KD-Informer: A Cuff-Less Continuous Blood Pressure Waveform Estimation Approach Based on Single Photoplethysmography

Chenbin Ma, Peng Zhang, Fan Song, Yangyang Sun, Guangda Fan, Tianyi Zhang, Youdan Feng, and Guanglei Zhang, *Member, IEEE*

***Abstract*—Ambulatory blood pressure (BP) monitoring plays a critical role in the early prevention and diagnosis of cardiovascular diseases. However, cuff-based inflatable devices cannot be used for continuous BP monitoring, while pulse transit time or multi-parameter-based methods require more bioelectrodes to acquire electrocardiogram signals. Thus, estimating the BP waveforms only based on photoplethysmography (PPG) signals for continuous BP monitoring has essential clinical values. Nevertheless, extracting useful features from raw PPG signals for fine-grained BP waveform estimation is challenging due to the physiological variation and noise interference. For single PPG analysis utilizing deep learning methods, the previous works depend mainly on stacked convolution operation, which ignores the underlying complementary time-dependent information. Thus, this work presents a novel Transformer-based method with knowledge distillation (KD-Informer) for BP waveform estimation. Meanwhile, we integrate the prior information of PPG patterns, selected by a novel backward elimination algorithm, into the knowledge transfer branch of the KD-Informer. With these strategies, the model can effectively capture the discriminative features through a lightweight architecture during the learning process. Then, we further adopt an effective transfer learning technique to demonstrate the excellent generalization capability of the proposed model using two independent multicenter datasets. Specifically, we first fine-tuned the KD-Informer with a large and high-quality dataset (Mindray dataset) and then transferred the pre-trained model to the target domain (MIMIC dataset). The experimental test results on the MIMIC dataset showed that the KD-Informer exhibited an estimation error of 0.02 $\pm$ 5.93 mmHg for systolic BP (SBP) and 0.01 $\pm$ 3.87 mmHg for diastolic BP (DBP), which complied with the association for the advancement of medical instrumentation (AAMI) standard. These results demonstrate that the KD-Informer has high reliability and elegant robustness to measure continuous BP waveforms.**

***Index Terms*—Blood pressure, photoplethysmography, sequence learning, knowledge distillation.**

I. Introduction

ACCORDING to the World Health Organization (WHO), an average of 1.65 million people dies from cardiovascular diseases (CVD) each year. Longitudinal follow-up of blood pressure (BP) parameters can help improve prognostic outcomes and thus reduce mortality from CVD [1]. According to a recent large cohort study, hypertension awareness and treatment rates, a common early symptom of CVD, were only 36% and 22.9%, respectively [2]. BP can be accurately measured by an invasive method involving puncturing a blood vessel through a specialized intervention where a pressure transducer observes fluid pressure in real-time. However, this approach can cause severe pain, and monitoring in daily life is impractical. Therefore, devices that provide long-term non-invasive monitoring of BP capabilities are essential for the early diagnosis of CVD patients [3]. Routine clinical BP screening is commonly performed with non-invasive inflatable cuff-based measurements, including auscultation and oscillometric techniques.

However, cuff-based methods do not facilitate long-term monitoring due to inflating and blocking the blood vessels during testing. Therefore, many studies have proposed unobtrusive cuff-less BP measurement methods to achieve portable, comfortable, and long-term BP monitoring. There are even studies that combine physiological signals such as the electrocardiogram (ECG) [4]–[9], ballistocardiogram (BCG) [10], photoplethysmography (PPG) [11], etc. In particular, pulse wave velocity (PWV) theory is usually based on PPG signals, which can be acquired by the pulse oximeter.

Due to the convenience of measurement, PWV-based methods have been the most popular cuff-less BP estimation [11]. Specifically, the pulse arrival time (PAT) and pulse transient

Manuscript received 15 December 2021; revised 2 May 2022; accepted 5 June 2022. Date of publication 14 June 2022; date of current version 5 May 2023. This work was supported in part by the Fundamental Research Funds for Central Universities, in part by the National Natural Science Foundation of China under Grant 61871022, in part by the Beijing Natural Science Foundation under Grant 7202102, and in part by the 111 Project under Grant B13003. *(Corresponding author: Guanglei Zhang.)*

Chenbin Ma is with the Beijing Advanced Innovation Center for Biomedical Engineering, School of Biological Science and Medical Engineering, Beihang University, Beijing 100191, China, and also with the Shenyuan Honors College, Beihang University, Beijing 100191, China (e-mail: machenbin@buaa.edu.cn).

Peng Zhang, Fan Song, Yangyang Sun, Guangda Fan, Tianyi Zhang, Youdan Feng, and Guanglei Zhang are with the Beijing Advanced Innovation Center for Biomedical Engineering, School of Biological Science and Medical Engineering, Beihang University, Beijing 100191, China (e-mail: pengzhang@buaa.edu.cn; fansong@buaa.edu.cn; 1251976963@qq.com; a641261717@buaa.edu.cn; zhangtianyi@buaa.edu.cn; emilyfeng@buaa.edu.cn; guangleizhang@buaa.edu.cn).

Digital Object Identifier 10.1109/JBHI.2022.3181328

time (PTT), the time interval of pulse wave propagation in the vasculature, can be extracted in conjunction with ECG, and thus BP can be estimated based on PWV theory. Ma *et al.* [12] characterized the human arteries through Fung hyper-elastic model and established a simple formula between PWV and the pressure pulse. However, it has not been adequately demonstrated that PTT can accurately estimate BP. Many factors can influence the correlation between BP and PWV parameters, e.g., smooth muscle contraction and viscous effects, thus requiring frequent calibration with a cuff-based BP monitor [5], [13]. Furthermore, ECG measurements require access to electrodes or two-handed contact to form a closed-loop measurement, which is inconvenient. Therefore, PWV-based cuff-less PPG/ECG BP monitors are still not universally used in the medical and consumer communities.

In recent years, traditional machine learning methods have benefited from their strong nonlinear fitting ability and better BP prediction accuracy [14]–[18]. However, the feature extraction stage is necessary for machine learning methods. It primarily relies on the experience of experts, making it easy to suffer from missing or redundant information and leading to excessive errors in predicting BP values. Therefore, recent studies have been using deep learning methods to extract features directly from the raw PPG waveforms [10], [19]–[22], thus avoiding the deficiencies of manually defined features or shallow learning algorithms in fitting BP. In 2019, Baek et al. [5] proposed an end-to-end fully convolutional network model to predict the BP value. However, the standard deviation of the error (SD) of the systolic BP (SBP) was 12.80 mmHg, and the SD of the diastolic BP (DBP) was 7.54 mmHg. Although current PPG-based deep learning methods have improved the prediction accuracy and generalization ability, they are often limited by the quality of the dataset [23], the experimental design [20], and the bottlenecks in the network framework itself. For example, too strict quality judgment guidelines have significantly reduced the usability of the data, Slapničar *et al.* [23] even censored 60% of the entire dataset. In addition, the computational burden of neural networks also needs to be considered. Due to the different physiological significance, methods for BP value prediction require separate modeling for SBP and DBP values, which also increases the computational complexity. Fortunately, knowledge distillation (KD), as a novel model compression method, has the potential to enable model deployment to mobile devices [24].

This paper aims to formulate a novel Informer architecture with KD strategies (KD-Informer) for BP waveform estimation by simultaneously optimizing deep features and integrating morphological features from raw PPG signals. The KD-Informer differs from the existing BP prediction methods that only consider pulse cycles or statistical features. Moreover, compared to other studies that directly output discrete BP values [4]–[6], [10], [20], the KD-Informer continuously estimates the BP waveform, which is essential for reducing the risk of intraoperative hypotension [25]. The main contributions of this paper are summarized as follows:

1) For PPG analysis, we applied Transformer as the backbone of the network for the first time for real-time prediction of BP waveforms by making full use of the advantages of the Transformer model in timing dependencies analysis. At the same time, we extracted morphological features from the PPG signal as prior features and used the ResNet with a squeeze-and-excitation module (SE-ResNet) to fuse them with deep features. Introduced prior information can facilitate the model to utilize more discriminative features during the learning process and thus achieve higher prediction accuracy.
2) To reduce memory consumption and achieve lighter computing, we adopted a response-based KD strategy to transfer the knowledge learned by the teacher Informer to the lightweight student Informer.
3) Backward elimination feature selection (BEFS) and feature fusion methods were designed. These algorithms help obtain a proper sparse feature subset, which can be efficiently combined with deep features learned by KD-Informer during the learning process. In particular, we designed the modified longest common sub-sequence (MLCSS) distance and introduced the Shapley value to verify the utility of the prior features.

