Stabilizing PPG-based Blood Pressure Estimation through Global-to-Personalized Learning

Gahyeon Kim [1] **Daeun Gwon** [2] **Minkyu Ahn** [1]

Abstract

This study investigates the limitations of subject-dependent models for PPG-based blood pressure estimation and proposes a population-level pretraining and fine-tuning strategy. With limited subject-specific data, conventional models often exhibit high inter-subject variability and unstable performance. To address this, a population-level model was first trained on stable subjects and then fine-tuned for subjects with unstable predictions. The proposed approach reduced the mean absolute error of systolic blood pressure (SBP) from 15.79 to 4.85 mmHg and diastolic blood pressure (DBP) from 4.98 to 3.14 mmHg. These results demonstrate substantially improved estimation stability compared with purely subject-dependent training.

1. Introduction

Blood pressure (BP) is a key physiological marker associated with severe complications such as cardiovascular disease and renal failure, and hypertension remains a leading cause of mortality worldwide (Elsayed et al., 2025). Continuous BP monitoring is therefore essential for early detection and chronic disease management.

However, BP fluctuates continuously due to physical activity, sleep, and posture (Morris et al., 2013). Conventional cuff-based sphygmomanometers measure BP intermittently via arm compression, causing discomfort (Pilz et al., 2024) and limiting continuous monitoring.

To address these limitations, wearable approaches using photoplethysmography (PPG) have gained attention. PPG measures peripheral blood flow using optical sensors and offers advantages in simplicity, portability, and cost-effectiveness (Chen et al., 2024).

Despite its potential, PPG-based BP estimation is challenged by substantial inter-subject variability. PPG waveforms differ across individuals due to physiological factors (e.g., vascular properties, age) and personal characteristics (e.g., skin properties and measurement sites) (Elgendi et al., 2024), leading to inconsistent mappings between waveform patterns and BP. As a result, population-level models trained on pooled data often suffer performance degradation (Zhang et al., 2020), while subject-dependent models are limited by insufficient or biased individual data (González et al., 2023).

In this work, we examine this problem from the perspective of estimation stability under limited personal data. Rather than proposing a new model architecture, we investigate how population-level pretraining and subject-specific fine-tuning interact in a global-to-personalized learning framework. In particular, we focus on subjects whose purely subject-dependent models exhibit unstable predictions and examine whether population-level initialization can serve as a rescue strategy in difficult personalization scenarios.

The objectives of this study are as follows:

- To investigate population-level learning and subject-dependent learning as complementary stages of model adaptation rather than independent alternatives.
- To compare models initialized from population-level training with purely subject-dependent models using the same subject-specific held-out test data.
- To analyze how global-to-personalized fine-tuning influences estimation accuracy and blood pressure variability tracking in subjects exhibiting unstable subject-dependent predictions.

2. Related Work

2.1. Physics-based (PTT-based) BP Estimation

Early studies on noninvasive BP estimation leveraged the relationship between pulse wave velocity (PWV) and blood pressure. PWV increases with BP due to vascular stiffening

[1]Department of Computer Science and Electrical Engineering, Handong Global University, Pohang, Republic of Korea [2]Institute for Information and Communication Technology, Handong Global University, Pohang, Republic of Korea. Correspondence to: Minkyu Ahn <minkyuahn@handong.edu>.

Proceedings of the Workshop on Structured Data for Health at the 43rd International Conference on Machine Learning, Seoul, South Korea. Copyright 2026 by the author(s).

(Ma et al., 2018), but its direct measurement is impractical. As a result, pulse transit time (PTT), an indirect surrogate inversely related to PWV, has been widely used (Choi et al., 2004). PTT is typically derived from ECG-PPG pairs or dual PPG signals and regressed against BP (Wang et al., 2014; Ding et al., 2017; Ghosh et al., 2016).

Despite its physiological interpretability, this approach shows limited inter-subject generalization (Butlin et al., 2018), requires per-subject calibration (Mukkamala et al., 2015), and struggles to capture long-term trends or rapid hemodynamic changes.

2.2. Machine Learning with Hand-crafted PPG Features

To overcome limitations of PTT-based methods, subsequent studies extracted temporal and morphological features from PPG signals and applied machine learning models (Zhang & Feng, 2017; Hasanzadeh et al., 2019). These features include waveform slope, amplitude, area, and derivatives.

However, such approaches remain sensitive to inter-subject variability. Age-related vascular changes and sensor conditions significantly alter waveform morphology, affecting feature consistency (Yousef et al., 2012; Ho et al., 2025). Consequently, hand-crafted features are often unstable and fail to generalize across subjects and measurement conditions.

2.3. Deep Learning-based End-to-End BP Estimation

Recent work has adopted deep learning models to estimate BP directly from raw PPG signals (Hong et al., 2021; Khajehpiri et al., 2024). These approaches eliminate the need for manual feature design and learn representations directly from data.

Nevertheless, their performance is highly dependent on data availability. With limited subject-specific data, models often fail to capture individual PPG–BP relationships, resulting in unstable and variable performance.

3. Method

3.1. Problem Setup

We formulate cuffless blood pressure estimation from 10-second PPG segments as a regression task. Given an input PPG segment, the model predicts the corresponding systolic blood pressure (SBP) and diastolic blood pressure (DBP) values.

3.2. Network Architecture

To focus on the effect of training strategy rather than architectural novelty, we use the same CNN–BiLSTM–Attention–FC backbone (Table 3) across all experiments. The CNN layers extract local waveform features, the BiLSTM captures temporal dependencies, and the attention module emphasizes informative temporal representations. The final feature representation is mapped to SBP and DBP through fully connected layers.

3.3. Subject-dependent Model (SD)

As a subject-specific baseline, SD models are trained independently for each subject using only that subject's data. This setting represents personalization without any population-level initialization.

Based on the subject-dependent evaluation results, we quantified prediction instability using two criteria: the mean absolute error (MAE) and the ratio between the variance of the predicted BP signal and that of the reference BP signal. Together, these indicators capture cases with high estimation error and poor tracking of BP variability, and they are later used in the experimental protocol to identify subjects requiring personalization.

3.4. Subject-independent Model (Global BP)

The Global BP model is trained on pooled data from estimation-stable subjects to learn population-level PPG–BP relationships. To reduce dominance by subjects with large numbers of segments, 10% of the available data is randomly sampled from each training subject. This model serves as a population-level reference and as an initialization for subsequent subject-specific personalization.

3.5. Fine-tuned Model (FT)

For each estimation-unstable subject identified through the experimental protocol, the Global BP model is adapted to the target subject to achieve personalized estimation. During fine-tuning, the CNN feature extractor is frozen, while the BiLSTM, attention, and fully connected layers are updated. This design preserves population-level waveform representations while adapting higher-level temporal and regression layers to subject-specific BP characteristics.

