P_{k,k-1}$ is the covariance matrix at time k for $\hat{X}_{k,k-1}$, an intermediate value reflecting the covariance matrix's updating changes. Q is the process noise covariance matrix. Together with the state transition matrix, it participates in the updating calculation of the $P_{k,k-1}$ covariance matrix. K_k is defined as the Kalman gain. Z_k represents the current observation data value at time k, while H and R represent the observation matrix and observation noise covariance matrix, respectively. The observation matrix maps the state variables to the observation data space, and the observation noise covariance matrix embodies the uncertainty in Z_k.

In the Kalman filtering process, we set the initial state estimate to the mean of the IPPG signal. In cases where observational data is null, the Kalman filtering prediction process is sustained. The post-Kalman filtering IPPG signal, illustrated as the Kalman signal in Fig. 5, predicts abnormal points in the absence of data, ensuring the continuity of the IPPG signal and thus rectifying signal anomalies. Simultaneously, the application of Kalman filtering eliminates numerous white noise components from the signal, resulting in an IPPG signal devoid of anomalies and characterized by reduced noise levels.

4.2.3.   EEMD decomposition

As depicted in Fig. 5 for the two sample signals post-Kalman filtering, it is evident that the signals still exhibit considerable environmental or motion noise affecting the amplitude. To address this, we employed EEMD [21] to decompose the signals, aiming to further eliminate

additional noise and extract the characteristic features of the pulse wave. Following Kalman filtering, the IPPG signal was subjected to EEMD decomposition, resulting in multiple IMFs. Through experimentation, we determined that selecting the third-order IMF component from the EEMD decomposition as the output retained more pulse wave signal features while effectively removing other noise components.

EEDM is an improved method proposed to address the issue of mode mixing in Empirical Mode Decomposition (EMD) [22]. By incorporating Gaussian white noise with a zero mean to aid in the analysis, EEDM leverages the characteristic of a uniformly distributed spectrum of white noise to mitigate mode mixing during the EMD decomposition process. This enhancement improves the robustness of EMD by addressing the phenomenon of mode mixing resulting from the uneven distribution of signal extrema [23].

EMD can decompose complex signals into multiple IMFs and a residual term. It is an adaptive method capable of decomposing nonlinear signals into IMF components of different frequencies. Initially, local maxima and minima of the signal are identified, and the upper and lower envelope lines are obtained through spline interpolation. The mean is then calculated to determine the average envelope. The IMF components are obtained by subtracting the calculated average envelope from the initial signal. The resulting IMF components must satisfy the following conditions:

    a. The difference in the number of local maxima and minima within the signal range is at most one or equal.

    b. At any given moment, the mean of the upper and lower envelope lines, determined by local maxima and minima, is zero.

If the above IMF conditions are not met, the operation of extracting IMF needs to be repeated. In the end, several IMF components and a residual term are obtained [24]. Since pulse wave signals are weak pulsation signals, EMD struggles to accurately decompose them. EEMD improves the accuracy of decomposing weak oscillation components by introducing white noise. The initial signal $X(t)$ is expressed as $X'(t)$ by adding Gaussian white noise $\omega(t)$:

$$X'(t) = X(t) + \omega(t) \tag{6}$$

Applying EMD to $X'(t)$ yields multiple IMF components. The decomposed IMF is denoted as $c_j(t)$, with the residual represented as $r_n(t)$:

$$X'(t) = \sum_{j=1}^{n} c_j(t) + r_n(t) \tag{7}$$

To eliminate the impact of Gaussian white noise on the signal, the IMF component $c_{ij}(t)$ of the N times added white noise decomposition is then averaged; $c_j(t)$ is represented as the jth order IMF component of the EEMD decomposition:

$$c_j(t) = \frac{1}{N} \sum_{i=1}^{N} c_{ij}(t) \tag{8}$$

In this study, we experimentally determined the use of the third-order IMF component obtained through Empirical Mode Decomposition with EEMD for output, as illustrated in EEMD3 in Fig. 5. The third-order IMF component, after decomposition, successfully eliminates the influence of other environmental noise, extracting the subtle pulse wave signal components within the IPPG signal.

4.2.4. Bandpass filtering

During the EEMD decomposition process, there is a possibility of some added white noise being mistakenly identified as part of the IMF components. As illustrated in Fig. 5, even after the third-order IMF decomposition in EEMD3, a certain amount of noise signal persists. To address this, we applied a Fourier transform and constructed a bandpass filter to eliminate the residual noise signal. By examining the spectral plot after Fourier transformation, we designed a bandpass filter within the 0.5-9Hz range. The output of the bandpass filter is depicted in Fig. 5 as the Bandpass signal. With this, we have completed the four-stage filtering process of the original IPPG signal, resulting in a signal applicable for training the CNN-BiLSTM-GRU network model.