Finally, we evaluated the proposed method for BP waveform estimation on two independent multicenter datasets and compared it with state-of-the-art (SOTA) methods.

II. Methodology

A. Problem Formulation

Predicting BP waveforms based on PPG signals could be generalized as a real-time long sequence prediction problem. Therefore, the model needs to learn and obtain the mapping relationship between the two sequences through a sliding window with a fixed window length L_x. In this study, for the input PPG sequence (i.e., PPG waveform) $X = \{x_1^t, x_2^t, \ldots, x_{L_x}^t | x_i^t \in \mathbb{R}^{d_x}\}$ at time window *t*, the output is the corresponding BP sequence (i.e., BP waveform) $Y = \{y_1^t, y_2^t, \ldots, y_{L_y}^t | y_i^t \in \mathbb{R}^{d_y}\}$. Unlike the univariate prediction problem, the feature dimension d_x of this model is not limited to univariate cases. The derivative series of the PPG series can be chosen as the input variables, i.e., the velocity of PPG (VPG) and the acceleration of PPG (APG). The proposed model accepts X as input and Y as output, given by (1):

$$Y = F(X; M; \theta) \tag{1}$$

where $F(\cdot)$ is the model function, $M = \{m_1^t, m_2^t, \ldots, m_s^t | m_s^t \in \mathbb{R}\}$ is optional parameters representing the introduced handcrafted prior knowledge, i.e., the morphological characteristics of the pulse waveform, and θ is the hyperparameter of the network.

B. Feature Metrics and Discretization

In this study, 75 interpretable features, including time-domain, frequency-domain, and nonlinear features, were extracted from the PPG signals. These morphological features are commonly defined in the existing paper, such as the time parameter signal skewness (2), dimensionless metrics margin

factor (3), and area parameter K value [26] (4).

$$x_{skew} = \frac{\frac{1}{L_x}\sum_{i=1}^{L_x}(x_i - x_{mean})^3}{\left[\frac{1}{L_x}\sum_{i=1}^{L_x}(x_i - x_{mean})^2\right]^{3/2}} \tag{2}$$

$$MF = x_{max}/\frac{1}{L_x}\sum_{i=1}^{L_x}(x_i)^2 \tag{3}$$

$$K = \frac{p_{mean} - p_{min}}{p_{max} - p_{min}} \tag{4}$$

where p_{max}, p_{min} and p_{mean} represent the maximum, minimum, and mean values of the PPG amplitude during the cycle to quantify the variability of the area, respectively.

In addition, we additionally calculated the pulse rate variability (PRV) to describe the patterns of variation within the sequence window, such as the approximate entropy (ApEn) describing the degree of sequence disorder. For the PPG sequence $\boldsymbol{x}$, we use a window truncation of length $m = 2$ to obtain the embedded sequence $\boldsymbol{W} = \{w(i), i = 1, 2, \ldots, N + m - 1\}$, the similarity tolerance is $r = 0.1 \times SD$, and ApEn is defined as:

$$ApEn(\boldsymbol{x}, m, r) = lnC_{Ap}^{m}(r) - lnC_{Ap}^{m+1}(r) \tag{5}$$

where $C_{Ap}^{m}(r)$ denotes the average similarity rate for all m-length sub-fragments of the entire sequence under the similarity criterion r. The calculation formula is as follows:

$$C_{Ap}^{m}(r) = count(w(j)) \tag{6}$$

where $w(j)$ satisfies $max(|w(i) - w(j)|) \leq r)/(N - m + 1)$.

In addition, we perform a supervised ChiMerge split-box discretization [27] of the computed continuous features, mainly considering three benefits: 1) the robustness of discrete features to anomalous data, 2) the efficient operation of sparse vector inner product multiplication, 3) simplifying the model and reducing the risk of overfitting. ChiMerge relies on chi-square tests that combine adjacent intervals with lower chi-square values (indicating that they have similar class distributions) until the stopping criterion is satisfied. The calculation of χ^2 is shown in (7), A_{ij} indicates the number of patterns in the *i-th* interval, *j-th* category, correspondingly, E_{ij} indicates the expected frequency of A_{ij}.

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{k}\frac{(A_{ij} - E_{ij})^2}{E_{ij}} \tag{7}$$

The set of continuous features can be expressed as $T_c = \{(M, Z)|M \in \mathbb{R}^{t \cdot d_m}, Z \in \mathbb{R}^t\}$, where $M^t = \{f_{m_1}^t, f_{m_2}^t, \ldots, f_{m_s}^t\}$ is called s-dimension feature set at time t, and $Z^t = \{(1/L_y)\sum_{i=1}^{L_y} y_i^t\}$ is the mean BP value for the t-th set. Thus, it is a multi-instance question because only coarse-grained label information is available [28]. We proposed a modified ChiMerge algorithm (the pseudo-code shown in https://github.com/Ma-Chenbin/Pseudo-code-of-Algorithms-in-KD-Informer) suitable for the BP waveform estimation. The algorithm first splits the BP values equidistantly into bins, then the χ^2 value of the corresponding category is calculated to merge the minimum interval, thus achieving feature discretization.

C. Feature Selection and Fusion

The complete features may have redundant information, so we designed the BEFS method to obtain the valuable features. Specifically, the BEFS algorithm is divided into the baseline set generation and backward elimination phase. The Wrapper and Embedded methods filter the feature subsets with the top k for estimating the SBP and DBP, respectively, and use the merged set as the baseline set. Then, in the backward elimination phase, the features with the lowest importance are eliminated in turn, and the remaining feature set is input to the regressor to fit the BP values. The final number of feature sets is considered according to the optimized regression results. Pseudo-code of BEFS is shown in https://github.com/Ma-Chenbin/Pseudo-code-of-Algorithms-in-KD-Informer.

After obtaining the optimized feature set, we need to design the fusion strategy to integrate the deep features from the KD-Informer encoder with the traditional morphological features. The classical fusion approach determines the proportion of features from different sets by setting feature weight ε, which is calculated as follows:

$$F_f = \varepsilon \cdot F_m + (1 - \varepsilon) \cdot F_d \tag{8}$$

where F_f is the fusion features, F_m and F_d indicate morphological features and deep features, respectively. However, although this linear fusion method is simple to deploy, the hyperparameters are challenging to adjust automatically according to the data distribution. Extensive experiments are required to select the appropriate scaling factors. For this purpose, we have designed a new fusion method that adaptively adjusts feature scale factors. This approach updates the feature weights by training a multi-layer perceptron, and the fusion operation is defined as follows:

$$F_f = max\left(0, \sum_{i}^{s} w_i f_{m_i} + \sum_{j}^{n} dw_j f_{d_j} + b\right) \tag{9}$$

where $F_m = \{f_{m_1}, f_{m_2}, \ldots, f_{m_s}|f_{m_i} \in \mathbb{R}\}$, $F_d = \{f_{d_1}, f_{d_2}, \ldots, f_{d_n}|f_{d_i} \in \mathbb{R}\}$, b is the bias. The nonlinear spatial mapping of the fully connected layer enables a better trade-off between prior knowledge and information from deep features.