4. Experiments

4.1. Datasets and Preprocessing

4.1.1. Dataset

This study uses the MIMIC Database (Moody & Mark, 1992), which contains multi-channel physiological signals collected from patients admitted to the intensive care unit (ICU) at Beth Israel Deaconess Medical Center. It includes PPG, arterial blood pressure (ABP), electrocardiogram (ECG), and other vital signals. Only records containing simultaneously acquired PPG and ABP signals were

selected for analysis.

4.1.2. Preprocessing

PPG and ABP signals were segmented into non-overlapping 10-second windows. Signal quality was assessed using the minimum standard deviation (θ_{std}) computed from short sliding windows within each segment. Segments with $\theta_{\mathrm{std}} < 1 \times 10^{-2}$ were discarded due to insufficient pulsatility.

Valid segments were filtered using a 4th-order Butterworth bandpass filter (0.5–8 Hz) to remove noise (Ezzat et al., 2023). SBP and DBP labels were extracted from ABP waveforms as local maxima and the minima between adjacent peaks, respectively. Z-score–based outlier removal (z = 2.0) and physiological constraints (SBP: 60–200 mmHg, DBP: 30–120 mmHg) were applied. After preprocessing, 57 records (48 subjects) with over 1,000 valid segments were retained.

4.2. Experimental Protocol

To clearly describe the subject selection and training pipeline, the experimental protocol was organized as follows:

1. After preprocessing, 48 subjects were retained for analysis.
2. Subject-dependent (SD) models were trained and evaluated for all retained subjects using subject-specific train/test splits.
3. Subjects were ranked according to two instability indicators derived from the SD results: MAE and the variance ratio between predicted and reference BP signals.
4. The seven subjects with the highest combined instability ranking were designated as estimation-unstable subjects and reserved for personalization experiments.
5. The remaining 41 subjects were treated as estimation-stable candidates; 36 were used for Global BP training and 5 for validation.

All valid segments were then split into training and test sets with an 8:2 ratio, consistently applied across all experiments. For each subject, the SD model was developed using only that subject's training split, with 5-fold cross-validation performed within the training portion and final evaluation conducted on the held-out test split. For each unstable subject, the same subject-specific train/test split was reused for fine-tuning to enable a direct comparison between SD and FT on the same held-out test data.

Normalization was performed differently depending on the setting. In subject-dependent experiments, Z-score normalization was computed from each subject's training data and applied to its test data. For Global BP and fine-tuning, normalization parameters were computed only from the 36 training subjects and applied to the validation set and the unstable-subject test set.

All models were optimized using the Huber loss. Training hyperparameters are summarized in Table 2.

4.3. Evaluation Metrics

In this study, we evaluated blood pressure estimation performance using mean absolute error (MAE), root mean square error (RMSE), and Pearson correlation coefficient r as the primary metrics. For each subject, these metrics were computed by aggregating predictions across that subject's test segments. The reported mean $\pm$ standard deviation summarizes these subject-level metrics across the seven selected unstable subjects. For qualitative analysis, we visualized the relationship between estimates and references using scatter plots and presented Bland–Altman plots to examine error distributions and systematic bias.

4.4. Main Results

As summarized in Table 1, within the selected unstable-subject subset, the proposed global-to-personalized training strategy achieved 4.85 ± 1.55 mmHg for SBP and 3.14 ± 0.99 mmHg for DBP on the seven estimation-unstable test subjects. These values represent the mean and standard deviation of subject-level performance across the selected unstable subjects.

Table 1 further demonstrates the benefits of subject-specific fine-tuning. After fine-tuning, both MAE and RMSE decreased across all selected unstable subjects, with SBP MAE reductions exceeding 10 mmHg for several subjects. Pearson correlation coefficients (r) also increased consistently, indicating improved tracking of subject-specific BP variability. Notably, subjects that exhibited low or negative correlation under the subject-dependent model showed substantially stronger linear relationships between predicted and reference blood pressure following fine-tuning.

Paired comparisons on the same seven subjects revealed statistically significant improvements of the fine-tuned model over the subject-dependent model across all evaluation metrics, including SBP/DBP MAE, RMSE, and r, as confirmed by a Wilcoxon signed-rank test ($p = 0.0156$). Taken together, these findings suggest that population-level initialization followed by targeted fine-tuning can effectively recover cases in which purely subject-dependent training produces unstable or flattened predictions. However, the results are specific to the selected estimation-unstable subjects and

Table 1. Subject-level comparison of SBP and DBP estimation performance between subject-dependent (SD) and fine-tuned (FT) models. Δ denotes performance improvement after fine-tuning.

Subject	Performance Metrics																	
	MAE (mmHg)						RMSE (mmHg)						Pearson Correlation (r)					
	SBP			DBP			SBP			DBP			SBP			DBP		
	SD	FT	Δ	SD	FT	Δ	SD	FT	Δ	SD	FT	Δ	SD	FT	Δ	SD	FT	Δ
211	18.68	8.08	**10.59**	5.70	3.09	**2.62**	22.45	11.12	**11.33**	6.96	4.54	**2.43**	0.06	0.88	**0.82**	0.03	0.77	**0.74**
220	5.84	4.62	**1.22**	5.18	3.84	**1.34**	7.70	6.02	**1.68**	6.55	4.83	**1.73**	0.28	0.68	**0.40**	0.42	0.72	**0.31**
230	12.81	3.35	**9.46**	6.20	2.11	**4.09**	15.19	4.29	**10.90**	7.52	2.73	**4.79**	0.57	0.96	**0.39**	0.60	0.94	**0.34**
240	27.42	3.72	**23.70**	6.55	4.68	**1.87**	29.82	5.11	**24.71**	7.95	5.87	**2.08**	0.07	0.90	**0.84**	0.27	0.69	**0.42**
248	7.67	5.12	**2.55**	2.76	2.50	**0.26**	9.53	6.70	**2.83**	3.79	3.33	**0.46**	0.26	0.70	**0.44**	0.11	0.45	**0.34**
259	14.41	4.70	**9.71**	4.23	2.05	**2.19**	16.37	6.83	**9.54**	5.24	2.88	**2.36**	-0.15	0.89	**1.04**	0.39	0.83	**0.44**
417	23.67	4.38	**19.29**	4.24	3.74	**0.50**	24.55	5.54	**19.01**	5.53	4.92	**0.61**	-0.20	0.46	**0.65**	-0.04	0.29	**0.32**

should not be interpreted as cohort-wide evidence across all subjects in the dataset.

Representative prior studies provide useful context for positioning this work, although direct numerical comparison remains difficult because datasets, subject cohorts, and evaluation protocols differ substantially. Some prior studies employed subject-specific fine-tuning of cross-subject models (Hong et al., 2021; Khajehpiri et al., 2024), whereas others evaluated population-level models without personalization (Slapničar et al., 2019). A reference-level summary is provided in Table 4 in the appendix for contextual comparison rather than as a strict head-to-head benchmark.

