**Related Works for PPG-based BP Waveform Estimation**

Accurately estimating BP waveforms is essential for proper care and patient safety in various medical settings, particularly in intensive care units (ICUs) and for intraoperative hemodynamic monitoring. Standard cuff-based BP measurements only provide systolic blood pressure (SBP) and diastolic blood pressure (DBP) values, whereas the BP waveform can offer additional information about the patient's health and help estimate other physiological parameters. Therefore, several techniques have been proposed to accurately estimate BP waveforms, including deep learning-based methods. This literature review aims to provide a comprehensive overview of the current state of BP waveform estimation research, focusing on noninvasive deep learning-based methodologies. We explore the historical context of BP waveform measurements, standard clinical procedures, and recent innovations in noninvasive BP waveform monitoring and discuss the feasibility of employing these methods in both home and ICU settings.

***1) Significance of Noninvasive BP Waveform Estimation***

For more than five decades, indwelling arterial catheterization has served as the standard method for continuous arterial pressure measurement and waveform analysis at the bedside. This invasive BP waveform monitoring has been used widely in critical patients in ICU and operating rooms to facilitate the quick diagnosis of cardiovascular insufficiency and assess the response to the medication before the onset of hypotension or hypertension. The BP waveform is an invaluable source of information on an individual's dynamic cardiovascular state. Specifically, the unique venous BP waveform architecture relates closely to significant right cardiac activities, while each peak and valley in the BP waveform represents a distinct left heart function. BP waveform analysis can determine numerous critical parameters, including predicting vascular resistance, left ventricular stroke volume (SV), variation of SV, and pulse pressure values during positive-pressure respiration [30]. Real-time BP variation monitoring is possible using an arterial catheter equipped with a pressure transducer, allowing for earlier detection of intraoperative hypotension and reliable venous access for blood sampling. Despite its excellent accuracy, invasive BP waveform estimation is too intrusive for routine inspections due to the risks associated with extended periods of cannulation, patient discomfort, hematoma formation, and catheter-related infections.

The drawbacks of invasive BP waveform monitoring have led to extensive research efforts for noninvasive BP estimation. Noninvasive techniques establish a clinical balance between the arterial catheter and cuff-based techniques. However, only a few studies have attempted to establish methodologies for noninvasive BP waveform estimation. Vascular unloading techniques based on the radar [9] and ultrasound [10] can potentially estimate the BP waveforms. However, these methods face numerous technical challenges, and more validation is required to assess their accuracy and reliability.

The vast amount of medical data available in electronic health records provides an opportunity to utilize machine learning or deep learning algorithms with physiological signals like PPG [4-8], electrocardiogram (ECG) [13-18], or both for more convenient BP waveform estimation. However, none of the proposed noninvasive methods have been validated yet. Given the risks associated with invasive BP waveform estimation, developing reliable and accurate noninvasive techniques will be crucial for safer and more effective continuous BP monitoring in home and critical care settings. Therefore, further research should focus on validating these methods and determining their clinical utility. In conclusion, the recent development of noninvasive BP estimation methods has the potential to revolutionize continuous BP monitoring, enabling earlier detection of hypotension or hypertension and improved patient outcomes.

***2) PPG-based BP Waveform Estimation***

Deep learning algorithms that leverage biomedical signals have gained significant traction in estimating BP waveforms in recent years. These techniques have demonstrated promise in automatically learning essential features from a wealth of available biomedical data, making them increasingly popular for BP waveform estimation. The abundance of biomedical data available makes it possible to develop highly accurate and reliable deep learning-based techniques that can estimate BP waveforms noninvasively.

Our review discovered that most studies employed deep learning algorithms for BP waveform estimation using physiological signals like PPG and ECG. Arterial BP waveforms were used as the gold standard and reference values in these papers. As shown in Fig. S1, obtaining ECG signals requires body patches, BP waveforms can be acquired invasively via an arterial catheter from the radial artery or noninvasively from the brachial arteries, and PPG signals can be generated by a photodiode embedded in the finger, with multiple acquisition points available (e.g., wrist, toe, earlobe, etc.). The shape of the waveform varies depending on the measurement location, although fingertip PPG and BP waveforms have almost identical shapes.

图示

描述已自动生成

**Fig. S1.** Conventional photoplethysmograph (PPG) and arterial blood pressure (BP) measurement technique. Source from Athaya et al. [20].