D. Sequence Learning Details

As the most commonly used sequence learning model, the Transformer has a strong ability to capture long-range dependencies [29]. However, the traditional Transformer still has some shortcomings, such as the computational complexity of multi-head self-attention (MHSA), large memory consumption bottleneck, and slow step-by-step decoding prediction. The Informer model proposed by Zhou et al. [30] addresses these problems well. As shown in Fig. 1(a), Informer extends Transformer with Kullback-Leibler divergence-based sparse attention and reduces the model size using the stacked multistage approach. The BP waveform estimation mainly focuses on the mapping relationship between the physiological sequences of different modalities, so the masked ProbSparse MHSA structure (Fig. 1(d))

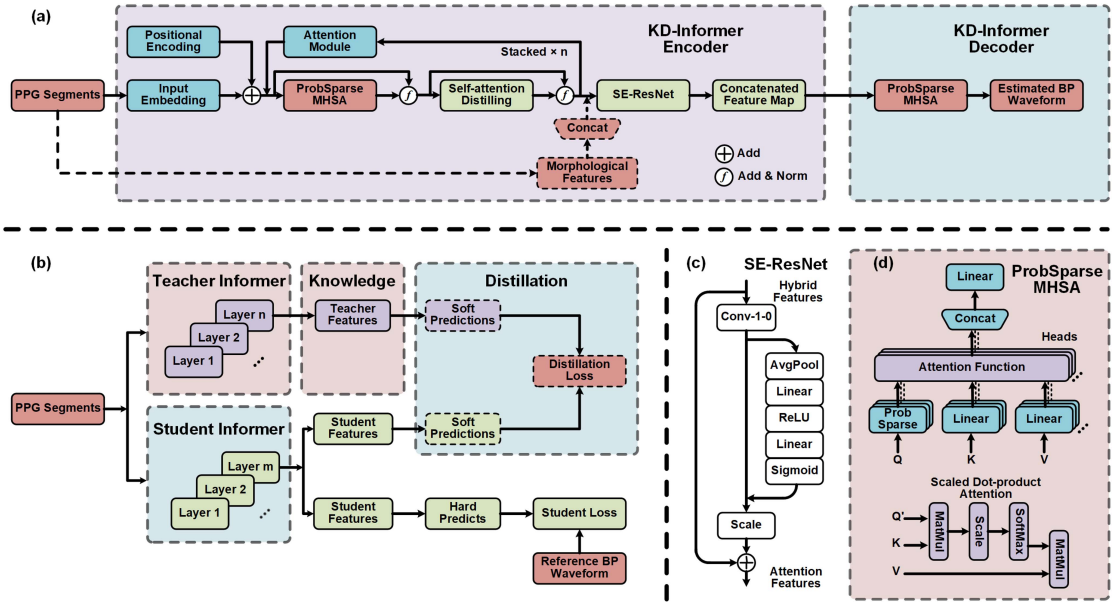

Fig. 1. The overall architecture of the proposed cuff-less continuous BP waveform estimation approach, (a) Architecture of KD-Informer. (b) teacher-student framework for response-based knowledge distillation. (c) the backbone of the SE-Resnet. (d) the ProbSparse MHSA module.

in the original paper is deleted to avoid sequence information leakage. The overall KD-Informer framework is shown in Fig. 1. Specifically, the proposed KD-Informer is composed of three main components: 1) Informer as the backbone network for learning the shared high-level features (Fig. 1(a)); 2) ResNet structure with SE module to fuse prior information of PPG signals (Fig. 1(c)); 3) response-based KD strategy to transfer the knowledge learned by the teacher Informer to the lightweight student Informer (Fig. 1(b)).

E. Knowledge Distillation

Considering the higher computational complexity and large storage requirements of deep models, we adopt a response-based KD strategy to transfer the knowledge learned by the teacher Informer to the lightweight student Informer. The main idea of this method is to train an efficient student model under the guidance of the teacher model to obtain comparable accuracy. The dark knowledge learned from the pre-trained teacher model can help the student model imitate the behavior of the teacher model. The specific distillation process is shown in Fig. 1(b).

As we know, the soft logits from the teacher model contain useful information among particular categories for a softmax regression task [24]. Temperature parameter T can control the importance of each label in the classification task, e.g., higher temperatures produce weaker probability distributions p_i on specific categories z_i (show in (10)). However, the output of the BP prediction task is a time series, but not a probability distribution or single value over classes in classification tasks. No such logits are softening in regression tasks; hence, we discard the temperature parameter. However, the predictions of teacher can help students reach the approximate solution space.

Therefore, we consider the predicted output of the teacher model, $\hat{y}_i^t$, as the soft labels to guide the training for the student. The KD-Informer loss can be defined as (11). The first term is the distillation loss, which measures the L_2 distance between the teacher prediction $\hat{y}_i^t$ and the student prediction $\hat{y}_i^s$. The second term is the student loss, which measures the distance between the student prediction $\hat{y}_i^s$ and the reference BP sequence y_i^{ref}. Hyperparameter μ is used to adjust the degree of contribution between the loss functions.

$$p_i = \frac{\exp(z_i/T)}{\sum_j \exp(z_j/T)} \tag{10}$$

$$L_{KD} = \mu \cdot \sqrt{\sum\nolimits_j \left(\hat{y}_i^t - \hat{y}_i^s\right)^2} + (1-\mu) \cdot \sqrt{\sum\nolimits_j \left(\hat{y}_i^s - y_i^{ref}\right)^2} \tag{11}$$

III. Experimental Setup

A. Data Description

We chose two datasets for validating the proposed method.

1) Mindray Dataset: This dataset was from the Shenzhen Mindray Bio-Medical Electronics Co., Ltd. The dataset was collected from 7 hospitals containing 467 patients. PPG was sampled at 125 Hz, and the invasive BP waveform was collected simultaneously as a reference signal for the model. Data collection also included the changes in drive current, which were used to reduce the effects of variability caused by changes in the environment, equipment, and sensor location. The length of the signals was from 30 min to 24 h.
2) MIMIC (Multi-parameter Intelligent Monitoring in Intensive Care) Dataset: This publicly available dataset contains more than 20000 waveform records sampled at 125 Hz from intensive care unit (ICU) patients [31]. Each recording typically contains between 24 and 48 h of continuous data recorded from patient monitors in the medical, surgical, and ICU of Boston's Beth Israel Hospital.

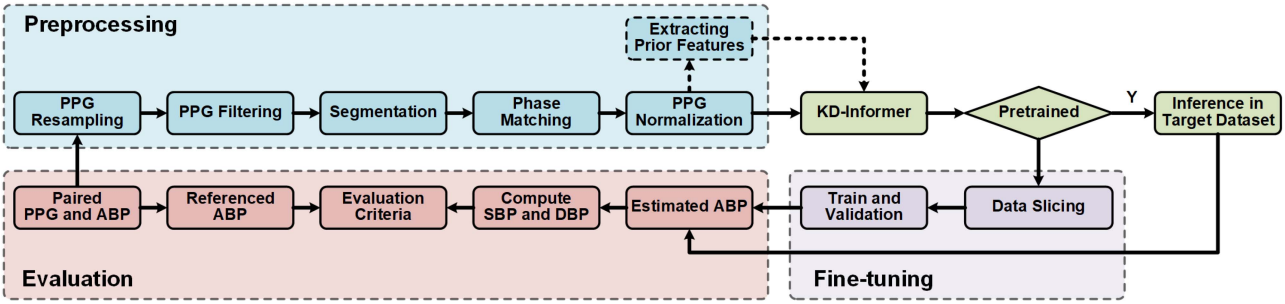

Fig. 2. Flow chart of overall experimental design. The pipeline for data pre-processing of the PPG signal includes resampling, filtering, segmentation, phase matching and normalization, and extracting prior features based on PPG waveform morphology is an optional step. If the proposed KD-Informer has been pre-trained on the source dataset (e.g., large and high-quality Mindray dataset), it can be directly transferred to perform inference on the target dataset (e.g., MIMIC dataset or collected signals from users) to obtain the estimated BP waveform. Training can be fine-tuned on the pre-processed dataset if the model is not pre-trained. Finally, the model utility can be evaluated by extracting the SBP and DBP parameters from the estimated BP waveform.

B. Data Prepossessing

Fig. 2 illustrates the flowchart of experimental design. The PPG and BP signals collected in the dataset have varying lengths and are disturbed by outliers and baseline drift, so signal pre-processing is required, as shown in the blue box in Fig. 2. For signals with different sampling rates, we first resample them to 125 Hz, which helps the sequence alignment. The wavelet transform is suitable for analyzing non-stationary signals and has better time-frequency localization in signal mutation, compression reconstruction, and signal denoising. This study considered different wavelet functions, including haar, daubechies, and symlet wavelets. Finally, we selected sym4 as the base wavelet, and the PPG signals are decomposed independently in two stages based on the soft threshold function. Then, the cleaned signals are segmented using a time lag window with a fixed size of 8 s and a sliding step of 3 s. Besides, the training data is augmented and aligned to avoid information loss.

There is a large variation in PPG amplitudes among different individuals or even the same person with varying sensor positions. In practice, it is found that the deep learning algorithms tend to converge better when all inputs have similar distributions. So we adopt the z-score normalization to eliminate the effects of different amplitudes among PPG recordings and ensure that the model could converge quickly. Finally, the morphological features are extracted from the normalized PPG signals to provide prior information for model decision-making. After data pre-processing, Mindray dataset consists of 210472 segments with a mean SBP of 123.93 $\pm$ 26.40 mmHg and DBP of 69.70 $\pm$ 17.91 mmHg. The MIMIC dataset consists of 108626 segments with a mean SBP of 131.26 $\pm$ 30.15 mmHg and DBP of 69.56 $\pm$ 18.84 mmHg.