5. Conclusion

In this study, we investigated whether subject-specific fine-tuning from a population-level model can improve PPG-based blood pressure estimation when purely subject-dependent models produce unstable predictions under limited personal data. For the selected subject-unstable group, the proposed approach provided more stable estimation performance than both the population-level model and the purely subject-dependent model. In particular, it alleviated flattened predictions that failed to respond sufficiently to blood pressure changes, thereby enabling more faithful tracking of true blood pressure variability.

These results suggest that global-to-personalized fine-tuning can serve as a practical rescue strategy for subject-unstable individuals whose subject-dependent models fail to adequately learn individual BP dynamics. At the same time, this study has several limitations. First, the findings are restricted to the selected unstable-subject subset, and validation on the full cohort and on additional datasets is needed before broader conclusions can be drawn. Second, even after fine-tuning, estimation errors in terms of MAE and RMSE remained relatively high for some subjects. Third, subject selection and fine-tuning strategies relied on simple rule-based criteria, leaving room for more adaptive and data-driven personalization approaches.

Future work will evaluate generalization through multi-dataset validation and conduct finer-grained analyses of subjects for whom prediction instability persists. We also plan to investigate adaptive personalization strategies, such as dynamically adjusting the fine-tuning scope or layer-wise learning rates according to individual data characteristics, with the goal of achieving more stable and reliable cuffless blood pressure monitoring across diverse user settings.

Acknowledgements

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the National Program for Excellence in SW (2023-0-00055) supervised by the IITP(Institute of Information & Communications Technology Planning & Evaluation) in 2026 and also by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(RS-2025-25412061).

Impact Statement

This work aims to advance machine learning methods for non-invasive blood pressure estimation using physiological signals. Improved stability and personalization of blood pressure estimation models may contribute to earlier detection and monitoring of cardiovascular risk, potentially supporting clinical decision-making and long-term health management.

At the same time, inaccurate or misinterpreted estimations

could lead to inappropriate health-related decisions if such systems are used without proper clinical oversight. Therefore, this work is intended as a supportive tool rather than a replacement for standard medical measurement or diagnosis.

We acknowledge the importance of responsible deployment, data privacy, and careful validation in diverse populations before real-world use. We do not foresee significant negative societal impacts beyond those commonly associated with medical machine learning systems when used appropriately.

References

Butlin, M., Shirbani, F., Barin, E., Tan, I., Spronck, B., and Avolio, A. P. Cuffless estimation of blood pressure: importance of variability in blood pressure dependence of arterial stiffness across individuals and measurement sites. *IEEE Transactions on Biomedical Engineering*, 65 (11):2377–2383, 2018.

Chen, G., Zou, L., and Ji, Z. A review: Blood pressure monitoring based on PPG and circadian rhythm. *APL Bioengineering*, 8(3), 2024.

Choi, B. C., Jung, D. K., Jeong, D. U., Ro, J. H., and Jeon, G. R. Measurement of cardiac pulse transit time using photoplethysmography sensor. *Journal of Sensor Science and Technology*, 13(5):383–391, 2004.

Ding, X., Yan, B. P., Zhang, Y. T., Liu, J., Zhao, N., and Tsang, H. K. Pulse transit time based continuous cuffless blood pressure estimation: A new extension and a comprehensive evaluation. *Scientific Reports*, 7(1), 2017.

Elgendi, M., Haugg, F., Fletcher, R. R., Allen, J., Shin, H., Alian, A., and Menon, C. Recommendations for evaluating photoplethysmography-based algorithms for blood pressure assessment. *Communications Medicine*, 4 (1), 2024.

Elsayed, A. E. M., Almani, N. Y. F., Asaad, O. M., Mohamed, L. H. S., Alghamdi, R. S. A., Moussa, N. A., Al-Majmaie, Z. N., Murayyi, R. M. A., Alsaluli, B. A., Alnuwaiqi, Z. A., Qutqut, H. M. H., Almehmadi, A. S., and Alhibshi, A. H. M. Cardiovascular complications in patients with hypertension: A retrospective study. *American Journal of Psychiatric Rehabilitation*, 28(5):523–528, 2025.

Ezzat, A., Omer, O. A., Mohamed, U. S., and Mubarak, A. S. Blood pressure estimation from photoplethysmogram using hybrid bidirectional long short-term memory and convolutional neural network architecture. *Traitement du Signal*, 40(6), 2023.

Ghosh, S., Banerjee, A., Ray, N., Wood, P. W., Boulanger, P., and Padwal, R. Continuous blood pressure prediction from pulse transit time using ECG and PPG signals. In *Proceedings of the IEEE Healthcare Innovation Point-Of-Care Technologies Conference (HI-POCT)*, pp. 188–191, 2016.

González, S., Hsieh, W. T., and Chen, T. P. C. A benchmark for machine-learning based non-invasive blood pressure estimation using photoplethysmogram. *Scientific Data*, 10(1), 2023.

Hasanzadeh, N., Ahmadi, M. M., and Mohammadzadeh, H. Blood pressure estimation using photoplethysmogram signal and its morphological features. *IEEE Sensors Journal*, 20(8):4300–4310, 2019.

Ho, M. Y., Pham, H. M., Saeed, A., and Ma, D. WF-PPG: A wrist-finger dual-channel dataset for studying the impact of contact pressure on PPG morphology. *Scientific Data*, 12(1), 2025.

Hong, J., Gao, J., Liu, Q., Zhang, Y., and Zheng, Y. Deep learning model with individualized fine-tuning for dynamic and beat-to-beat blood pressure estimation. In *Proceedings of the IEEE 17th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, pp. 1–4. IEEE, 2021.

Khajehpiri, B., Granger, E., De Zambotti, M., Baker, F. C., Yuksel, D., and Forouzanfar, M. SleepBP-Net: A time-distributed convolutional network for nocturnal blood pressure estimation from photoplethysmogram. *IEEE Sensors Journal*, 24(12):19590–19600, 2024.

Ma, Y., Choi, J., Hourlier-Fargette, A., Xue, Y., Chung, H. U., Lee, J. Y., Wang, X., Xie, Z., Kang, D., Wang, H., Han, S., Kang, S.-K., Kang, Y., Yu, X., Slepian, M. J., Raj, M. S., Model, J. B., Feng, X., Ghaffari, R., Rogers, J. A., and Huang, Y. Relation between blood pressure and pulse wave velocity for human arteries. *Proceedings of the National Academy of Sciences*, 115(44):11144–11149, 2018.

Moody, G. B. and Mark, R. G. The MIMIC database. PhysioNet, 1992.