Due to the structural similarity, most papers discussed in Table SI used PPG signals to estimate BP waveforms. Each of the works discussed thus far, as summarized in Table SI, demonstrates a more or less similar process. Table SII summarizes the advantages and disadvantages of deep learning-based methods.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Table SI Performance Summary of SOTA Works on BP Waveform Estimation (ME ± SDE) | | | | | | | |
| Algorithm | # of subject | Input | Input Length | Paradigm | Calibration | SBP | DBP |
| WNN [23] | > 90 (MIMIC) | PPG | - | - | - | 2.32 ± 2.91 | 1.92 ± 2.47 |
| LSTM [24] | 42 (MIMIC) | PPG | - | Sample-wise | - | - | - |
| NARX [15] | 15 (MIMIC) | PPG, ECG | 100 samples | Sample-wise | Free | -1.20 ± 9.10 | -0.60 ± 4.30 |
| DCAE [19] | 18 (Custom) | PPG | 5 s | Sample-wise | - | -1.66 ± 5.84 | 0.67 ± 3.98 |
| UNet [20] | 100 (MIMIC) | PPG | 2.048 s | Sample-wise | Free | 3.68 ± 4.42 | 1.97 ± 2.92 |
| RDAE [25] | 1, 227 (MIMIC) | PPG | 5 s | Subject-wise | Tuning | 1.28 ± 3.74 | -0.30 ± 3.41 |
| UNet [26] | 948 (MIMIC) | PPG | 32 samples | Sample-wise | - | 0.59 ± 4.78 | 0.43 ± 4.78 |
| V-Net [21] | 264 (MIMIC) | PPG, ECG,  hand-crafted features | 4 s | Subject-wise | Free | 4.30 ± 6.53 | -3.11 ± 4.57 |
| 110 (UCLA) | Fine-tuning in MIMIC | 2.40 ± 5.62 | -2.50 ± 3.79 |
| CycleGAN [27] | 92 (MIMIC) | PPG | 2.048 s | Subject-wise | - | 0.67 ± 4.52 | 1.78 ± 4.67 |
| KD-Informer [22] | 467 (Mindray) | PPG, hand-crafted features | 8 s | Subject-wise | Free | 0.02 ± 5.93 | 0.01 ± 3.87 |
| 241 (MIMIC) | Fine-tuning in MIMIC | 0.03 ± 6.38 | 0.02 ± 4.49 |
| PPG2BP-cGAN | 683 (Mindray) | PPG | 9 s | Subject-wise | Free | 0.73 ± 4.83 | -0.46 ± 2.77 |
| 200 (MIMIC) | Fine-tuning in Mindray | 0.72 ± 4.34 | 0.41 ± 2.48 |
| Note, WNN denotes wavelet neural network; LSTM denotes long short-term memory; NARX denotes nonlinear autoregressive models with exogenous input; DCAE denotes deep convolutional autoencoder; RDAE denotes deep convolutional autoencoder; CycelGAN denotes cycle generative adversarial network; KD-Informer denotes Transformer-based method with knowledge distillation. "-" is used where the information is not provided or the item is not used. | | | | | | | |