5. Results

Firstly, we conducted comparative experiments on the merged dataset (VV-Small+CN-BP) on the different stages of filtering methods used in the signal preprocessing process and the selection of different lengths of signal slices. It was used to verify the effectiveness of the preprocessing process in improving the accuracy of estimated blood pressure. Following the signal preprocessing steps, we conducted training and testing of the network model on both the combined datasets (VV-Small + CN-BP) and MIMIC-II. Throughout the experimental process, various existing network models were compared to validate the effectiveness and reliability of the proposed method in this study.

5.1. Experiments on combined datasets

5.1.1. Model and evaluation metrics

The network model constructed in this study integrates Convolutional Neural Network (CNN), Bidirectional Long Short-Term Memory (BiLSTM), and Gated Recurrent Unit (GRU). The architectural depiction of the network model is illustrated in Fig. 1. In the CNN layer, the input signal is normalized through BatchNormalization. Subsequently, a series of convolutional operations is applied to extract local pulsatile features from the IPPG signal. The Leaky ReLU activation function is introduced for non-linearity. Following BatchNormalization and convolutional operations, the channel count is increased, and the kernel size is reduced progressively to extract more complex features. Max-pooling operations are employed to reduce dimensions and mitigate overfitting risks.

Following the CNN layer, a BiLSTM layer with 128 units has been incorporated. BiLSTM, equipped with forget gates, input gates, and output gates, simultaneously acquires forward and backward information from temporal data. This aids the network model in comprehending the periodic patterns and characteristics of the pulse wave. To extract richer and more abstract features from the signal, the output of the BiLSTM layer serves as the input for the GRU layer, further modeling temporal dependencies. The GRU layer captures long-term dependencies in temporal data through update and reset gates. With fewer parameters, the GRU layer accommodates rapid sequence variations. The combination of BiLSTM and GRU layers enables the network model to consider two distinct types of temporal dependencies concurrently, enhancing its generalization capabilities.

Toward the end of the network model, we have defined a Regression Layer responsible for generating the model's output. In the Regression Layer, a Dropout layer with a dropout rate of 0.3 has been incorporated to mitigate the risk of model overfitting. A fully connected layer is employed to amalgamate features from preceding layers, and a non-linear activation function is applied to alleviate the risk of gradient vanishing.

We partitioned the participants in the VV-Small+CN-BP dataset, with training data comprising data from 80% of the participants, while data from the remaining 20% of participants was used for the test set. A total of 324 IPPG signal data segments, derived from 108 participants

post-cropping, were employed as the training set. Simultaneously, 78 IPPG signal segments from 26 participants, also post-cropped, were designated as the test set. During the model training process, we utilized the following parameters: a batch size of 8, a learning rate of 0.001, and 250 training epochs, and employed Adam as the optimizer. Standard Deviation (STD), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE) were employed as evaluation metrics for model testing, with the calculation formulas as follows:

$$STD = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N - 1}} \tag{9}$$

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i| \tag{10}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \tag{11}$$

In this context, x_i represents the error between the predicted blood pressure value and the standard blood pressure value for the ith prediction, and $\bar{x}$ denotes the average error. y_i and $\hat{y}_i$ respectively signify the ith standard blood pressure value and the ith blood pressure value predicted by the model, with N representing the total amount of data.

5.1.2. Comparative Experiments on Preprocessing Procedures

To compare the impact of different filtering methods at various stages and different signal slice lengths during the preprocessing of IPPG signals on the accuracy improvement of blood pressure estimation, we conducted comprehensive comparative experiments using the constructed CNN+BiLSTM+GRU network on the merged dataset (VV-Small+CN-BP). The experimental results are presented in Table 1, with dataset partitioning and model parameters consistent with those described in Section 6.1.1.

Table 1. Comparative Experiments in the Preprocessing Stage.

Signal	Filtering Methods	Slice Length	SBP(mmHg)			DBP(mmHg)		
			STD	RMSE	MAE	STD	RMSE	MAE
Original Signal	—	Unsliced (900)	21.05	21.98	18.82	13.33	13.27	10.60
		150	**19.62**	20.50	17.28	11.54	11.50	9.12
		250	19.79	19.82	**15.20**	**10.70**	**10.61**	**8.40**
		350	19.83	**19.79**	16.28	10.93	10.93	8.72
Two-layer Filtering	Removal, Kalman	Unsliced (900)	20.20	20.69	16.92	11.63	11.43	9.29
		150	20.08	20.09	15.60	10.57	10.54	8.49
		250	**17.88**	**17.70**	**13.62**	**10.29**	**10.22**	**8.19**
		350	18.26	18.08	14.18	10.55	10.46	8.55
Four-layer Filtering	**Removal, Kalman, EEMD 3, Bandpass**	Unsliced (900)	19.58	19.63	16.04	9.04	9.23	7.17
		150	18.30	18.28	13.50	8.93	8.95	6.13
		250	**17.08**	**17.12**	**12.40**	**8.33**	**8.36**	**5.74**
		350	17.76	17.84	13.02	8.75	8.68	6.14

When using IPPG signals from different filtering stages for non-invasive blood pressure estimation, we observed that, compared to the original signal, the use of two-stage filtering (Removal and Kalman) enhances the overall accuracy of SBP and DBP estimation. Further decomposition of the signal through EEMD to extract the third IMF component, followed by bandpass filtering, results in a further reduction of blood pressure estimation errors. The IPPG signal after four layers of filtering, as opposed to signals with fewer filtering layers or the original signal, contains more information reflecting changes in the pulse wave. This contributes to the improvement of non-contact blood pressure estimation.