Morris, C. J., Hastings, J. A., Boyd, K., Krainski, F., Perhonen, M. A., Scheer, F. A. J. L., and Levine, B. D. Day/night variability in blood pressure: influence of posture and physical activity. *American Journal of Hypertension*, 26(6):822–828, 2013.

Mukkamala, R., Hahn, J. O., Inan, O. T., Mestha, L. K., Kim, C. S., Töreyin, H., and Kyal, S. Toward ubiquitous blood pressure monitoring via pulse transit time: theory and practice. *IEEE Transactions on Biomedical Engineering*, 62(8):1879–1901, 2015.

Pilz, N., Picone, D. S., Patzak, A., Opatz, O. S., Lindner, T., Fesseler, L., Heinz, V., and Bothe, T. L. Cuff-based blood pressure measurement: challenges and solutions. *Blood Pressure*, 33(1), 2024.

Slapničar, G., Mlakar, N., and Luštrek, M. Blood pressure estimation from photoplethysmogram using a spectro-temporal deep neural network. *Sensors*, 19(15), 2019.

Wang, R., Jia, W., Mao, Z. H., Sclabassi, R. J., and Sun, M. Cuff-free blood pressure estimation using pulse transit time and heart rate. In *Proceedings of the 12th International Conference on Signal Processing (ICSP)*, pp. 115–118, 2014.

Yousef, Q., Reaz, M. B. I., and Ali, M. A. M. The analysis of PPG morphology: investigating the effects of aging on arterial compliance. *Measurement Science Review*, 12(6), 2012.

Zhang, L., Hurley, N. C., Ibrahim, B., Spatz, E., Krumholz, H. M., Jafari, R., and Bobak, M. J. Developing personalized models of blood pressure estimation from wearable sensors data using minimally-trained domain adversarial neural networks. In *Proceedings of the 5th Machine Learning for Healthcare Conference*, pp. 97–120, 2020.

Zhang, Y. and Feng, Z. A SVM method for continuous blood pressure estimation from a PPG signal. In *Proceedings of the 9th International Conference on Machine Learning and Computing*, pp. 128–132, 2017.

A. Training Configurations

A.1. Model Training Setup

Table 2. Training configurations for subject-dependent, Global BP pre-training, and Global BP fine-tuning models.

Configuration	Subject-Dependent	GlobalBP Pre-training	GlobalBP Fine-tuning
k-fold	5	–	5
Epochs	50	100	30
Batch size	32	64	16
Early stopping (patience)	7	10	7
Learning rate	1×10^{-3}	1×10^{-3}	5×10^{-4}

The training configurations were designed to reflect the distinct roles of each learning strategy while maintaining stable and efficient optimization. The subject-dependent (SD) models were trained using 5-fold cross-validation to ensure robust evaluation under limited subject-specific data.

For the Global BP pre-training, a larger batch size and increased number of epochs were adopted to facilitate stable learning of population-level representations from diverse subjects. In contrast, the fine-tuning (FT) stage used fewer epochs and a smaller batch size to enable efficient adaptation to subject-specific characteristics without overfitting.

Different early stopping criteria were applied to balance training stability and generalization across settings. Additionally, a reduced learning rate was used during fine-tuning to allow gradual adaptation while preserving the pretrained representations.

A.2. Model Architecture

Table 3. Network architecture of the proposed CNN–BiLSTM–Attention model for blood pressure estimation.

Block	Layer / Operation	Output Dimension	Remarks
Input	Raw PPG segment	$(B, 1, 1250)$	10-second PPG window
CNN Feature Extractor	Conv1D ($k = 7$, $s = 2$) + BN + ReLU	$(B, 32, 625)$	Temporal downsampling
	Conv1D ($k = 5$, $s = 2$) + BN + ReLU	$(B, 64, 313)$	Hierarchical feature learning
	Conv1D ($k = 3$, $s = 2$) + BN + ReLU	$(B, 128, 156)$	Local morphology encoding
Temporal Modeling	2-layer BiLSTM ($h = 128$)	$(B, 156, 256)$	Bidirectional temporal dependency
	LayerNorm	$(B, 156, 256)$	Sequence-level stabilization
Attention	Scaled Additive Attention	$(B, 256)$	Learnable temperature, temporal weighting
Shared FC Head	FC (256 $\rightarrow$ 64) + LN + Tanh	$(B, 64)$	Bounded latent representation
	Dropout	$(B, 64)$	Regularization
Output Heads	FC (64 $\rightarrow$ 1) $\times$ 2	$(B, 2)$	SBP and DBP estimation

The network architecture is designed to capture both local morphological patterns and long-range temporal dependencies in PPG signals. The CNN feature extractor performs hierarchical temporal downsampling while encoding local waveform characteristics. The BiLSTM layer models bidirectional temporal dependencies to capture sequential dynamics across the signal.

An attention mechanism is applied to emphasize informative temporal segments, enabling adaptive weighting of features relevant to blood pressure estimation. The shared fully connected head projects the learned representation into a compact latent space, with layer normalization and nonlinear activation improving stability and expressiveness.

Finally, separate output heads are used to jointly estimate systolic and diastolic blood pressure (SBP and DBP), allowing the model to share representations while learning task-specific mappings.

B. Reference-Level Comparison with Prior Work

Table 4. Reference-level comparison with representative prior studies on PPG-based blood pressure estimation. Direct numerical comparison should be interpreted with caution because datasets, subject compositions, and evaluation protocols differ across studies.

Ref.	Cross-subject train	Subject FT	Performance (MAE, mmHg)	
			SBP	DBP
(Hong et al., 2021)	O	O	9.49	5.54
(Slapničar et al., 2019)	O	X	9.43	6.88
(Khajehpiri et al., 2024)	O	O	7.20	5.70
This work	O	O	4.85 ± 1.55	3.14 ± 0.99