|  |  |  |
| --- | --- | --- |
| Table SII Advantages and Disadvantages of SOTA Works on waveform Estimation | | |
| Algorithm | Advantages | Disadvantages |
| WNN [23] | Only a PPG signal is needed. | High computational complexity. Data redundancy. |
| LSTM [24] | Only a PPG signal is needed. | The length of the input is not defined. The network architecture and overall process are not described. Only one-point output with respect to multiple input points. The same subjects are used for training and testing. |
| NARX [15] | A feedback loop is used to predict BP values. BP waveform estimation is shown from the ECG signal as less sensitive to artifacts. | BP data is needed for model training. Few subjects. The delay removal process is not applicable to all cases. Different ECG, PPG, and BP peak ranges were not unified while training. cCoss-correlation analysis was performed to quantify any delay between the predicted and the measured BP. Two BP waveform points are needed for the input of the model. The same subjects are used for training and testing. |
| DCAE [19] | Only a PPG signal is needed. Custom data has been used. The number of subjects is less. | The same subjects are used for training and testing. Custom data has been used. The number of subjects is less. |
| UNet [20] | Only a PPG signal is needed. A comparatively good result is obtained using only one model. | The same subjects are used for training and testing. |
| RDAE [25] | Only a PPG signal is needed. The proposed model requires fewer parameters than other methods. Subjects of training and testing sets are different. | The calibrated model gives a better result. |
| UNet [26] | The model is implemented on a Raspberry Pi 4 device. Only a PPG signal is needed. | The device implementation process is not described. PPG signal artifacts can provide wrong results. The same subjects are used for training and testing. |
| V-Net [21] | SBP and DBP estimation process is shown. Results were shown for two different datasets. Training and validation sets include different patients. | Both the PPG and ECG waveforms and several constants are needed as input. |
| CycleGAN [27] | Only a PPG signal is needed. 5-fold cross-validation is obtained with the data. | PPG signal artifacts can provide wrong results. A constant value of λ is used, which is set to 10. |
| KD-Informer [22] | Results were shown for two different datasets. Training and validation sets include different patients. SBP and DBP estimation process is shown. 5-fold cross-validation is obtained with the data. | Features that need to be defined manually. A two-stage knowledge distillation is required. A feature selection algorithm needs to be designed. A portion of custom data is used. |
| PPG2BP-cGAN | Only a PPG signal is needed. Results were shown for two different datasets. Training and validation sets include different patients. SBP and DBP estimation process is shown. Different BP categories can be effectively identified. 5-fold cross-validation is obtained with the data. | A portion of custom data is used. Confidence coefficients for the BP categories were not available. |
| Note, WNN denotes wavelet neural network; LSTM denotes long short-term memory; NARX denotes nonlinear autoregressive models with exogenous input; DCAE denotes deep convolutional autoencoder; RDAE denotes deep convolutional autoencoder; CycelGAN denotes cycle generative adversarial network; KD-Informer denotes Transformer-based method with knowledge distillation. | | |

In noninvasively estimating BP waveforms, various machine learning algorithms have been proposed. Li et al. [23] proposed an optimized wavelet neural network (WNN) with PPG signals as input, which was preferable to a three-layered WNN because of its simpler structure of cache hidden layer nodes without multipliers. Inhomogeneous resilient backpropagation was also implemented to reduce computational complexity and convergence time. However, as mentioned in Table SII, this model has high computational complexity and data redundancy.

Sideris et al. [24] employed a Long Short-Term Memory (LSTM) recurrent neural network as the input of the proposed model, but no detailed network architecture and process were described. The training process is lengthy and inefficient for implementing a device, and the model is not generalizable since it is trained individually for each patient, producing a single BP point as output when multiple inputs are used.

Landry et al. [15] proposed a nonlinear autoregressive model with exogenous input (NARX) for estimating BP waveforms using ECG signals. The ECG and PPG signals achieved the highest Pearson correlation coefficient. However, two sets of BP data are required for model training, and the delay removal procedure is inappropriate for all circumstances. During training, the ECG, PPG, and BP peak ranges were not unified, and cross-correlation analysis was used to quantify the difference between predicted and measured BP.

Athaya et al. [20] proposed a one-dimensional (1D) modified UNet network for estimating BP waveforms, which differs from Vardhan et al. [26] due to the signal preprocessing techniques used. Although Vardhan et al. [26] demonstrated the feasibility of implementing their proposed model on a Raspberry Pi 4 device with an inference time of 4.25 ms, the numerous parameters required for the UNet model make device implementation difficult. Qin et al. [25] proposed a deep autoencoder based on regularized convolution, RDAE, which requires fewer parameters than alternative methods and provides estimated BP waveforms that can then be used to predict SBP, DBP, and MAP.

In [19], a comparison of two deep convolution autoencoders named LeNet-5 and UNet to estimate the BP waveforms is shown. To investigate data generalization, the cross-validation (CV) technique was used. The results indicate that the UNet outperforms other estimation methods for SBP values. Meanwhile, the LeNet-5 is marginally more accurate at estimating DBP values. Finally, a genetic algorithm-based optimization deep convolutional autoencoder (GDCAE) is used to optimize the ensemble of CV models. The findings show that the GDCAE outperforms the LeNet-5 and the UNet. Thus, this review discusses the outcome of the best-performing model GDCAE. However, combining two deep learning algorithms to obtain two distinct values requires many parameters, which is inefficient computationally. Additionally, two values can be obtained by blending two different models, but no optimized model for predicting BP waveforms is shown.