In the model training and validation phases, we divide the dataset into the training set, validation set and test set with the ratio of 7:1.5:1.5 using inter-subject division, which ensures that each set does not contain data from the same patient, thus avoiding information leakage. To ensure the reliability of the test results, the results shown in this paper are all evaluation metrics obtained on the test set.

C. Evaluation Criteria

This paper designed an MLCSS algorithm to compare the estimated BP waveform and the reference waveform. Traditional distance calculation methods, such as the Euclidean distance, are susceptible to outliers because they are calculated with equal weights for all corresponding points. These disturbances can also lead to problems in measuring similarity if the predicted BP has a similar pattern to the referenced BP sequence for most of the period but with sudden outliers. Although the accuracy of the predicted values is essential, this study is also concerned with the ability to track the trend of the predicted waveforms. And LCSS can solve these problems well by defining the length of the longest common subsequence of two BP sequences by (12), and then obtaining the similarity formula (13) based on the common subsequence.

$$D_{LCSS} = 1 - (LCSS(Y', Y))/min(len_{Y'}, len_Y) \tag{12}$$

$$LCSS(Y', Y) = \begin{cases} 0, if\ Y' = \emptyset\ or\ Y = \emptyset \\ 1 + LCSS(y'_{t-1}, y_{t-1}), if(dist(y'_t, y_t) < \gamma) \\ \max(LCSS(y'_{t-1}, y_t), LCSS(y'_t, y_{t-1})), otherwise \end{cases} \tag{13}$$

where γ denotes the membership similarity threshold.

To further improve the accuracy of BP value estimation, we propose a time coefficient to reflect the proportion of predicted BP with similar values over a specific period. Its calculation formula is:

$$Tcor = \frac{\sum_{i=1}^{len_{Y'}} \sum_{j=1}^{len_Y} (\Delta t - |t_i(y') - t_j(y)|) \cdot \delta_{ij}(p_i^{Y'}, p_j^Y)}{\sum_{i=1}^{len_{Y'}} \sum_{j=1}^{len_Y} (\Delta t - |t_i(y') - t_j(y)|)} \tag{14}$$

where Δt denotes the time precision, and $t_i(y')$ denotes the time when the predicted BP value y' is at Δt. δ_{ij} is the Kronecker delta, which represents the overlap equation, and the function value is 1 when the predicted BP $p_i^{Y'}$ is the same as the measured BP value p_j^Y. To combine the time factor, the similarity algorithm

TABLE I
PERFORMANCE COMPARISON WITH REFERENCED FEATURE SELECTION ALGORITHMS

Algorithm	Redundancy rate	MLCSS distance	SBP			DBP		
			ME (mmHg)	SD (mmHg)	MAE (mmHg)	ME (mmHg)	SD (mmHg)	MAE (mmHg)
MCFS [36]	0.297	0.217	0.019	6.355	4.650	0.010	4.920	3.117
ReliefF [33]	0.316	0.328	0.017	6.418	4.817	0.019	4.542	3.078
mRMR [34]	0.265	0.312	0.036	6.296	4.562	0.007	4.972	3.290
WSF [35]	0.322	0.297	0.022	6.353	4.799	0.015	4.683	3.175
BEFS (Proposed)	**0.233**	**0.175**	**0.016**	**6.242**	**4.180**	**0.010**	**4.456**	**2.993**

of MLCSS is calculated as:

$$D_{MLCSS} = \frac{1 - (LCSS(Y', Y))}{min(len_{Y'}, len_Y)} \times Tcor \tag{15}$$

In addition, we also used commonly used metrics to assess the accuracy of predicted BP values. For instance, the mean error (ME), SD, and mean absolute error (MAE) between estimated BP value and reference BP value were adopted for measuring the regression performances. The most widely used standard for evaluating BP devices, the American National Standards of the Association for the Advancement of Medical Instrumentation (AAMI), was used to validate the performance of the proposed method. This standard requires that the ME and SD of the predicted and the reference values be less than 5 and 8 mmHg. In addition, the correlation and Bland-Altman analysis were performed to verify the agreement between the model and the reference.

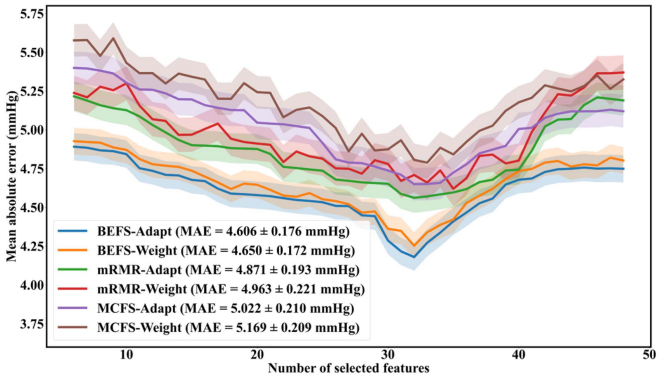

Fig. 3. Estimation error curve of feature selection method after feature fusion strategies. BEFS-Adapt donates BEFS algorithm combined with proposed fusion method for adaptively adjusting feature scale factors. BEFS-Weight means the traditional method of setting fixed weight coefficients.

IV. EXPERIMENTAL RESULTS

A. Feature Selection and Fusion

To verify the performance of the proposed BEFS algorithm, supervised feature weighting algorithm Relief [33], max-relevance and min-redundancy algorithm (mRMR) based on mutual information [34], weakly supervised feature selection algorithm (WSF) based on spectral analysis [35], and unsupervised MCFS [36], were chosen for comparison. In addition to quantifying the error in BP estimation by the AAMI standard, we also used the redundancy rate based on mutual information to calculate the redundancy of the formula measurement features, as shown in (16). Mutual information is expressed by (17), which measures the degree of bivariate dependence and can be viewed as the amount of information contained in one feature that contains another feature. The smaller the correlation between the selected features, the lower the redundancy of the subset.

$$R(F_m) = \frac{1}{s(s-1)} \sum_{I,J \in F_m, i<j} MI(I, J) \tag{16}$$

$$MI(I, J) = \sum_{f_i \in I} \sum_{f_j \in J} p(f_i, f_j) log\left(\frac{p(f_i, f_j)}{p(f_i)p(f_j)}\right) \tag{17}$$

Table I shows the test results obtained using multiple feature selection methods, the redundancy rate for the corresponding number of feature subsets, and the MLCSS distance. BEFS algorithm outperforms other methods with a maximum MAE decrease of 0.638 and 0.306 mmHg in SBP and DBP. And the SD of SBP was reduced by a maximum of 0.176 mmHg, the redundancy rate was reduced by 27.64%, and the MLCSS distance was reduced by 46.65%.

We further evaluate the MAE versus different feature selection methods and feature fusion strategy with SBP as an example. The results are shown in Fig. 3. Here, three feature selection methods are selected for demonstration, combined with our proposed fusion method for adaptively adjusting feature scale factors and the traditional method of setting fixed weight coefficients. As shown in Fig. 3, the proposed BEFS algorithm has the lowest MAE for all feature sets compared to the other feature selection methods. Compared with the fixed weight approach, the performance curve of our proposed fusion method is smoother and has better optimization performance. Most feature selection methods have the lowest estimation error on the subset of 32 features.