C. Effect of Fine-Tuning Data Ratio on Prediction Performance

C.1. Population-Level Analysis

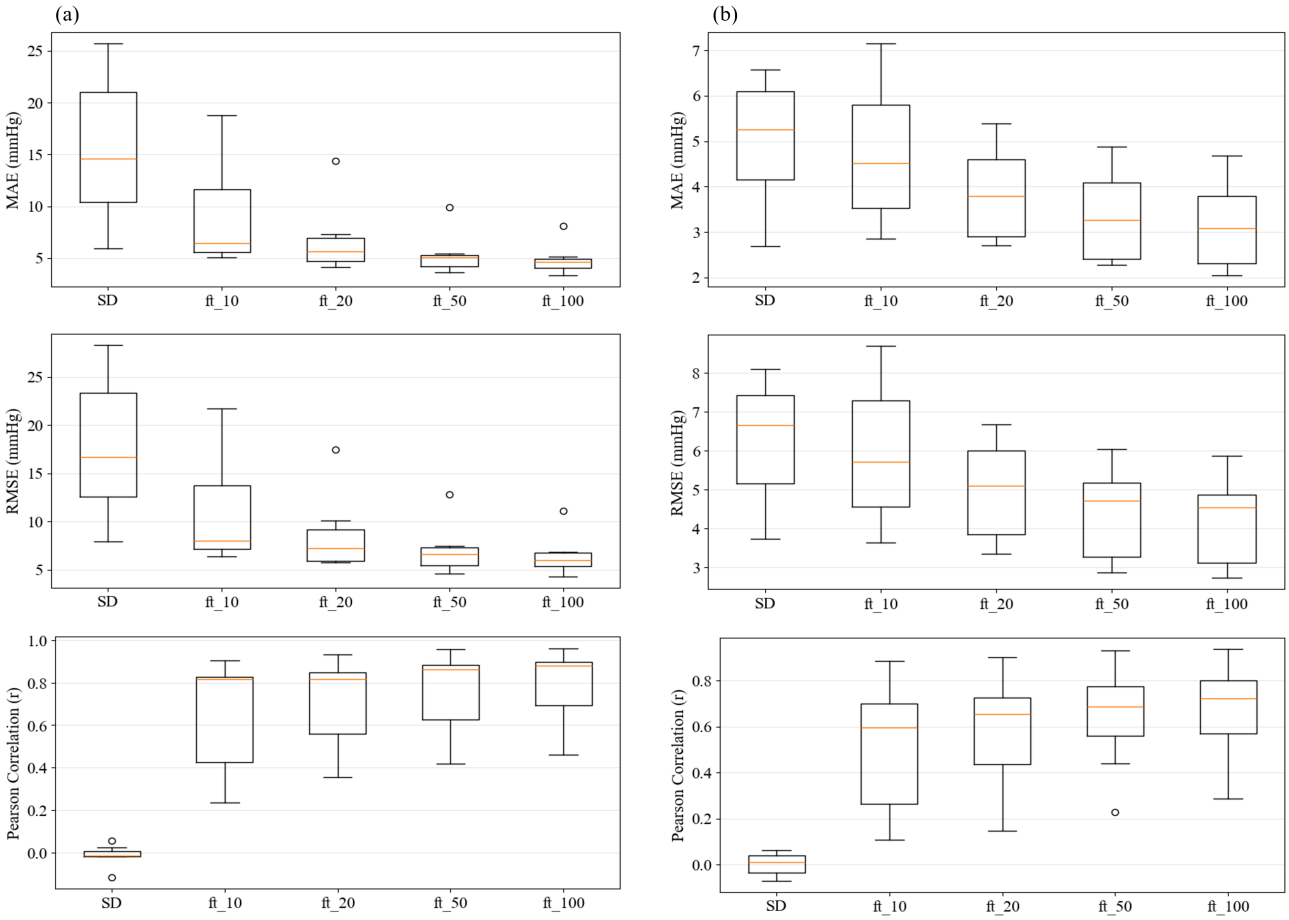

Figure 1. Distribution of estimation performance metrics (MAE, RMSE, and Pearson correlation coefficient) across different fine-tuning data ratios. Boxplots summarize population-level performance for the subject-dependent (SD) model and fine-tuned models using increasing proportions of subject-specific data (10%, 20%, 50%, and 100%). (a) SBP; (b) DBP.

Figure 1 presents the population-level distribution of blood pressure estimation performance across different fine-tuning data ratios. The boxplots compare the subject-dependent (SD) model with fine-tuned models using increasing proportions of subject-specific data (10%, 20%, 50%, and 100%) for both SBP and DBP.

Overall, the SD model shows relatively large MAE and RMSE values, with a wide distribution across subjects, indicating substantial variability in estimation performance. In contrast, fine-tuning from the population-level model substantially

reduces MAE and RMSE, even when only 10% of the subject-specific data is used. As the fine-tuning data ratio increases, the error metrics further decrease and gradually stabilize, suggesting that additional personal data provides diminishing gains beyond a certain point.

For the Pearson correlation coefficient (r), the fine-tuned models generally show higher values than the SD model. This indicates that population-level initialization improves the ability to track blood pressure variability across subjects, in addition to reducing absolute estimation errors. The improvement in r is particularly evident in the early fine-tuning stages and tends to stabilize as more subject-specific data is used.

Overall, these results demonstrate that fine-tuning initialized from population-level learning can improve both estimation accuracy and stability under limited subject-specific data conditions, compared with purely subject-dependent learning.

C.2. Subject-Level Analysis

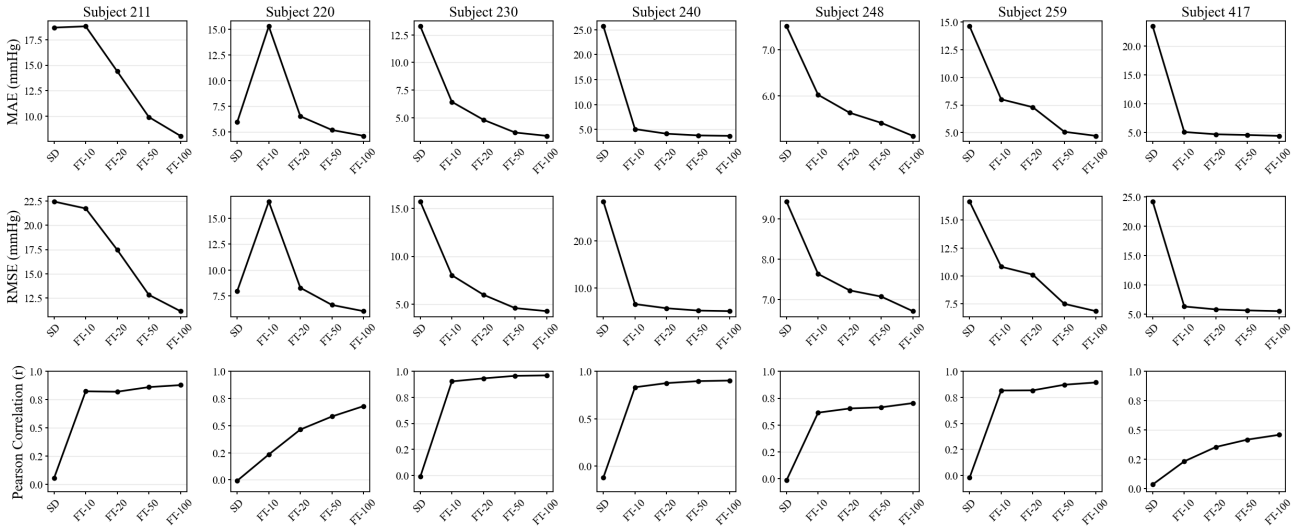

Figure 2. Subject-level trends of systolic blood pressure (SBP) estimation performance across different fine-tuning data ratios. For each subject, the left, middle, and right panels show MAE, RMSE, and Pearson correlation coefficient (r), respectively, for the subject-dependent (SD) model and fine-tuned models using increasing proportions of subject-specific data (10%, 20%, 50%, and 100%).