Meanwhile, a 1D V-Net deep learning algorithm was proposed by Hill et al. [21] for BP waveform estimation, utilizing ECG and PPG signals as inputs along with several constant values. The model is designed to learn the residual error between the PPG and BP waveforms. However, the model requires a large number of input variables and noninvasive BP measurements.

Recent works by Mehrabadi et al. [27] and Ma et al. [22] offer novel techniques for BP waveform estimation. Mehrabadi et al. [27] proposed using a cycle generative adversarial network (CycleGAN) to estimate BP waveforms with PPG data as input, utilizing both a generator and discriminator network. Meanwhile, Ma et al. [22] proposed a Transformer-based method with knowledge distillation (KD-Informer) for BP waveform estimation, integrating hand-crafted features extracted from PPG signals using a backward elimination algorithm into the knowledge transfer branch for superior performance in a subject-wise paradigm.

***3) Data Preprocessing Pipeline***

Preprocessing data is critical for deep learning methods to provide an accurate model estimation. Even when similar algorithms and data are used, subtle differences in preprocessing techniques result in noticeable differences in the results, as shown in Table SIII. Typically, data preprocessing entails one or more of the following:

1. Segmenting the data to train the model.
2. Removing erroneous physiological signals that are inaccurate for measurement.
3. Filtering the physiological signals to remove the baseline wandering and high-frequency noises.
4. Normalizing inputs and outputs for accurate training of the model.

Physiological signals contain a variety of artifacts. If those artifacts are used to train deep learning algorithms, the algorithm may produce incorrect results. As a result, the erroneous data containing artifacts must be deleted. As all of the papers listed in Table SI used either PPG or ECG signals or a combination of the two, signal filtering is necessary to remove high-frequency noise and extract the necessary portion of the signal for the algorithm. Normalization of the data is an additional step in data preprocessing. The signals must be normalized to eliminate the range difference between different physiological signals. This is accomplished through the use of a variety of different normalization techniques. Table SIII illustrates the differences in the data preprocessing methods used by various works.

|  |  |  |
| --- | --- | --- |
| Table SIII Different Data Preprocessing Techniques of SOTA Works on waveform Estimation | | |
| Algorithm | Preprocessing Steps | Normalization Methods |
| WNN [23], LSTM [24],  UNet [26] | Filtering the physiological signals Removing erroneous physiological signals | Normalization is not implemented |
| NARX [15], DCAE [19] | Removing erroneous physiological signals | Normalization is not implemented |
| UNet [20] | Filtering the physiological signals Segmentation  Phase matching Removing erroneous physiological signals Normalization |  |
| V-Net [21] | Filtering the physiological signals Segmentation Removing erroneous physiological signals Normalization |  |
| RDAE [25] | Filtering the physiological signals Removing the erroneous portion of physiological signals Segmentation Normalization |  |
| CycleGAN [27] | Removing the erroneous portion of physiological signals Filtering the physiological signals Segmentation | Normalization is not implemented |
| KD-Informer [22],  PPG2BP-cGAN | Resampling Filtering the physiological signals Segmentation Phase matching Normalization |  |
| Note, WNN denotes wavelet neural network; LSTM denotes long short-term memory; NARX denotes nonlinear autoregressive models with exogenous input; DCAE denotes deep convolutional autoencoder; RDAE denotes deep convolutional autoencoder; CycelGAN denotes cycle generative adversarial network; KD-Informer denotes Transformer-based method with knowledge distillation. | | |

The studies reviewed in this paper emphasize the importance of preprocessing techniques in accurately modeling BP waveforms using PPG signals. WNN [23] used a low-pass filter for the Daubechies wavelet to process PPG signals with scales 21 and 22, while Sideris et al. [24] removed baseline drift from both PPG and BP signals. Landry et al. [15] excluded noisy data from the dataset by using double derivation to determine the signal's erroneous portion. Daubechies wavelet denoising, as described in [23], involved setting the coefficients for decomposition to zero to neglect too low and too high-frequency components. Mean normalization was used for the PPG signal. For segmentation, Landry et al. [15] used an input sample size of 100 to obtain one BP point, while [19] used a fixed length of 5 s. In [20], signals were segmented into 350 samples with a 100-sample overlap using a bandpass Equiripple FIR filter with a frequency range of 0.5-8 Hz. Vardhan et al. [26] excluded irregular BP waveforms and downsampled the data to meet computational requirements, while Hill et al. [21] employed a comprehensive preprocessing technique that included filtering, normalization, segmentation, and artifact removal. Fast Fourier transform (FFT) was used to remove unimportant information in [27], and a standardized pipeline was established by KD-Informer [22]. The significance of normalization was highlighted since it could eliminate the interference of PPG signal amplitude due to individual differences. By using various preprocessing techniques, these works attempted to improve the performance of models based on PPG signals for BP waveform prediction.