In Table 1, we conducted comparative experiments with signals of different lengths. We utilized the original signal length (900) without any signal clipping, as well as signals trimmed to lengths of 150, 250, and 350. When clipping the signals, segments shorter than 250 or 350 were discarded. By contrasting the accuracy of estimating SBP and DBP with varying signal lengths, we observed that the unclipped signals, containing excessive pulse wave variations, hindered the network from effectively learning crucial gradient information. Consequently, the accuracy of predicting SBP and DBP improved for signals that underwent clipping during various filtering stages. Furthermore, when comparing signals trimmed to lengths of 150, 250, and 350, we found that excessively short clipped signals resulted in suboptimal model predictions due to inadequate information capture, in contrast to longer clipped signal segments.

In the end, we discovered that trimming the signal into sequences of length 250 was most advantageous for the neural network model to capture the temporal information of the pulse wave during the training process, thereby enhancing the performance and generalization ability of the network model. When trimming the original signal, without filtering, to a length of 250 for predicting SBP, the indicators of STD and RMSE were slightly higher compared to sequences of lengths 150 and 350. However, with an increase in the number of filtering layers, the IPPG signal contained more pronounced variations in pulse wave information. At this point, trimming the signal to a length of 250, as opposed to other lengths, demonstrated a clear advantage in improving the accuracy of the network model's estimation of SBP and DBP blood pressure values. Therefore, as depicted in Fig. 1, we employed a four-layer filtering approach during the preprocessing of the IPPG signal. Subsequently, the filtered signal was sliced into sequences of length 250 before being inputted into the CNN+BiLSTM+GRU network model for training and testing.

5.1.3. Experimental comparison of models

To evaluate the performance of the constructed models in this study, we conducted experiments using different neural network architectures on preprocessed IPPG signals, and the experimental results are presented in Table 2. We experimented with various network models while reproducing the structures proposed by Schrumpf et al. [25], Mahardika T et al. [26], and Mou et al. [27] Comparative analysis from the experiments indicates that our constructed CNN+BiLSTM+GRU achieved the highest accuracy, with the three metrics—STD, RMSE, and MAE—all reaching their lowest values. Specifically, the MAE values for SBP and DBP are 12.40 and 5.74, respectively. It is noteworthy that, compared to the model proposed by Mou et al. [27], which performed the best, our model showed a significant improvement in the accuracy of estimating SBP and DBP. The MAE for SBP decreased by 13.6%, and the MAE for DBP decreased substantially by 16.4%. After preprocessing, the IPPG signals trained through the deep learning network models can predict blood pressure values within a certain margin of error. However, due to the substantial variability in SBP among individuals and the smaller variability in DBP, the model exhibits better performance in predicting DBP.

Concurrently, we conducted cross-dataset comparative experiments, evaluating the accuracy of predicting blood pressure through the extraction of IPPG signals from facial videos using different methods. The comparative data were obtained from the accuracies of the best models

Table 2. Comparative results of experiments with different models on the combined dataset.

	Models	STD	RMSE	MAE
SBP (mmHg)	VGG-16	21.10	21.03	17.20
	RestNet-18	18.70	18.74	15.56
	BiLSTM	21.67	21.63	18.07
	GRU	21.42	21.34	17.47
	Schrumpf et al. [25]	19.74	19.73	15.58
	Mahardika T et al. [26]	19.60	19.62	14.95
	Mou et al. [27]	18.06	17.95	14.36
	Ours(CNN+BiLSTM+GRU)	**17.08**	**17.12**	**12.40**
DBP (mmHg)	VGG-16	12.09	12.06	9.74
	RestNet-18	9.67	11.16	8.28
	BiLSTM	10.99	11.35	8.72
	GRU	11.77	12.47	8.67
	Schrumpf et al. [25]	9.69	10.69	7.71
	Mahardika T et al. [26]	9.57	10.27	7.58
	Mou et al. [27]	9.22	9.54	6.87
	Ours(CNN+BiLSTM+GRU)	**8.33**	**8.36**	**5.74**

and methods identified in existing literature, as illustrated in Table 3. Despite the complexity of the dataset employed in this study, which originates from diverse natural environments, and covers various age groups, different ethnicities, and a wide range of skin tones, the method employed yielded an MAE of 12.40 for measuring SBP, second only to the 12.35 achieved by Chen et al. [7]. As for measuring DBP, the MAE was 5.74, demonstrating a notable improvement compared to existing methods.