Figure 2 shows subject-level changes in SBP estimation performance with increasing fine-tuning data ratios for unstable subjects. The subject-dependent (SD) model exhibits large MAE and RMSE with high variability, and in some cases very low Pearson correlation, indicating poor tracking of BP variability. In contrast, fine-tuning rapidly reduces MAE and RMSE for most subjects, with substantial improvements observed even with 10% of personal data. Performance generally stabilizes beyond a 20% data ratio, with diminishing returns from additional data.

For the Pearson correlation coefficient (r), sharp improvements are observed in early fine-tuning stages, suggesting that variability tracking is restored prior to reductions in absolute error. As the data ratio increases, r further improves or stabilizes, indicating enhanced consistency and stability.

Figure 3 shows subject-level changes in DBP estimation performance with increasing fine-tuning data ratios for unstable subjects. The subject-dependent (SD) model generally exhibits relatively small MAE and RMSE, but low Pearson correlation (r) for some subjects, indicating insufficient tracking of DBP variability. Fine-tuning gradually reduces MAE and RMSE for most subjects. Although the magnitude of error reduction is smaller than in SBP, both error and variability consistently decrease as the data ratio increases, showing stable convergence.

For the Pearson correlation coefficient (r), sharp improvements are observed in early fine-tuning stages, suggesting that variability tracking is restored prior to reductions in absolute error. As the data ratio increases, r further improves or stabilizes, indicating enhanced consistency and reliability.

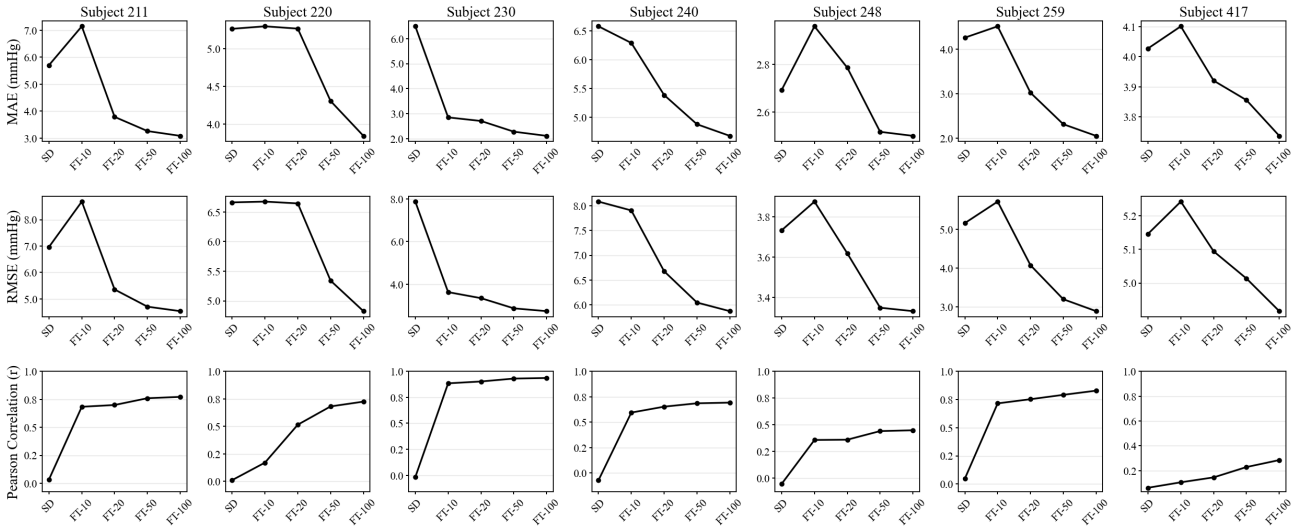

Figure 3. Subject-level trends of diastolic blood pressure (DBP) estimation performance across different fine-tuning data ratios. For each subject, the left, middle, and right panels show MAE, RMSE, and Pearson correlation coefficient (r), respectively, for the subject-dependent (SD) model and fine-tuned models using increasing proportions of subject-specific data (10%, 20%, 50%, and 100%).

D. Subject-Level Scatter and Bland–Altman Analysis

This section presents subject-level scatter and Bland–Altman analyses comparing the subject-dependent (SD) and fine-tuned (FT) models. Because the distributions of SBP and DBP in the test data vary substantially across subjects, scatter plots alone are insufficient to fully interpret estimation behavior. Therefore, the test data distributions are presented alongside the scatter and Bland–Altman plots to provide clearer context for subject-specific estimation characteristics.

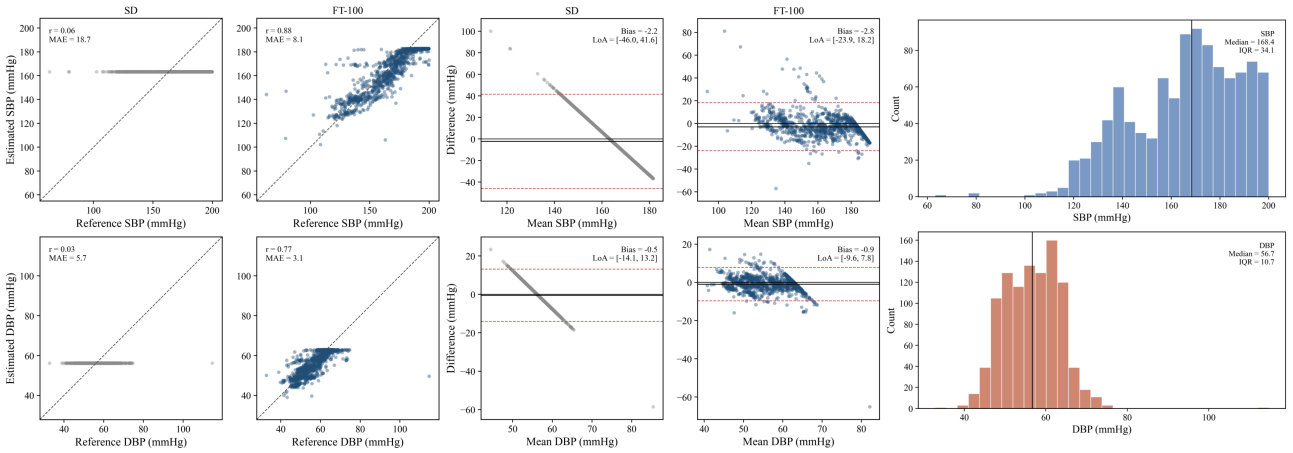

Figure 4. Estimation behavior for Subject 211. Scatter plots, Bland–Altman plots, and BP histograms from left to right. SD: subject-dependent model; FT: fine-tuned model; FT100: fine-tuned using the same amount of subject-specific data as SD.