***4) Data Availability***

The effectiveness of deep learning algorithms is highly dependent on the quantity and quality of data used for training, validation, and testing. Table SIV summarizes the use of data for different methods in the reviewed studies. The majority of papers utilized data from the MIMIC, MIMIC II, and MIMIC III waveform datasets. However, the number of subjects and the total amount of data collected varied across the studies. To determine the generalizability of the proposed algorithm, several papers used K-fold cross-validation (CV), where each fold's training data was partitioned into training and validation data for the K-fold CV. In contrast, some papers divided the data set into the train, validation, and test sets, while others split the data at the subject level. Notably, the number of patients, the data size, and the method of data splitting can all impact the results obtained from the proposed algorithms.

Moreover, the diversity of the dataset is also crucial to ensure the generalizability of the model. The more diverse the data set, the more accurately the proposed algorithm can predict the BP waveform from PPG signals. Various demographic factors, such as age, sex, and ethnicity, may affect the transferability of the model to different populations. Therefore, it is important to include data from different populations to evaluate the proposed algorithms' performance in different clinical settings. Furthermore, it is crucial to consider the quality of the recorded data to avoid noisy or erroneous signals. Proper curation of the dataset can improve the accuracy of the algorithm and enhance its clinical utility.

In conclusion, the availability of high-quality data is vital for developing effective deep learning algorithms for predicting BP waveforms from PPG signals. The proper division of data for training, validation, and testing, along with the use of K-fold cross-validation, can also impact the generalizability of the model. Including data from diverse populations is essential to evaluate the algorithm's performance in different clinical settings. Finally, carefully considering the dataset's quality is necessary to ensure accurate model predictions.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Table SIV Summary of the Datasets Used to Train, Validate, and Test the Deep Learning Models | | | | | |
| Algorithm | Dataset | # of Subject | Total Data (hours) | K-Fold Cross-Validation | Train:Val:Test |
| WNN [23] | MIMIC | >90 | - | No | Not given |
| LSTM [24] | MIMIC | 42 | - | No | 80:10:10 (in total data) |
| NARX [15] | MIMIC II | 15 | - | No | 70:15:15 (in total data) |
| DCAE [19] | Custom | 18 | ≈50.72 | Yes (10 Folds) | 85:-:15 (in total data) |
| UNet [20] | MIMIC, MIMIC III Waveform | 100 | ≈195 | No | 70:15:15 (in total data) |
| RDAE [25] | MIMIC II | 1227 | ≈54.53 | No | 60:20:20 (subject-wise) |
| UNet [26] | MIMIC II Waveform dataset | 948 | ≈353.5 | Yes (10 Folds) | 78.58:-:21.42 |
| V-Net [21] | MIMIC III, UCLA | MIMIC-264, UCLA-110 | ≈2516.48 | No | 66:-:33 (in subjects of MIMIC) |
| CycleGAN [27] | MIMIC II Waveform dataset | 92 | ≈7.67 | Yes (5 Folds) | 80:-:20 |
| KD-Informer [22] | MIMIC III, Custom | MIMIC-241, Custom-467 | ≈73.30 | Yes (5 Folds) | 80:-:20 (subject-wise) |
| PPG2BP-cGAN | MIMIC III, Mindray | MIMIC-200, Custom-683 | ≈81.79 | Yes (5 Folds) | 80:-:20 (subject-wise) |
| Note, WNN denotes wavelet neural network; LSTM denotes long short-term memory; NARX denotes nonlinear autoregressive models with exogenous input; DCAE denotes deep convolutional autoencoder; RDAE denotes deep convolutional autoencoder; CycelGAN denotes cycle generative adversarial network; KD-Informer denotes Transformer-based method with knowledge distillation. "-" is used where the information is not provided or the item is not used. | | | | | |