Table 3. Comparative results with existing methods on their respective datasets.

Method	Year Published	SBP(mmHg) MAE	DBP(mmHg) MAE
Schrumpf et al. [25]	2021	13.60	10.30
Chen et al. [7]	2023	12.35	9.54
Hamoud et al. [28]	2023	13.75	11.17
Ours(CNN+BiLSTM+GRU)	**2024**	**12.40**	**5.74**

5.1.4. Different subjects' blood pressure estimation results

In our study, we averaged the predicted blood pressure values based on multiple IPPG signals, each of length 250, for each participant in the test set. The resulting average is considered the predicted blood pressure for that participant, including both the average values for Systolic Blood Pressure (SBP) and Diastolic Blood Pressure (DBP). We further compared the accuracy of blood pressure estimation among different ethnicities, age groups, and skin tones in the test set, as depicted in Fig. 6.

In the test dataset, we included 9 individuals of European descent (Caucasian), 9 individuals of Asian descent (Yellow race), and 8 individuals of African descent (Black race), covering various skin tones across different ages and Fitzpatrick skin types ranging from 1 to 6. Upon analyzing populations of different ethnicities, we observed a higher accuracy in blood pressure prediction for individuals of Asian descent. However, we believe that the primary factor contributing to the differences in blood pressure prediction accuracy among different ethnicities lies in the samples
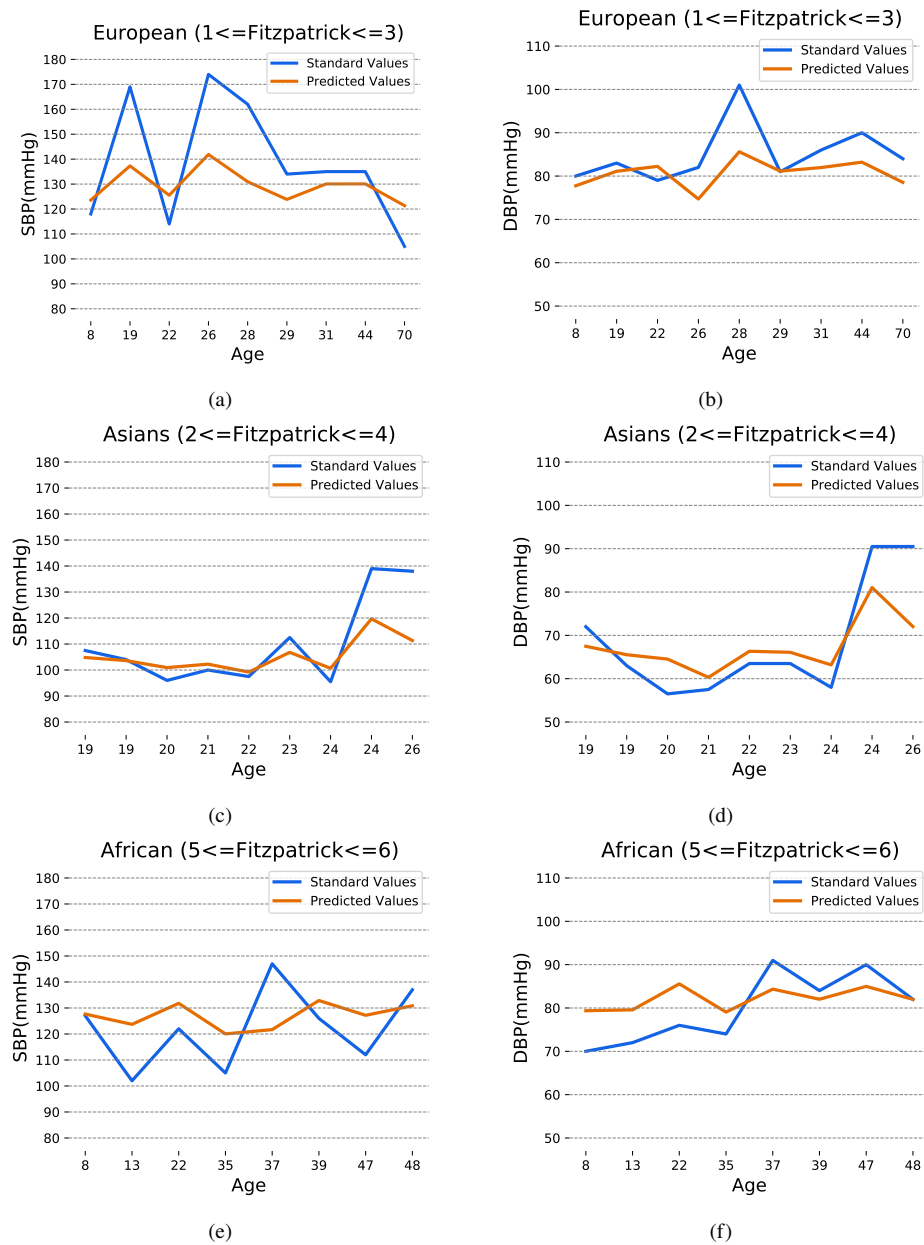
Fig. 6. Comparison of predicted and standardized values of blood pressure in the test set according to different races, skin colors, and ages. (a) and (b) represent the SBP and DBP blood pressure test results for Europeans (Caucasians), respectively. (c) and (d) represent the SBP and DBP blood pressure test results for Asians (Yellow race), respectively. (e) and (f) represent the results of SBP and DBP blood pressure tests for Africans (Black race), respectively.