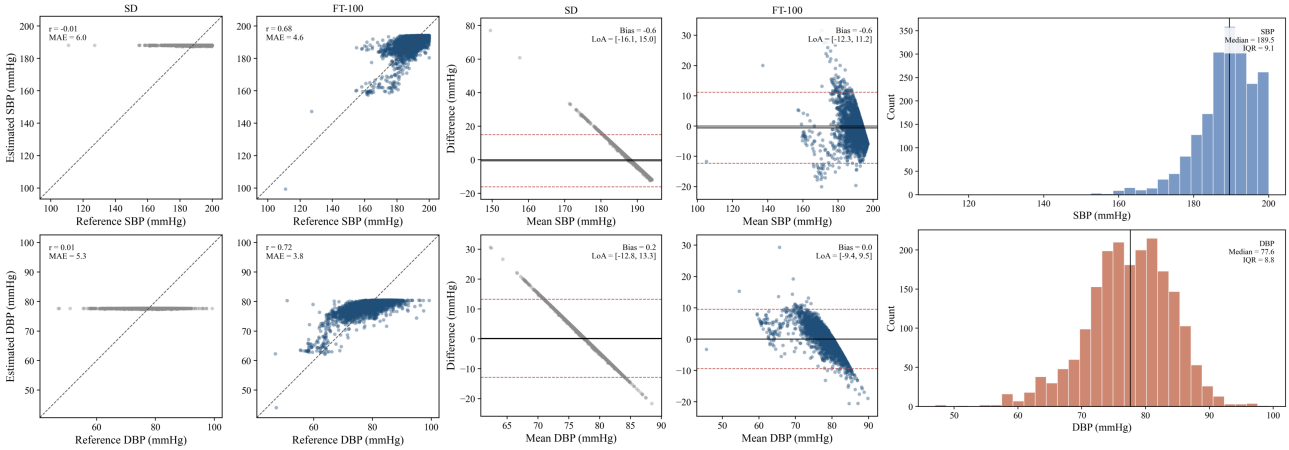

Figure 5. Estimation behavior for Subject 220. Scatter plots, Bland–Altman plots, and BP histograms from left to right. SD: subject-dependent model; FT: fine-tuned model; FT100: fine-tuned using the same amount of subject-specific data as SD.

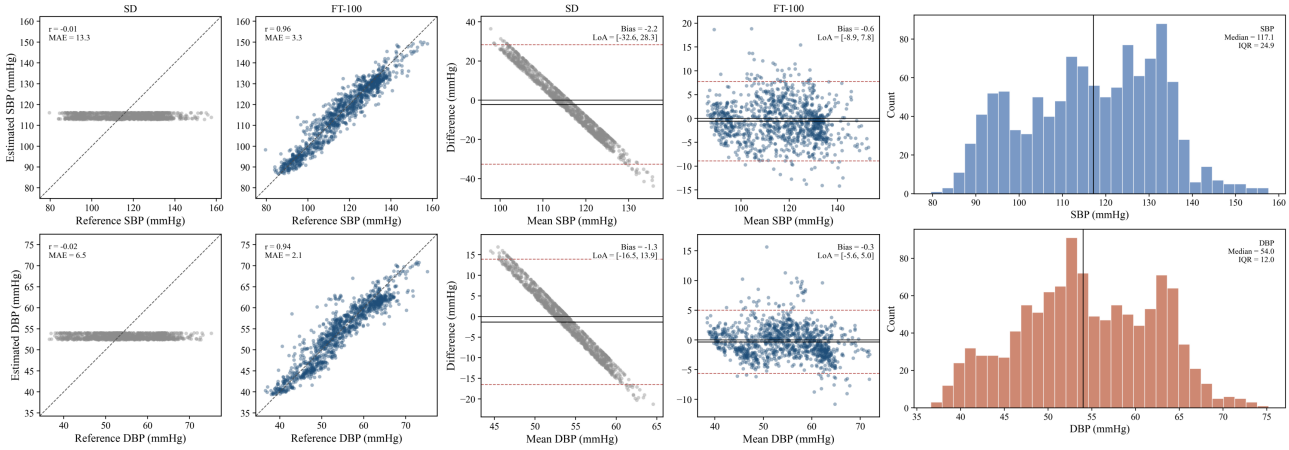

Figure 6. Estimation behavior for Subject 230. Scatter plots, Bland–Altman plots, and BP histograms from left to right. SD: subject-dependent model; FT: fine-tuned model; FT100: fine-tuned using the same amount of subject-specific data as SD.

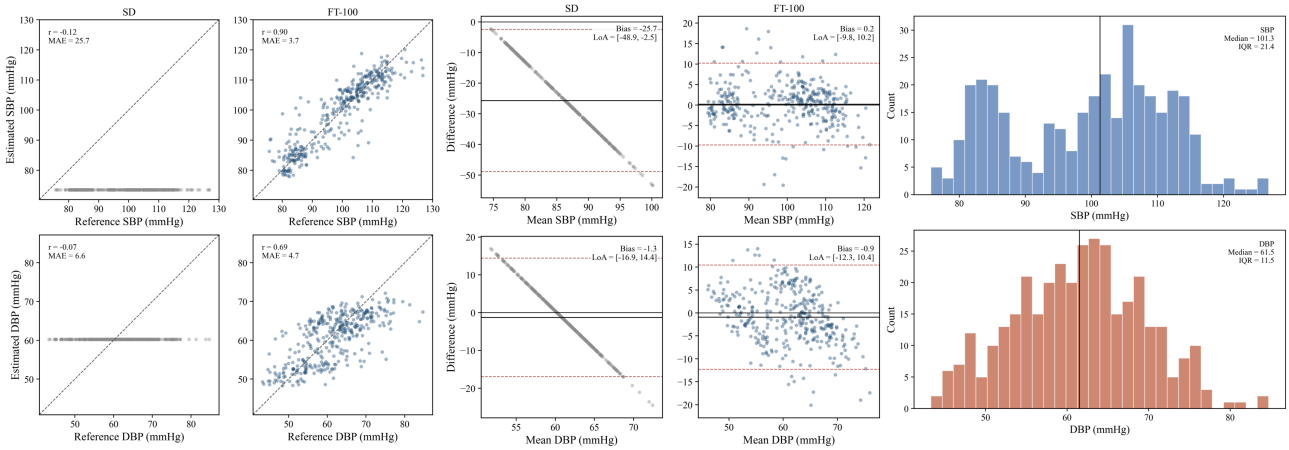

Figure 7. Estimation behavior for Subject 240. Scatter plots, Bland–Altman plots, and BP histograms from left to right. SD: subject-dependent model; FT: fine-tuned model; FT100: fine-tuned using the same amount of subject-specific data as SD.