Further, we evaluated the effectiveness of the optimal subset screened by the BEFS algorithm using the Shapley value. The biggest advantage of Shapley value over feature importance is that the former reflects the influence of the feature in each sample and shows the positivity and negativity of the influence. However, the traditional Shapley value must compute all possible subsets of features to estimate the importance, and the computational complexity increases when there are more features. Therefore, the approximation of Monte Carlo sampling was used in this experiment [37]. The Shapley sampling values

TABLE II
CONTRIBUTION OF KEY DESIGN MODULES FOR EACH MODEL (ME ± SD, MMHG) ON MINDRAY DATASET

Model	Single PPG sequence		PPG and VPG		PPG with morphological features		All modules	
	DBP	SBP	DBP	SBP	DBP	SBP	DBP	SBP
Student model	0.101 ± 6.422	0.053 ± 8.441	-0.024 ± 5.767	-0.022 ± 8.343	-0.017 ± 4.964	**0.205 ± 7.830**	**0.092 ± 4.532**	0.024 ± 8.129
Teacher model	0.010 ± 4.517	-0.087 ± 5.347	-0.092 ± 3.814	0.083 ± 5.109	**-0.042 ± 2.271**	0.051 ± 4.445	0.091 ± 3.867	**-0.172 ± 4.379**
KD-Informer	-0.031 ± 6.315	0.046 ± 7.652	-0.018 ± 4.510	-0.017 ± 7.177	**0.011 ± 4.453**	**0.013 ± 6.237**	0.019 ± 4.457	0.024 ± 6.559

of the features f_m^i are calculated as follows:

$$Shapley(f_m^i) = \sum_{s=1}^{S} (\hat{y}(F_m^{+i}) - \hat{y}(F_m^{-i})) \tag{18}$$

where, $\hat{y}(F_m^{+i})$ denotes the predicted BP value for the included feature set f_m^i, and $\hat{y}(F_m^{-i})$ denotes the predicted BP value for the non-included feature set f_m^i.

These results demonstrate the validity of some of the physiological parameters proposed in previous studies, including the K value [26], which quantifies the degree of variability in the pulse waveform amplitude, and temporal parameters that meticulously describe the systolic and diastolic phases [15], [38]. Interestingly, some of our newly added features of PRV, such as ApEn, also play a significant positive role. In addition, non-dimensional metrics that describe the time-domain characteristics of the waveform, such as margin factor, also have a high degree of contribution, ranking fourth in importance.

B. Ablation Study

It is known that multiple modal input signals may affect the prediction of BP sequences. For example, several studies have shown that VPG can provide sensitive BP waveform changes [19], [39]. From the hypothesis of this study, incorporating more manually defined morphological features also seems to help the KD-Informer capture the sequence mapping relationship between PPG and BP. This subsection reports our ablation studies to select critical design modules on the Mindray dataset (Table II). It is worth mentioning that these results were obtained by feeding the proposed models with features in different modes. The best results from the student net, teacher net, and the KD-informer are bold in Table II. These results demonstrate the validity of VPG and morphological features. Both fused VPG and morphological features effectively reduced the estimation error of BP compared to single PPG signals. The combination of PPG waveform and morphological features is optimal for the KD-Informer model with the KD strategy. The estimation errors for SBP and DBP are 0.013 ± 6.237 mmHg and 0.011 ± 4.453 mmHg, respectively. However, these enhancements are negligible for the optimal combination. Therefore, considering the derivative calculation and training expenses in practical application scenarios, we prefer the KD-Informer to choose the PPG combined with morphological features. In addition, the results of each model can demonstrate that the distillation strategy can further reduce the estimation error of the student model through the supervision of the large-scale teacher model.

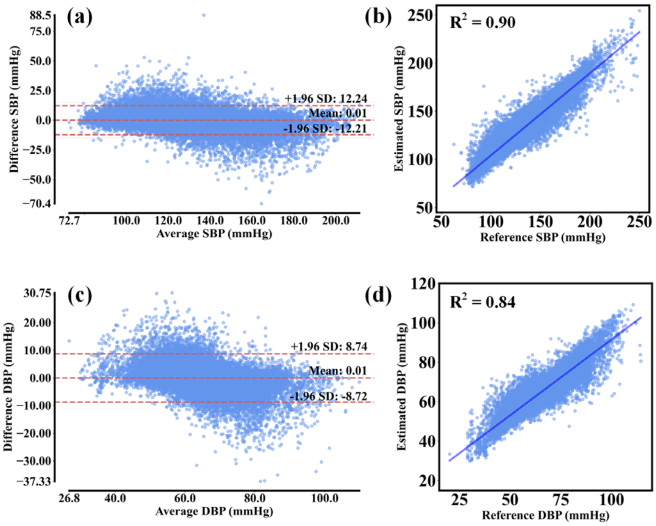

Fig. 4. Bland-Altman and the correlation plots of SBP and DBP with reference invasive BP. (a) Bland-Altman and the (b) Correlation plots in SBP estimation, (c) Bland-Altman and the (d) Correlation plots in DBP estimation. The Bland–Altman plot indicates that most of the estimated points for SBP and DBP are within the limits of agreement (bias ± 1.96 × SD), and the correlation plot indicates a high correlation between the estimated and the reference BP.

We evaluate the accuracy performance through the average recordings of the whole Mindray dataset. Fig. 4 presents the overall Bland-Altman and correlation plots for the KD-Informer estimations. Correlation plots show that the correlation coefficients of predicted SBP and DBP with reference BP are 0.90 and 0.84, respectively. This indicates that the estimated BP values are correlated with the reference BP values. The Bland-Altman plot shows that most of the estimated points for SBP and DBP were within 0.01 ± 12.24 and 0.01 ± 8.74 mmHg limits. However, in general, the method proposed in this study has a high agreement with the invasive BP measurement values, and both meet the AAMI standard.

Further, we randomly selected PPG and BP sequences aligned over time and inferred directly in the KD-Informer model obtained under the optimal training strategy and obtained the prediction results as shown in Fig. 5. As can be seen, the values are highly consistent, although there is a slight mismatch in individual rising or falling branches of the BP waveform. The output BP sequence of the model can accurately predict the trend of the reference BP sequence. However, the predicted BP may tend to follow the values of the previous cycle more closely, with a somewhat delayed response to abrupt waveform

TABLE III
PERFORMANCE COMPARISON WITH CLASSICAL ONE-DIMENSIONAL MODELS ON MINDRAY DATASET

Neural network regressor	Params (M)	FLOPs (Gmac)	MLCSS	DBP			SBP		
				ME (mmHg)	SD (mmHg)	MAE (mmHg)	ME (mmHg)	SD (mmHg)	MAE (mmHg)
RegNet [45]	11.89	0.56	0.633	0.419	7.349	5.194	-0.095	9.688	7.631
ResNet [46]	10.13	0.25	0.719	0.053	7.689	4.239	0.592	10.629	8.518
DenseNet [47]	0.63	0.15	0.660	-0.032	8.047	5.987	0.255	9.731	7.856
CNN-LSTM [20]	1.43	0.62	0.612	0.017	6.148	3.208	0.671	9.277	7.687
GRU based Seq2Seq [49]	**0.34**	**0.08**	0.540	0.048	7.933	4.191	0.040	8.692	7.928
SEMkResNet [48]	9.02	0.38	0.391	-0.032	5.549	4.588	-0.016	7.522	6.958
Transformer [29]	4.77	1.21	0.363	0.093	6.936	5.411	0.032	7.291	6.889
KD-Informer (Ours)	0.81	0.19	**0.175**	**0.011**	**4.453**	**2.997**	**0.013**	**6.237**	**4.178**

TABLE IV
PERFORMANCE COMPARISON WITH CLASSICAL ONE-DIMENSIONAL MODELS ON MIMIC DATASET

Neural network regressor	MLCSS	DBP			SBP		
		ME (mmHg)	SD (mmHg)	MAE (mmHg)	ME (mmHg)	SD (mmHg)	MAE (mmHg)
RegNet [45]	0.813	0.098	8.196	6.069	0.080	9.971	7.820
ResNet [46]	0.790	0.034	8.358	5.979	0.471	10.032	8.438
DenseNet [47]	0.728	0.031	8.076	5.851	0.372	9.905	7.367
CNN-LSTM [20]	0.671	0.022	6.906	4.721	-0.503	9.539	7.860
GRU based Seq2Seq [49]	0.608	-0.018	8.017	5.601	0.037	8.720	6.997
SEMkResNet [48]	0.417	-0.019	6.210	4.076	0.019	7.971	5.979
Transformer [29]	0.396	0.087	6.328	4.209	-0.043	7.863	6.319
KD-Informer (Transfer learning)	**0.158**	**0.010**	**3.867**	**2.970**	**0.018**	**5.931**	**4.008**
KD-Informer (Tuning on MIMIC)	**0.175**	**0.022**	**4.489**	**3.134**	**0.027**	**6.378**	**4.303**

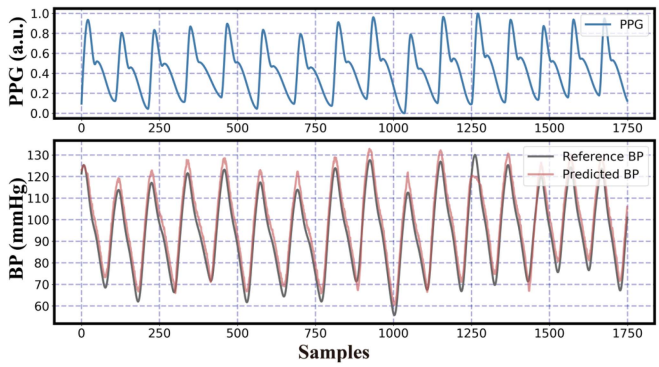

Fig. 5. The KD-Informer obtains the predicted BP waveform by direct inference from the input PPG signal.

changes. We can see that the PPG peak envelope and the BP waveform peak envelope have a consistent trend of change, which shows that the low-frequency signal has affected the BP estimation. These results confirm the previous conclusions about the importance of interaction between the low-frequency cardiovascular rhythms for the normal functioning of the cardiovascular system in humans [40]–[44]. The low-frequency component of signals reflects the frequency modulation of the heart rate by the autonomic nervous system.