representing high blood pressure and low blood pressure. When the test subjects' systolic blood pressure (SBP) and diastolic blood pressure (DBP) values are higher or lower, the limited number of samples for high blood pressure or low blood pressure in the dataset may lead to a decrease in

the model's performance when predicting abnormal samples. Especially when extreme samples occur, as illustrated in Fig. 6(a), the model may exhibit noticeable errors in predicting extreme samples due to the limited inclusion of such samples in the training set. However, the trend of the blood pressure prediction curve remains fundamentally consistent with standard blood pressure values.

Excluding the interference from extreme samples, we observed that the error in predicting blood pressure for individuals of African descent is slightly larger than that for individuals of Asian and European descent. We believe that for individuals with darker skin tones, there may be an impact on extracting effective IPPG signals. However, with the method proposed in this study, the error in blood pressure estimation for individuals of African descent remains within a certain range and does not result in significant deviations.

By examining the age distribution on the x-axis of Fig. 6, we observe that for individuals across different age groups, the non-contact blood pressure estimation method is not significantly affected by age differences in terms of estimating blood pressure errors. The primary factor influencing blood pressure estimation errors continues to be abnormal blood pressure samples. In summary, the proposed non-contact blood pressure estimation method in this study is capable of estimating blood pressure within a certain margin of error and is applicable to individuals of various races, skin tones, and age groups.

We have also generated Bland-Altman plots for the test results of 26 participants, as illustrated in Fig. 7. These plots include the mean line of errors and the 95% limits of agreement represented by ±1.96 times the standard deviation. Due to significant variations in SBP values among the participants and a scarcity of high SBP samples in the dataset, the accuracy in estimating SBP values demonstrates an overall higher precision for participants with lower blood pressure values. However, this phenomenon is less pronounced when estimating DBP values.
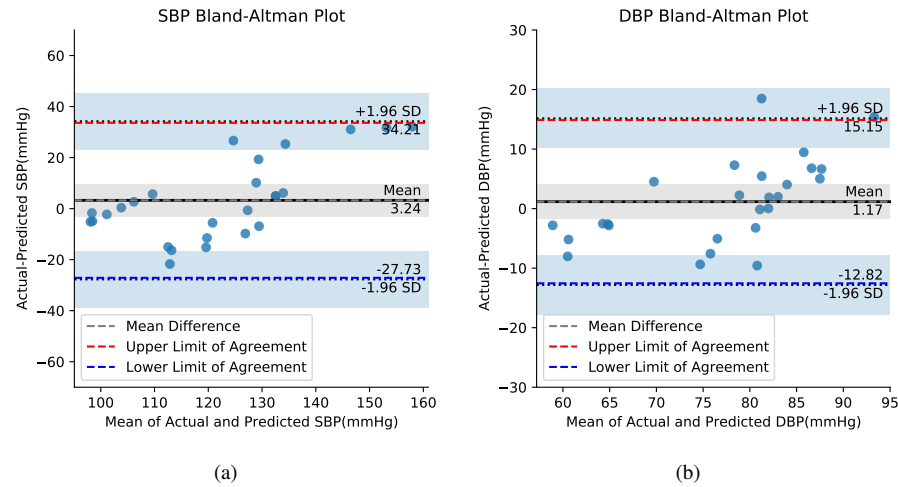
Fig. 7. The Bland-Altman plots for SBP and DBP of the 26 participants in the test set are respectively presented in (a) and (b).

## 5.2. Experiments on the MIMIC-II dataset

To assess the applicability of the method employed in this study to PPG signals, we applied the same four-layer filtering approach to 12,200 PPG signal data points from 80 patients in the MIMIC-II dataset. Similarly, we performed a 250-length cropping of the PPG signals. Unlike the previous process, during the EEMD filtering, we chose the fourth-order IMF component as the output. The bandpass filter in the filtering process had a frequency range of 0.2-2.5 Hz.