***5) Results Comparison***

All reviewed papers included figures to visualize their estimated BP waveforms, demonstrating the effectiveness of their proposed algorithms. To quantitatively evaluate the performance of the proposed algorithms, various performance metrics were utilized, as summarized in Table SV. The most commonly derived metrics were SBP and DBP values, with some studies also calculating MAP. Mean absolute error (MAE), mean error (ME), standard deviation error (SDE), root mean square error (RMSE), Pearson's correlation coefficient (R2), and average mean squared error (AMSE) were frequently used as performance metrics.

Furthermore, to evaluate the proposed algorithms' performance more objectively, most papers compared their results to established standards, such as the British Hypertension Society (BHS) and the Association for the Advancement of Medical Instrumentation (AAMI). According to the AAMI standard, the ME should be within ±5 mmHg, and the SDE should be less than or equal to 8 mmHg for large datasets involving more than 85 subjects. Most of the papers in this review defined the AAMI standard using the ME, while some used the MAE [20, 23]. However, it is crucial to note that several studies have shown that using only the ME can lead to incorrect results, as a lower ME may result in a higher MAE [19, 22, 25, 26]. Therefore, Table SV presents the outcomes of the proposed works based on both the ME and the MAE according to the AAMI standard.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table SV Performance Summary of the SOTA BP Waveform Estimation Methods | | | | | | | | |
| Algorithm | # of subject | Paradigm | Waveform | SBP | DBP | MAP | BHS Grade | AAMI |
| WNN [23] | > 90 (MIMIC) | - | Mean: 3.4094 AMSE: 4.4797 | MAE ± SDE: 2.32 ± 2.91 | MAE ± SDE: 1.92 ± 2.47 | - | - | Failed (subjects < 85) |
| LSTM [24] | 42 (MIMIC) | Sample-wise | RMSE: 6.042 ± 3.26 R2: 0.95 MAE: 5.98 ME: -0.214 | RMSE: 2.575 | RMSE: 1.977 | - | - | Failed (subjects < 85) |
| NARX [15] | 15 (MIMIC) | Sample-wise | - | ME ± SDE: -1.20 ± 9.10 | ME ± SDE: -0.60 ± 4.30 | - | - | Failed (subjects < 85) |
| DCAE [19] | 18 (Custom) | Sample-wise | RMSE: 3.46 MAE: 2.33 R2: 0.984 | ME ± SDE: -1.66 ± 5.84 MAE: 2.54 R2: 0.981 | ME ± SDE: 0.67 ± 3.98 MAE: 1.48 R2: 0.979 | - | - | Failed (subjects < 85) |
| UNet [20] | 100 (MIMIC) | Sample-wise | R2: 0.993 | MAE ± SDE: 3.68 ± 4.42 RMSE: 5.75 R2: 0.976 | MAE ± SDE: 1.97 ± 2.92 RMSE: 3.52 R2: 0.970 | MAE ± SDE: 2.17 ± 3.06 RMSE: 3.75 R2: 0.976 | A | Passed (MAE, SDE)\* |
| RDAE [25] | 1, 227 (MIMIC) | Subject-wise | - | ME ± SDE: 1.28 ± 3.74 MAE: 5.424 | ME ± SDE: -0.30 ± 3.41 MAE: 3.144 | ME ± SDE: 1.28 ± 3.74 | SBP: B | Failed (MAE) |
| UNet [26] | 948 (MIMIC) | Sample-wise | - | ME ± SDE: 0.59 ± 4.78 MAE: 5.16 | ME ± SDE: 0.43 ± 4.78 MAE: 2.89 | ME ± SDE: 0.59 ± 4.78 | SBP: B | Failed (MAE) |
| V-Net [21] | 264 (MIMIC) | Subject-wise | RMSE: 6.961 R2: 0.947 | ME ± SDE: 4.30 ± 6.53 | ME ± SDE: -3.11 ± 4.57 | - | - | Passed (ME, SDE)† |
| 110 (UCLA) | ME ± SDE: 2.40 ± 5.62 | ME ± SDE: -2.50 ± 3.79 | - | - | Passed (ME, SDE)† |
| CycleGAN [27] | 92 (MIMIC) | Subject-wise | - | ME ± SDE: 0.67 ± 4.52 | ME ± SDE: 1.78 ± 4.67 | - | A | Failed (subjects < 85) |
| KD-Informer [22] | 467 (Mindray) | Subject-wise | RMSE: 2.78# MAE: 2.23# R2: 0.99# | ME ± SDE: 0.01 ± 6.24 MAE: 4.18 | ME ± SDE: 0.01 ± 4.45 MAE: 3.00 | - | - | Passed |
| 241 (MIMIC) | ME ± SDE: 0.03 ± 6.38 MAE: 4.30 | ME ± SDE: 0.02 ± 4.49 MAE: 3.13 | - | - | Passed |
| PPG2BP-cGAN | 683 (Mindray) | Subject-wise | RMSE: 3.46 MAE: 2.33 R2: 0.984 | ME ± SDE: 0.73 ± 4.83 MAE: 3.41 R2: 0.980 | ME ± SDE: -0.46 ± 2.77 MAE: 2.40 R2: 0.941 | ME ± SDE: 0.48 ± 3.08 MAE: 2.61 | A | Passed |
| 200 (MIMIC) | ME ± SDE: 0.72 ± 4.34 MAE: 3.15 | ME ± SDE: 0.41 ± 2.48 MAE: 2.21 | ME ± SDE: 0.46 ± 2.39 MAE: 2.12 | A | Passed |
| Note, WNN denotes wavelet neural network; LSTM denotes long short-term memory; NARX denotes nonlinear autoregressive models with exogenous input; DCAE denotes deep convolutional autoencoder; RDAE denotes deep convolutional autoencoder; CycelGAN denotes cycle generative adversarial network; KD-Informer denotes Transformer-based method with knowledge distillation. "-" is used where the information is not provided or the item is not used. \* indicates that ME is not available. † indicates that MAE is not available. # indicates that only one random sample of evaluation metrics is shown. No unit for R2, mmHg for Others. | | | | | | | | |