C. Comparison With Other Models

To further determine the effectiveness of the proposed method, Tables III and IV list the classical one-dimensional networks for comparative analysis. These networks are highly influential in computer vision and are generally used as a significant reference in deep learning research. We transformed it into a one-dimensional convolution to make it more suitable for the time series estimation task. These models include: 1) classical representative networks, including RegNet with the optimal structure found from network design spaces [45], ResNet [46], and DenseNet [47]; 2) cascade networks, CNN-LSTM with fused time-space information [20]; 3) the latest further improved novel model, SEMkResNet, ResNet with a fused channel attention mechanism [48]. In addition, there is the GRU-based encoder-decoder architecture commonly used in machine translation, Seq2Seq [49] and Transformer [29]. It should be noted that the use of the same evaluation metrics and the same validation scheme (five-fold cross-validation) was ensured during the comparison. The experiments measured

TABLE V
SUMMARY OF STUDIES ON BP PREDICTION BASED ON SEQUENCE LEARNING USING MIMIC/PRIVATE DATASET

Author (Year)	Signals	Database (number of people included)	Dataset division	Composition of feature set	Neural network regressor	Prediction results	
						SBP (ME ± SD)	DBP (ME ± SD)
Shimazaki et al. (2018) [4]	PPG, VPG, ECG	Private Data (1363 cases)	Sample-wise	1s segments and demographic features	Auto Encoder	N/A ± 11.86	N/A
Baek et al. (2019) [5]	PPG, ECG	MIMIC (N/A)	Sample-wise	4s segments and spectral features	CNN	-1.23 ± 12.80	0.13 ± 7.54
Landry et al. (2020) [6]	PPG, ECG	MIMIC (15 cases)	Subject-wise	1s segments	ANN	-1.20 ± 17.80	-0.60 ± 8.40
Penwar et al. (2020) [20]	PPG	MIMIC (1557 cases)	Sample-wise	8s segments	CNN-RNN	-1.25 ± 5.65	1.55 ± 5.41
Eom et al. (2020) [10]	PPG, ECG, BCG	Private Data (15 cases)	Sample-wise	5s segments	CNN-RNN	-0.20 ± 5.83	-0.02 ± 4.91
This work (2021)	**PPG**	**Private and MIMIC Data (708 cases)**	**Subject-wise**	**8s segments and morphological features**	**KD-Informer (Transfer learning)**	**0.02 ± 5.93**	**0.01 ± 3.87**
		MIMIC (241 cases)			**KD-Informer (Tuning on MIMIC)**	**0.03 ± 6.38**	**0.02 ± 4.49**
						SBP (MAE ± SD)	DBP (MAE ± SD)
Tanveer et al. (2019) [7]	PPG, ECG	MIMIC (39 cases)	Subject-wise	2s segments and morphological features	ANN-RNN	1.10 ± 1.56	0.58 ± 0.85
Yan et al. (2019) [8]	PPG, ECG	MIMIC (604 cases)	Sample-wise	10s segments	CNN	3.09 ± 2.76	2.11 ± 2.00
Shimazaki et al. (2019) [9]	PPG, ECG	Private Data (78 cases)	Subject-wise	1s segments and demographic features	CNN	N/A ± 11.40	N/A
Slapničar et al. (2019) [50]	PPG, VPG	MIMIC (78 cases)	Subject-wise	5s segments	CNN	9.43 ± N/A	6.88 ± N/A
Aguiree et al. (2021) [22]	PPG, VPG	MIMIC (1131 cases)	Sample-wise	5s segments and demographic features	Seq2Seq	12.08 ± 15.67	5.56 ± 7.32
This work (2021)	**PPG**	**Private and MIMIC Data (708 cases)**	**Subject-wise**	**8s segments and morphological features**	**KD-Informer (Transfer learning)**	**0.02 ± 5.93**	**0.01 ± 3.87**
		MIMIC (241 cases)			**KD-Informer (Tuning on MIMIC)**	**0.03 ± 6.38**	**0.02 ± 4.49**

the complexity of the algorithm by the total number of model parameters and floating-point operations (FLOPs).

It can be observed that the proposed KD-Informer outperforms other counterparts in most of the evaluation metrics. Compared with the Transformer model based on the MHSA mechanism, the number of parameters of our proposed KD-Informer is reduced by 83.02% to 0.81 M. Meanwhile, FLOPs account for only 15.70% of the former. While being lightweight, our proposed model has the lowest estimation error. For BP waveform estimation, the KD-Informer has the lowest MLCSS, which is 51.79% lower than the Transformer, indicating a high agreement between the predicted BP waveform of the KD-Informer and the reference BP waveform. For DBP prediction on the Mindray dataset shown in Table III, most models pass the AAMI standard except DenseNet. However, only SEMkResNet, Transformer, and KD-Informer satisfy the AAMI standard in the prediction of SBP. These results demonstrate the effectiveness of our proposed approach. Compared to Transformer, KD-Informer only requires 15.70% of the number of FLOPs to achieve comparable accuracy, which has the potential for application to mobile wearable devices.

D. Robustness Performance

As we know, due to the nonlinear and nonconvex nature of neural networks, models may depend on the distribution characteristics of the training data and may risk poor generalization of new data. Models in the medical health domain require higher robustness and output more accurate BP even if there is some deviation in the data. To verify the robustness of the proposed KD-Informer, we further train and test it on the MIMIC dataset and make a detailed comparison with classical one-dimensional networks (shown in Table IV) and related research works (shown in Table V). Two experiments are conducted in this section. The first experiment obtains the optimal KD-Informer model by pre-training on the Mindray dataset and then fine-tuning on the MIMIC dataset and obtaining the prediction results of transfer learning on the latter dataset (shown in Fig. 2). The second experiment is directly trained and tested on the MIMIC dataset to keep the same dataset as similar studies. The results show that the transfer learning paradigm obtains the highest estimation accuracy, and the estimation errors for SBP and DBP are 0.018 $\pm$ 5.931 mmHg and 0.010 $\pm$ 3.867 mmHg, respectively. The experimental results demonstrate the effectiveness of transfer learning. Compared with the method trained directly in the MIMIC dataset, the pre-trained KD-Informer is superior in estimating BP waveform because it captures more physiological information from the subjects. However, the segments of the MIMIC dataset are only 51.61% of that of the Mindray dataset, and the BP prediction results of the KD-Informer are still ahead of the other classical models on the small dataset.

Notably, the method proposed in this study has the lowest BP estimation error compared to most studies using the subject-wise method [6], [9], [50]. In addition, Tanveer *et al.* [7] used

PPG and ECG data from 39 ICU patients in MIMIC, and their constructed hybrid cascade model predicted MAE errors of 1.10 $\pm$ 1.56 and 0.58 $\pm$ 0.85 mmHg for SBP and DBP, respectively. Although these results were far superior to similar studies, the small amount of data may influence the results, leading to more optimistic results. On the other hand, there are also many studies based on training samples for slicing data [4], [5], [8], [10], [20], [22], and the test set constructed in this way suffers from individual information leakage problems, so the generalization ability of this result still needs to be verified by collecting more datasets.

In addition, we also summarize the input signal modalities and feature sets in Table V. It is found that longer signal segments and incorporating prior knowledge, such as demographic features or morphological features, could predict BP values more accurately. Adding ECG signals to model inputs [5], [8], [10] also helps reduce the error in estimating BP. However, considering the cost of signal collection and computation, we suggest further investigating the utility of multimodal inputs for BP estimation.