The preprocessed PPG signals were input into the CNN+BiLSTM+GRU network model. We randomly selected 80% of them as the training set and the remaining 20% as the test set. The training involved 550 rounds with a batch size of 64, a fixed learning rate of 0.001, and the Adam optimizer. The experimental results are presented in Table 4, showing MAE test results of 3.11 for SBP and 1.60 for DBP. In Table 4, we compared various methods from recent years and their accuracy in predicting SBP and DBP on public datasets. The results indicate that the signal preprocessing approach and network model adopted in this study exhibit good applicability to PPG signals. The errors for SBP and DBP are improved compared to recent literature, with a significant enhancement in DBP accuracy.

Table 4. Comparison of MAE performance of methods on public datasets in recent years.

| Method	Year Published	Dataset	Signal Type	MAE(mmHg)	
				SBP	DBP
CNN + LSTM [29]	2021	MIMIC-II	PPG	5.59	3.36
CNN + LSTM [30]	2021	MIMIC-II	PPG, ECG	4.41	2.91
Modified U-net [31]	2021	MIMIC- I, MIMIC-III	PPG	3.68	1.97
ARIU [32]	2022	MIMIC-III	PPG	4.75	2.81
CycleGAN [33]	2022	MIMIC-III	PPG	2.29	1.93
U-net+Unet [34]	2022	MIMIC-III	PPG	5.73	3.45
CNN + LSTM [26]	2023	MIMIC-III	PPG	3.64	2.39
MLR [35]	2023	MIMIC-II	PPG, ECG	4.46	4.20
LSTM-ANN [36]	2023	MIMIC-II	PPG	3.39	1.79
Ours(CNN+BiLSTM+GRU)	**2024**	**MIMIC-II**	**PPG**	**3.11**	**1.60**

We generated scatter plots with linear regression for the model-predicted blood pressure values and the standard blood pressure values, as depicted in Fig. 8. The standard values for SBP and DBP show a substantial linear correlation with the predicted values. The correlation coefficient for SBP is 0.95, and for DBP, it is 0.93.
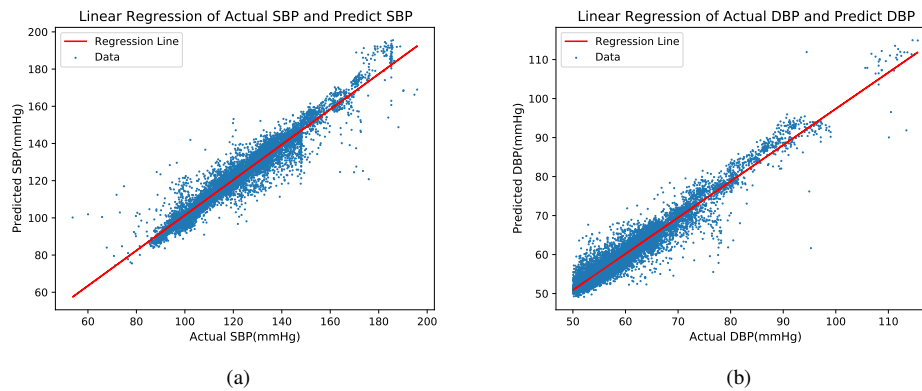
Fig. 8. The linear regression plots for SBP and DBP in the MIMIC-II dataset are shown in (a) and (b) respectively.

Bland-Altman plots were created concurrently for the standard and projected values of SBP and DBP, as shown in Fig. 9. The red and blue lines represent the upper and lower limits of 95% agreement, corresponding to 1.96 times the standard deviation, while the grey line indicates the mean error. The results indicate that the average prediction error for SBP is −0.37, with 95% of

SBP data errors concentrated within the interval [−9.96, 9.22]. Similarly, the average error for DBP is −0.15, and 95% of the errors for DBP fall within the interval [−4.79, 4.48].
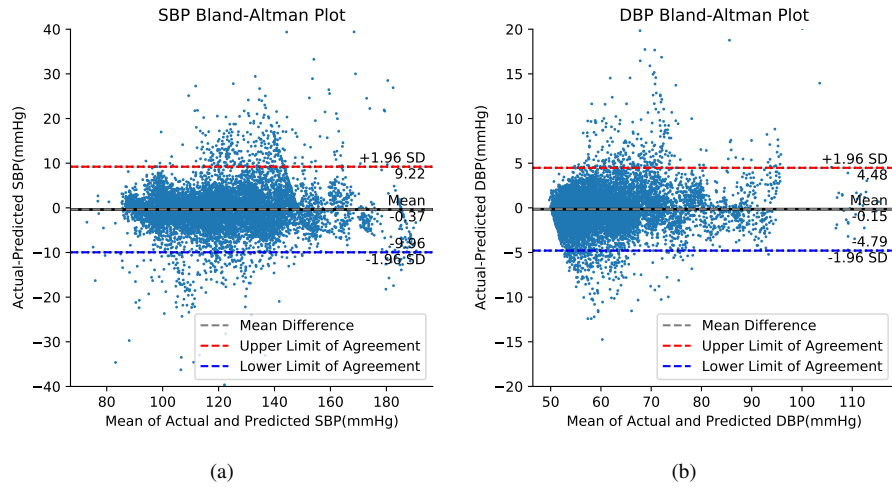
Fig. 9. The Bland-Altman plots for SBP and DBP in the MIMIC-II dataset are displayed in (a) and (b) respectively.