Li et al. [23] demonstrated the performance of their proposed algorithm with mean and average mean squared error, estimated against the MIMIC dataset waveforms, which are 3.4094 mmHg and 4.4797 mmHg, respectively. The mean absolute error (MAE) and standard deviation error (SDE) for their approach fell within the range specified by the AAMI standard. Sideris et al. [24] presented the root mean square error (RMSE), MAE, and mean square error for their estimated waveforms, and for SBP and DBP, only the RMSE was displayed. In [15], the authors explored three different methods to estimate BP waveforms: using only the ECG, only the PPG, or both. However, no precise numerical values for the performance metrics were provided. The Deep Convolutional Autoencoder (DCAE) proposed in [19] produced the best results among the three approaches described; nonetheless, the number of subjects tested was too small to meet the AAMI standard. Athaya et al. [20] compared the estimated waveform figures to the average Pearson's correlation coefficient (R2) value for BP waveforms and met both standards.

In [25], the estimated BP waveform was compared to the ground-truth BP waveform based on the RDAE, both with and without calibration. Five separate cases were presented to demonstrate the BP waveform results: those displaying a clear dicrotic notch, those without one, hypotension, a clearly abnormal cycle with a high degree of BP fluctuation, and those without a clear dicrotic notch. The calibrated result was superior to the uncalibrated result, and Table SV displays the best result that satisfies the AAMI error range for ME values but not for MAE values. This work received a grade of B for estimating SBP according to the BHS standards. Vardhan et al. [26] presented an example of an estimated BP waveform, but the obtained BP values did not conform to either of the standards. Hill et al. [21] reported their findings from two distinct datasets, with RMSE and correlation coefficient displayed for waveforms, and ME and SDE given for SBP and DBP values. Their work conformed to the AAMI standard according to ME values [21].

Mehrabadi et al. [27] did not display any results for waveforms or estimated waveforms. Results for SBP and DBP were presented, and they satisfied the respective standards when the given value was used. The results of Ma et al. [22] presented the BP waveform estimates, plotted for a randomly selected sample concerning the RMSE, MAE, and R2 of 7.27 mmHg, 2.23 mmHg, and 0.99, respectively. However, the average results of BP waveform estimation were not provided. The results of both SBP and DBP met the AAMI standard, but the MAP results were not provided nor compared with the BHS standard.