V. Discussion

Typically, deep neural networks can be viewed as a black box, although they can automatically learn some distinguishable features. Therefore, if they tend to fit some unimportant features, it will cause the model to partially collapse to some features that are not conducive to expressing physiological states. To address this problem, adding human-designed prior information to the model will allow the model to learn some key features. This study explicitly uses prior knowledge to formulate information migration rules and designs feature fusion strategies to further enhance the model's ability to estimate the BP waveform.

The results of testing the number of features that describe the subset of PPG morphology on BP prediction in Fig. 3 show that the estimation error decreases first and then increases in a smile curve as the number of prior features increases. This result validates the conclusion of previous studies [35], [51] that an excessive number of features instead increases the degree of information redundancy. In addition, the feature fusion method that adaptively adjusts the weights of prior features and deep features can also effectively reduce the estimation error.

The quantitative analysis of Table I verifies the effectiveness of the proposed BEFS method, which is optimal among all the feature selection methods. We screened the optimal number of feature subsets with this method and explained the degree of feature contribution using the Shapley value. Our results confirm the conclusion of previous studies that morphological features of PPG are an essential indicator of BP prediction [7], [15], [38]. In addition to these features often reported in articles, our study concludes that PRV describing changes in PPG waveforms could also help improve the accuracy of BP prediction.

On the other hand, our study supports the utility of VPG. Takazawa *et al.* [52] reported that the APG could characterize vascular function. Several studies [51], [53] have also found that increasing the magnitude characteristics of APG can significantly improve the accuracy of BP prediction. Unfortunately, although some studies used the magnitude parameter of VPG [19], [39], there is not sufficient evidence for the utility of VPG. However, in this study, the ablation experiments shown in Table II reveal that the VPG sequence also improves the BP prediction accuracy, reducing the SD of DBP and SBP by 1.822 and 0.468 mmHg, compared with the input PPG sequence alone. Although the improvement after further fusion of morphological features is not significant, considering the complexity of the feature calculation, the relatively simple difference operation seems to have practical application.

In terms of model applicability his study is the first to apply Transformer as the backbone of a network for real-time estimating the BP waveform. Meanwhile, a response-based KD strategy is adopted to transfer the knowledge learned by the teacher Informer to the lightweight student Informer. The experimental setup is rigorous, using the invasive BP waveform as the gold standard for modeling and strictly ensuring data separation between subjects during the model testing process, thus ensuring the reliability of the results. The prediction results (Figs. 4 and 5) show that our proposed method meets the AAMI standard and has strong robustness in invalidating the public MIMIC dataset. Table III shows that compared with other deep networks, such as Transformer, the proposed KD-Informer has the highest prediction accuracy for BP with only 0.81 M parameters. Moreover, the cuff-less continuous BP measurement method requires only a single PPG signal, which avoids the cumbersome wearing of ECG electrodes and has the potential for clinical application scenarios.

In addition, unlike the regressor that directly acquires BP values, the KD-Informer directly outputs a continuous dynamic BP waveform. A recent randomized trial showed that early detection of hypotension by continuous hemodynamic monitoring might reduce intraoperative hypotension [25]. Therefore, the Seq2Seq model proposed in this study has a broader application prospect.

Furthermore, the experimental design of this study is rigorous. It is commonly believed that slicing training data may lead to extreme prediction bias [9], [22], [50]. As can be seen from the experimental designs summarized in Table IV, nearly half of the sequence studies used an approach that randomly split the signal segments and then divided the training set, leading to a high degree of similarity between the training and test sets. Therefore, to avoid encountering data from the same subjects in the test set, a more rigorous approach would be to use a cut-off with inter-subject separation, thus ensuring absolute independence of the training and test data.

And the distribution of the dataset used needs to be carefully considered during data processing. Clinical or daily monitoring scenarios can generate many signals of uncontrollable quality, which can affect the robustness of the model if it is trained only on datasets of excellent quality and with concentrated data distribution. In this study, several datasets are used for validation, and the experiments of transfer learning and dataset replacement also validate the effectiveness of the proposed method well. As can be seen from Table IV, the proposed method has the optimal BP prediction accuracy compared to the SOTA work with sequence learning.

We conclude by discussing some limitations of the present study. First, the distribution of BP may be different in different

populations; for example, MIMIC [31] collected mostly from clinically critically ill patients with a wide range of fluctuations in BP variability, which is different from that of healthy individuals. However, some studies have collected healthy individuals with minimal fluctuation [9], [10], [51], [54], [55]. In this sense, our proposed approach may be more applicable to the clinical monitoring setting. Therefore, continuously collecting datasets of the healthy population may further validate the proposed method.

Furthermore, the data collection must come from a controlled clinical scenario. This is because measurements are often more susceptible to physiological parameter fluctuations and produce motion noise. Theoretically, a more extensive test sample and long-term monitoring duration can encompass a broader range of population information, which hides a large number of physiological details and improves the fitting effect and generalization ability of the algorithm. There is still a need to establish more multicenter datasets to support this type of study.

VI. Conclusion

This paper presents a novel deep learning method (KD-Informer) for BP waveform estimation using the non-invasive PPG signal. At the same time, we integrate the prior information of pulse patterns with the SE-ResNet module into the knowledge transfer branch. Our proposed method uses a KD strategy to obtain information about potential associations between physiological sequences in a more lightweight architecture than existing sequence learning deep networks for PPG.

We demonstrate outstanding performance for continuous BP estimation on two independent multicenter datasets compared with the SOTA works. In addition, we also design comprehensive experiments to verify the utility of the pulse morphological feature selection and fusion method proposed in this paper. The experimental results prove that the proposed method is accurate and robust, thus providing a novel BP measurement insight.

References

[1] S. L. Stevens *et al.*, "Blood pressure variability and cardiovascular disease: Systematic review and meta-analysis," *BMJ*, vol. 354, 2016, Art. no. i4098.

[2] J. Lu *et al.*, "Prevalence, awareness, treatment, and control of hypertension in China: Data from 1· 7 million adults in a population-based screening study (China PEACE million persons project)," *Lancet*, vol. 390, no. 10112, pp. 2549–2558, 2017.

[3] X. Ding, Y. Zhang, and H. K. Tsang, "Impact of heart disease and calibration interval on accuracy of pulse transit time–based blood pressure estimation," *Physiol. Meas.*, vol. 37, no. 2, pp. 227–237, Jan. 2016.

[4] S. Shimazaki, S. Bhuiyan, H. Kawanaka, and K. Oguri, "Features extraction for cuffless blood pressure estimation by autoencoder from photoplethysmography," in *Proc. IEEE 40th Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2018, pp. 2857–2860.

[5] S. Baek, J. Jang, and S. Yoon, "End-to-end blood pressure prediction via fully convolutional networks," *IEEE Access*, vol. 7, pp. 185458–185468, 2019.

[6] C. Landry, S. D. Peterson, and A. Arami, "Nonlinear dynamic modeling of blood pressure waveform: Towards an accurate cuffless monitoring system," *IEEE Sensors J.*, vol. 20, no. 10, pp. 5368–5378, May 2020.

[7] M. S. Tanveer and M. K. Hasan, "Cuffless blood pressure estimation from electrocardiogram and photoplethysmogram using waveform based ANN-LSTM network," *Biomed. Signal Process. Control*, vol. 51, pp. 382–392, May 2019.

[8] C. Yan *et al.*, "Novel deep convolutional neural network for cuff-less blood pressure measurement using ECG and PPG signals," in *Proc. IEEE 41st Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2019, pp. 1917–1920.

[9] S. Shimazaki, H. Kawanaka, H. Ishikawa, K. Inoue, and K. Oguri, "Cuffless blood pressure estimation from only the waveform of photoplethysmography using CNN," in *Proc. IEEE 41st Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2019, pp. 5042–5045.

[10] H. Eom *et al.*, "End-to-end deep learning architecture for continuous blood pressure estimation using attention mechanism," *Sensors*, vol. 20, no. 8, 2020, Art. no. 2338.

[11] B. Saugel *et al.*, "Continuous noninvasive pulse wave analysis using finger cuff technologies for arterial blood pressure and cardiac output monitoring in perioperative and intensive care medicine: A systematic review and meta-analysis," *Brit. J. Anaesth.*, vol. 125, no. 1, pp. 25–37, Jul. 2020.

[12] Y. Ma *et al.*, "Relation between blood pressure and pulse wave velocity for human arteries," *Proc. Nat. Acad. Sci.*, vol. 115, no. 44, pp. 11144–11149, 2018.