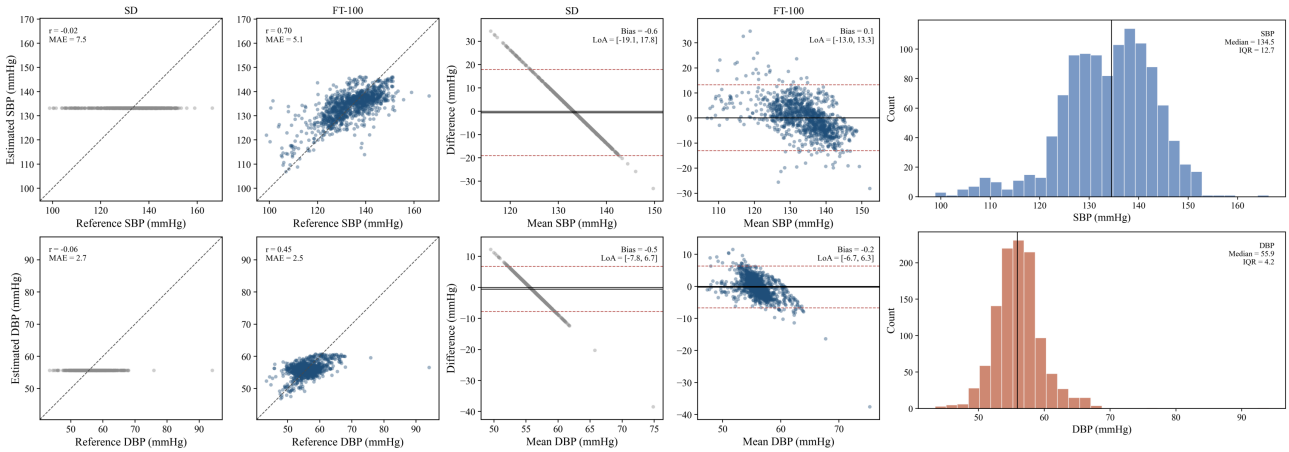

Figure 8. Estimation behavior for Subject 248. Scatter plots, Bland–Altman plots, and BP histograms from left to right. SD: subject-dependent model; FT: fine-tuned model; FT100: fine-tuned using the same amount of subject-specific data as SD.

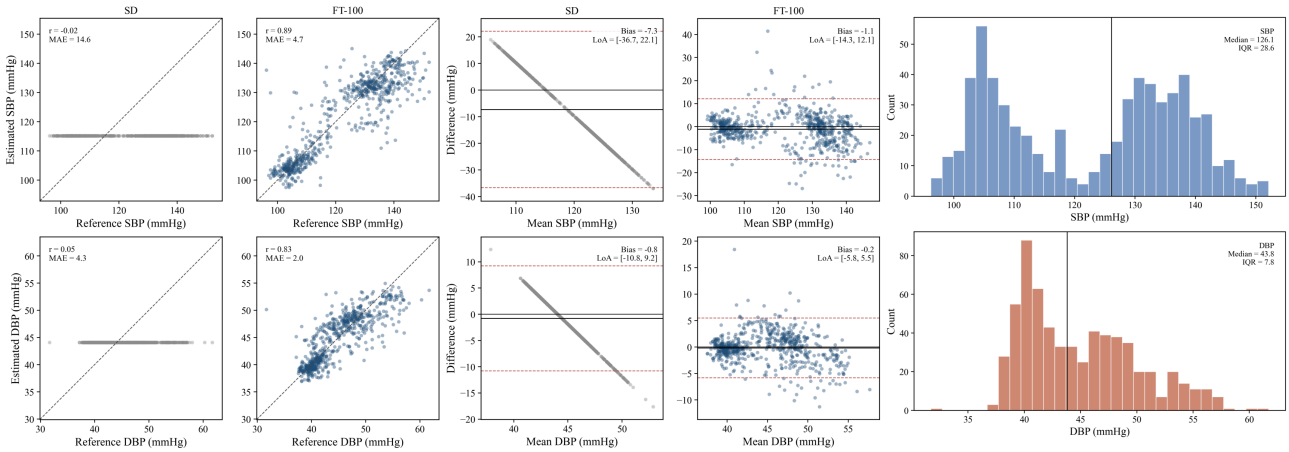

Figure 9. Estimation behavior for Subject 259. Scatter plots, Bland–Altman plots, and BP histograms from left to right. SD: subject-dependent model; FT: fine-tuned model; FT100: fine-tuned using the same amount of subject-specific data as SD.

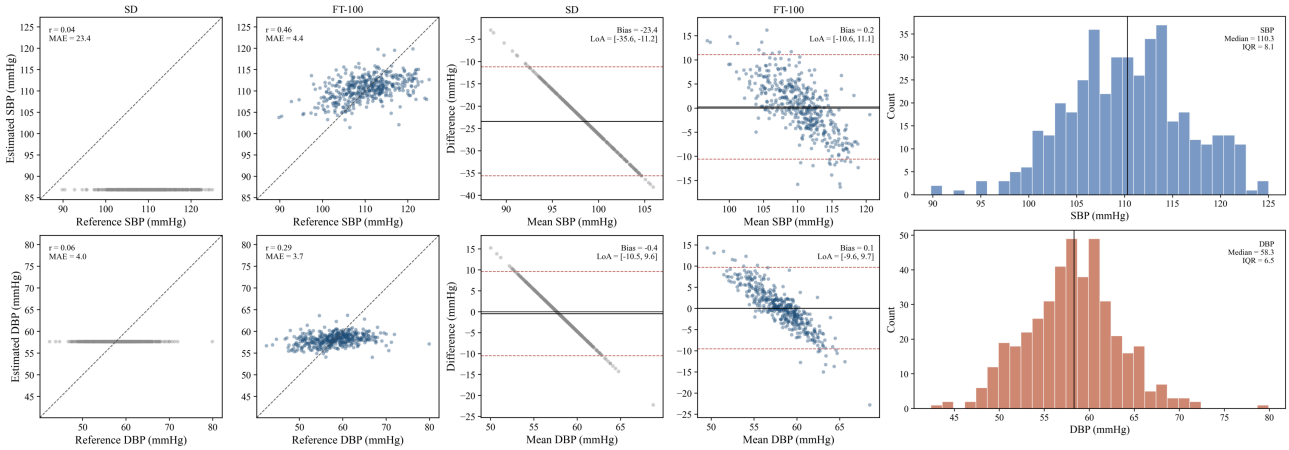

Figure 10. Estimation behavior for Subject 417. Scatter plots, Bland–Altman plots, and BP histograms from left to right. SD: subject-dependent model; FT: fine-tuned model; FT100: fine-tuned using the same amount of subject-specific data as SD.