6. Conclusions

This study successfully utilized IPPG signals extracted from facial videos, subjected to a four-layer filtering process, and trained using the CNN+BiLSTM+GRU network model. This achievement allowed for the detection of both SBP and DBP values in diverse natural environments, encompassing individuals of various races and skin tones. Notably, in the measurement of DBP, the method exhibited measurement errors of less than 15 mmHg, 10 mmHg, and 5 mmHg in 91%, 83%, and 67% of cases, respectively, meeting the B-grade criteria outlined by the British Hypertension Society (BHS). While this approach represents a groundbreaking advancement compared to many current research methods, it falls slightly short of achieving the A-grade criteria set by BHS standards, which necessitate measurement errors of less than 15 mmHg, 10 mmHg, and 5 mmHg in 95%, 85%, and 60% of cases, respectively. Regrettably, in the measurement of SBP, the proposed method in this study does not yet meet the BHS standards. Improving the accuracy of non-contact SBP measurement remains a challenging task for future research.

Moreover, this study conducted tests on the proposed method for blood pressure detection using PPG signals. The test results on the publicly available MIMIC-II dataset indicate that the method employed in this study is not only applicable to blood pressure detection from PPG signals but also demonstrates an improvement in accuracy compared to recent research methods. Particularly noteworthy is the significant enhancement in the accuracy of DBP detection achieved by this method.

In summary, the VV-Small+CN-BP dataset utilized in this study encompasses facial videos of diverse populations captured in various scenarios, aiming to enhance the generality and practicality of current non-contact blood pressure detection methods. Additionally, this study innovatively introduces a four-layer filtering method for extracting pulse wave signals from IPPG signals. The constructed CNN+BiLSTM+GRU network model effectively improves the accuracy of current non-contact blood pressure detection. Importantly, this model is equally applicable to the detection of blood pressure using PPG signals.