[13] X. Ding, B. P. Yan, Y.-T. Zhang, J. Liu, N. Zhao, and H. K. Tsang, "Pulse transit time based continuous cuffless blood pressure estimation: A new extension and a comprehensive evaluation," *Sci. Rep.*, vol. 7, no. 1, Sep. 2017, Art. no. 11554.

[14] S. Suzuki and K. Oguri, "Cuffless blood pressure estimation by error-correcting output coding method based on an aggregation of adaboost with a photoplethysmograph sensor," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2009, pp. 6765–6768.

[15] Y. Kurylyak, F. Lamonaca, and D. Grimaldi, "A neural network-based method for continuous blood pressure estimation from a PPG signal," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf.*, 2013, pp. 280–283.

[16] L. Wang, W. Zhou, Y. Xing, and X. Zhou, "A novel neural network model for blood pressure estimation using photoplethesmography without electrocardiogram," *J. Healthcare Eng.*, vol. 2018, Mar. 2018, Art. no. 7804243.

[17] S. S. Mousavi, M. Firouzmand, M. Charmi, M. Hemmati, M. Moghadam, and Y. Ghorbani, "Blood pressure estimation from appropriate and inappropriate PPG signals using a whole-based method," *Biomed. Signal Process. Control*, vol. 47, pp. 196–206, Jan. 2019.

[18] M. Kachuee, M. M. Kiani, H. Mohammadzade, and M. Shabany, "Cuffless blood pressure estimation algorithms for continuous health-care monitoring," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 4, pp. 859–869, Apr. 2017.

[19] D. Lee *et al.*, "Beat-to-beat continuous blood pressure estimation using bidirectional long short-term memory network," *Sensors*, vol. 21, no. 1, 2021, Art. no. 96.

[20] M. Panwar, A. Gautam, D. Biswas, and A. Acharyya, "PP-Net: A deep learning framework for PPG-based blood pressure and heart rate estimation," *IEEE Sensors J.*, vol. 20, no. 17, pp. 10000–10011, Sep. 2020.

[21] 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*, vol. 21, no. 5, 2021, Art. no. 1867.

[22] N. Aguirre, E. Grall-Maës, L. J. Cymberknop, and R. L. Armentano, "Blood pressure morphology assessment from photoplethysmogram and demographic information using deep learning with attention mechanism," *Sensors*, vol. 21, no. 6, 2021, Art. no. 2167.

[23] G. Slapničar, M. Luštrek, and M. Marinko, "Continuous blood pressure estimation from PPG signal," *Informatica*, vol. 42, no. 1, pp.33–42, 2018.

[24] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *Adv. Neural Inf. Process. Syst., Workshop*, vol. 27, 2014. *arXiv:1503.02531*.

[25] K. Maheshwari *et al.*, "A randomized trial of continuous noninvasive blood pressure monitoring during noncardiac surgery," *Anesth. Analg.*, vol. 127, no. 2, pp. 424–431, 2018.

[26] K. H. Wesseling, "A simple device for the continuous measurement of cardiac output. Its model basis and experimental varification," *Adv. Cardiovasc.*, vol. 5, pp. 16–52, 1983.

[27] R. Kerber, "ChiMerge: Discretization of numeric attributes," in *Proc. 10th Nat. Conf. Artif. Intell.*, 1992, pp. 123–128.

[28] Z.-H. Zhou, "A brief introduction to weakly supervised learning," *Nat. Sci. Rev.*, vol. 5, no. 1, pp. 44–53, 2018.

[29] A. Vaswani *et al.*, "Attention is all you need," *Adv. Neural Inf. Process. Syst.*, vol. 30, pp. 6000–6010, 2017.

[30] H. Zhou *et al.*, "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 11106–11115.

[31] M. Saeed *et al.*, "Multiparameter intelligent monitoring in intensive care II: A public-access intensive care unit database," *Crit. Care Med.*, vol. 39, no. 5, pp. 952–960, 2011.

[32] *American National Standards for Electronic or Automated Sphygmomanometers*, ANSI/AAMI SP 10-1987, Association for the Advancement of Medical Instrumentation, Needham, MA, USA, 1987.
[33] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," *Mach. Learn.*, vol. 53, no. 1, pp. 23–69, 2003.
[34] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.
[35] F. Miao, Z.-D. Liu, J.-K. Liu, B. Wen, Q.-Y. He, and Y. Li, "Multi-sensor fusion approach for cuff-less blood pressuremeasurement," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 1, pp. 79–91, Jan. 2020.
[36] D. Cai, C. Zhang, and X. He, "Unsupervised feature selection for multi-cluster data," in *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2010, pp. 333–342.
[37] E. Štrumbelj and I. Kononenko, "Explaining prediction models and individual predictions with feature contributions," *Knowl. Inf. Syst.*, vol. 41, no. 3, pp. 647–665, Dec. 2014.
[38] Z. Xu, J. Liu, X. Chen, Y. Wang, and Z. Zhao, "Continuous blood pressure estimation based on multiple parameters from eletrocardiogram and photoplethysmogram by back-propagation neural network," *Comput. Ind.*, vol. 89, pp. 50–59, Aug. 2017.
[39] W.-H. Lin, H. Wang, O. W. Samuel, G. Liu, Z. Huang, and G. Li, "New photoplethysmogram indicators for improving cuffless and continuous blood pressure estimation accuracy," *Physiol. Meas.*, vol. 39, no. 2, Feb. 2018, Art. no. 025005.
[40] A. S. Karavaev *et al.*, "Low-frequency component of photoplethysmogram reflects the autonomic control of blood pressure," *Biophys. J.*, vol. 120, no. 13, pp. 2657–2664, Jul. 2021.
[41] A. R. Kiselev *et al.*, "Method of estimation of synchronization strength between low-frequency oscillations in heart rate variability and photoplethysmographic waveform variability," *Russian Open Med. J.*, vol. 5, Q5 no. 1, 2016, Art. no. e0101.
[42] A. Karavaev *et al.*, "Phase and frequency locking of 0.1-Hz oscillations in heart rate and baroreflex control of blood pressure by breathing of linearly varying frequency as determined in healthy subjects," *Hum. Physiol.*, vol. 39, no. 4, pp. 416–425, 2013.
[43] A. Kiselev *et al.*, "Interaction of 0.1-Hz oscillations in heart rate variability and distal blood flow variability," *Hum. Physiol.*, vol. 38, no. 3, pp. 303–309, 2012.
[44] D. S. Goldstein, O. Bentho, M. Y. Park, and Y. Sharabi, "Low-frequency power of heart rate variability is not a measure of cardiac sympathetic tone but may be a measure of modulation of cardiac autonomic outflows by baroreflexes," *Exp. Physiol.*, vol. 96, no. 12, pp. 1255–1261, 2011.
[45] I. Radosavovic, R. P. Kosaraju, R. Girshick, K. He, and P. Dollár, "Designing network design spaces," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 10428–10436.
[46] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
[47] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," *in Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4700–4708.
[48] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.
[49] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," *Adv. Neural Inf. Process. Syst.*, vol. 27, pp. 496–527, 2014.
[50] G. Slapničar, N. Mlakar, and M. Luštrek, "Blood pressure estimation fromphotoplethysmogram using a spectro-temporal deep neural network," *Sensors*, vol. 19, no. 15, 2019, Art. no. 3420.
[51] M. Liu, L.-M. Po, and H. Fu, "Cuffless blood pressure estimation based on photoplethysmography signal and its second derivative," *Int. J. Comput. Theory Eng.*, vol. 9, no. 3, pp. 202–206, 2017.
[52] F. Miao *et al.*, "A novel continuous blood pressure estimation approach based on data mining techniques," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 6, pp. 1730–1740, Nov. 2017.
[53] K. Takazawa *et al.*, "Assessment of vasoactive agents and vascular aging by the second derivative of photoplethysmogram waveform," *Hypertension*, vol. 32, no. 2, pp. 365–370, 1998.
[54] K. N. G. Priyanka *et al.*, "Estimating blood pressure via artificial neural networks based on measured photoplethysmography waveforms," in *Proc. IEEE SENSORS*, 2018, pp. 1–4.
[55] K. Song, K. Chung, and J. Chang, "Cuffless deep learning-based blood pressure estimation for smart wristwatches," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 7, pp. 4292–4302, Jul. 2020.