References

1. F. D. Fuchs and P. K. Whelton, "High blood pressure and cardiovascular disease," Hypertension **75**(2), 285–292 (2020).
2. J. Slivnick and B. C. Lampert, "Hypertension and heart failure," Heart Failure Clinics **15**(4), 531–541 (2019).
3. P. Muntner, D. Shimbo, and R. M. Carey, "Measurement of blood pressure in humans: a scientific statement from the American Heart Association," Hypertension **73**(5), e35–e66 (2019).
4. S. Maqsood, S. Xu, M. Springer, *et al.*, "A benchmark study of machine learning for analysis of signal feature extraction techniques for blood pressure estimation using photoplethysmography (ppg)," IEEE Access **9**, 138817–138833 (2021).
5. A. A. Kamshilin, I. S. Sidorov, and L. Babayan, "Accurate measurement of the pulse wave delay with imaging photoplethysmography," Biomed. Opt. Express **7**(12), 5138–5147 (2016).
6. W. Xing, Y. Shi, and C. Wu, "Predicting blood pressure from face videos using face diagnosis theory and deep neural networks technique," Comput. Biol. Med. **164**, 107112 (2023).
7. Y. Chen, J. Zhuang, B. Li, *et al.*, "Remote blood pressure estimation via the spatiotemporal mapping of facial videos," Sensors **23**(24), 1 (2023).
8. S. Tulyakov, X. Alameda-Pineda, E. Ricci, *et al.*, "Self-adaptive matrix completion for heart rate estimation from face videos under realistic conditions," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2016), pp. 2396–2404.
9. Y. Li, M. Wei, and Q. Chen, "Hybrid d1dcnet using forehead ippg for continuous and noncontact blood pressure measurement," IEEE Sens. J. **23**(3), 2727–2736 (2023).
10. K. Iuchi, R. Miyazaki, and G. C. Cardoso, "Blood pressure estimation by spatial pulse-wave dynamics in a facial video," Biomed. Opt. Express **13**(11), 6035–6047 (2022).
11. S. Woo, J. Park, J.-Y. Lee, *et al.*, "Cbam: Convolutional block attention module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, (2018).
12. B. Lin, J. Tao, and J. Xu, "Estimation of vital signs from facial videos via video magnification and deep learning," iScience **26**(10), 107845 (2023).
13. J. Yang, J. Adu, H. Chen, *et al.*, "A facial expression recongnition method based on dlib, ri-lbp and resnet," J. Phys.: Conf. Ser. **1634**(1), 012080 (2020).
14. D. McDuff, "Camera measurement of physiological vital signs," ACM Comput. Surv. **55**(9), 1–40 (2023).
15. M. Saeed, M. Villarroel, and A. T. Reisner, "Multiparameter intelligent monitoring in intensive care ii (mimic-ii): a public-access intensive care unit database," Crit. Care Med. **39**(5), 952–960 (2011).
16. J. Yang, J. Adu, H. Chen, *et al.*, "A facial expression recongnition method based on dlib, ri-lbp and resnet," in *Journal of Physics: Conference Series*, vol. 1634 (IOP Publishing, 2020), p. 012080.
17. J. Harju, A. Tarniceriu, and J. Parak, "Monitoring of heart rate and inter-beat intervals with wrist plethysmography in patients with atrial fibrillation," Physiol. Meas. **39**(6), 065007 (2018).
18. C. Zhang, J. Tian, D. Li, *et al.*, "Comparative study on the effect of color spaces and color formats on heart rate measurement using the imaging photoplethysmography (ippg) method," Technol. Health Care **30**, 391–402 (2022).
19. G. Bishop and G. Welch, "An introduction to the Kalman filter," Proc of SIGGRAPH, Course **8**, 41 (2001).
20. Q. Li, R. Li, K. Ji, *et al.*, "Kalman filter and its application," in *2015 8th International Conference on Intelligent Networks and Intelligent Systems (ICINIS)*, (2015), pp. 74–77.
21. Z. Wei, K. G. Robbersmyr, and H. R. Karimi, "An EEMD aided comparison of time histories and its application in vehicle safety," IEEE Access **5**, 519–528 (2017).
22. X. Lang, Q. Zheng, Z. Zhang, *et al.*, "Fast multivariate empirical mode decomposition," IEEE Access **6**, 65521–65538 (2018).
23. W.-c. Wang, K.-w. Chau, D.-m. Xu, *et al.*, "Improving forecasting accuracy of annual runoff time series using arima based on eemd decomposition," Water Resour. Manag. **29**(8), 2655–2675 (2015).
24. J. Wang, G. Du, and Z. Zhu, "Fault diagnosis of rotating machines based on the emd manifold," Mech. Syst. Signal Process. **135**, 106443 (2020).
25. F. Schrumpf, P. Frenzel, and C. Aust, "Assessment of non-invasive blood pressure prediction from PPG and RPPG signals using deep learning," Sensors **21**(18), 6022 (2021).
26. N. Q. Mahardika, T. Y. N. Fuadah, D. U. Jeong, *et al.*, "PPG signals-based blood-pressure estimation using grid search in hyperparameter optimization of CNN–LSTM," Diagnostics **13**(15), 2566 (2023).
27. H. Mou and J. Yu, "Cnn-lstm prediction method for blood pressure based on pulse wave," Electronics **10**(14), 1664 (2021).

28. B. Hamoud, A. Kashevnik, W. Othman, *et al.*, "Neural network model combination for video-based blood pressure estimation: New approach and evaluation," Sensors **23**(4), 1753 (2023).

29. M. Rong and K. Li, "A multi-type features fusion neural network for blood pressure prediction based on photoplethysmography," Biomed. Signal Process. Control. **68**, 102772 (2021).

30. S. Baker, W. Xiang, and I. Atkinson, "A hybrid neural network for continuous and non-invasive estimation of blood pressure from raw electrocardiogram and photoplethysmogram waveforms," Comput. Methods Programs Biomed. **207**, 106191 (2021).

31. T. Athaya and S. Choi, "An estimation method of continuous non-invasive arterial blood pressure waveform using photoplethysmography: a U-net architecture-based approach," Sensors **21**(5), 1867 (2021).

32. M. Yu, Z. Huang, and Y. Zhu, "Attention-based residual improved u-net model for continuous blood pressure monitoring by using photoplethysmography signal," Biomed. Signal Process. Control. **75**, 103581 (2022).

33. M. A. Mehrabadi, S. A. H. Aqajari, A. H. A. Zargari, *et al.*, "Novel blood pressure waveform reconstruction from photoplethysmography using cycle generative adversarial networks," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (2022), pp. 1906–1909.

34. N. Ibtehaz, S. Mahmud, and M. E. H. Chowdhury, "Ppg2abp: Translating photoplethysmogram (PPG) signals to arterial blood pressure (ABP) waveforms," Bioengineering **9**(11), 692 (2022).

35. L.-H. Wang, K.-K. Sun, and C.-X. Xie, "Cuffless blood pressure estimation using dual physiological signal and its morphological features," IEEE Sens. J. **23**(11), 11956–11967 (2023).

36. N. Faris Ali and M. Atef, "An efficient hybrid LSTM-ANN joint classification-regression model for PPG based blood pressure monitoring," Biomed. Signal Process. Control. **84**, 104782 (2023).

37. H. Fang, J. Xiong, and L. He, "Fair non-contact blood pressure estimation using imaging photoplethysmography: dataset," Github, 2023, https://github.com/maoshanliulian/IPPG.