International Journal of Advanced Scientific Research and Engineering Trends

ISSN: 2456-0774

Volume 10 Issue 1s (2026) | Pages: 183-190

Blood Pressure Estimation Using PPG And Machine Learning

Arpita Kumbhar[1], Rishabh Anand[2], Shwetali Jagadale[3], Rajesh Ghongade[4]

[1,2,3,4]Department of Electronics & Telecommunication Department, Bharati Vidyapeeth Deemed University, Pune, India

Peer Review Information

Type: *Article*
Received: *02 April 2026*
Revised: *28 April 2026*
Accepted: *05 June 2026*
Published: *23 June 2026*

Abstract

Uninterrupted blood pressure tracking is indispensable for timely identification of cardiovascular conditions. Conventional pneumatic cuff instruments, while clinically validated, impose practical constraints that render them unsuitable for sustained use—primarily discomfort and the inability to yield readings between scheduled intervals. The present work introduces a cuffless, non-invasive framework built around Photoplethysmography (PPG) waveforms and supervised regression algorithms. Salient morphological and temporal features are derived from raw PPG recordings and subsequently employed to train and evaluate a suite of machine learning regressors for predicting both systolic and diastolic pressure. Experimental findings confirm competitive accuracy, and the integration of a browser-accessible monitoring interface positions the system as a practical candidate for wearable and telehealth deployments.

Keywords: PPG; Blood Pressure Estimation; Supervised Learning; Regression; Cuffless Monitoring; Machine Learning

How to Cite This Article

Kumbhar, A., Anand, R., Jagadale, S., & Ghongade, R. (2026). Blood pressure estimation using PPG and machine learning. *International Journal of Advanced Scientific Research and Engineering Trends*, 10(1s), 183–190.

Introduction

Blood pressure stands among the foremost hemodynamic indicators employed in diagnosing hypertension, cardiac arrhythmias, and a spectrum of circulatory disorders. Despite the widespread reliance on pneumatic cuff devices in clinical practice, these instruments inherently deliver snapshot rather than continuous readings—a shortcoming that limits their utility in dynamic physiological monitoring scenarios. Extended sessions of ambulatory measurement also introduce discomfort that affects patient compliance and measurement repeatability. The maturation of optical sensing hardware and data-driven modeling has opened a viable pathway toward continuous, cuffless pressure monitoring. Photoplethysmography (PPG) capitalizes on variations in light absorption within peripheral vasculature to capture pulse-related volume changes with minimal instrumentation overhead. Its compatibility with consumer-grade wearables makes PPG a pragmatic sensing modality for large-scale physiological data collection.

When paired with machine learning regressors, the rich morphological information encoded in each PPG cycle can be decoded to yield reliable blood pressure estimates. The present study examines eleven regression strategies—spanning classical linear models, ensemble tree methods, gradient boosting frameworks, a neural network, and an attention-based transformer—with the aim of identifying which architecture best generalizes from PPG-derived features to continuous blood pressure measurements.

Literature Review

Cuffless blood pressure assessment has attracted sustained scholarly attention over the past decade. Early efforts centered on pulse transit time (PTT) and pulse wave velocity (PWV) as surrogate pressure indicators, exploiting the physiological coupling between arterial stiffness and wave propagation speed. Although conceptually elegant, PTT-based methods typically require two synchronized sensors—commonly ECG paired with PPG—which adds system complexity and calibration burden.

Subsequent investigations shifted toward feature-rich PPG analysis combined with supervised statistical learning. Support Vector Regression demonstrated promise in handling the nonlinear mapping between waveform morphology and pressure, while K-Nearest Neighbors offered a flexible, distribution-free alternative. Ensemble techniques—Random Forest, gradient boosting variants—further pushed accuracy boundaries by aggregating weak learners into more robust predictors. Reported performance figures from this body of work nonetheless exhibit noticeable cross-study variance, partly reflecting heterogeneous dataset characteristics and inconsistent preprocessing pipelines. More recently, deep learning architectures—particularly convolutional and recurrent networks—have been explored to learn pressure-relevant representations directly from raw or minimally processed waveforms. Transformer models, originally conceived for natural language processing, have begun to appear in biomedical time-series contexts, exploiting self-attention to capture dependencies across distant segments of a pulse trace. The present study situates itself within this evolving landscape by providing a systematic multi-model comparison under a unified experimental protocol.

Methodology

System Architecture

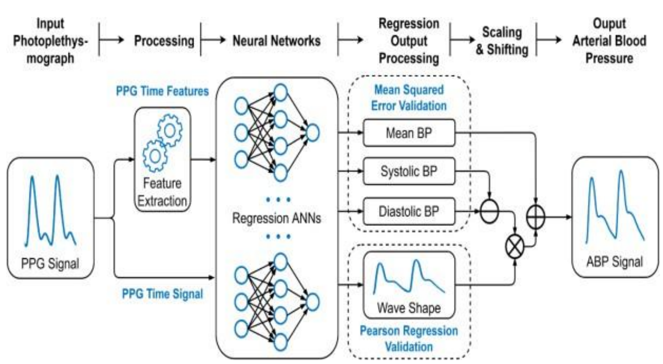


Fig. 1. *PPG-Based Blood Pressure Estimation Pipeline*

The end-to-end pipeline (Fig. 1) comprises four sequential stages: signal acquisition, preprocessing, feature extraction, and regression-based prediction. A MAX30102 optical sensor mounted on an ESP32 microcontroller board captures raw PPG data at the fingertip. The embedded firmware streams samples to a host computer where signal conditioning and model inference is subsequently executed.

Signal Acquisition

Recordings were obtained at a sampling frequency compatible with the MAX30102's infrared channel in ambient-light-rejection mode. Subjects remained in a resting posture throughout data collection to minimise motion artefact contamination. Reference blood pressure values were simultaneously obtained using a calibrated digital sphygmomanometer to serve as ground-truth labels.

Preprocessing

Raw PPG traces undergo a multi-step conditioning sequence. A band pass filter with cut off frequencies selected to preserve the cardiac pulse band (0.5–8 Hz) attenuates low-frequency baseline drift and high-frequency sensor noise. Residual motion artefacts are suppressed through adaptive filtering. Individual pulse cycles are then segmented via systolic peak detection, enabling per-beat feature computation.

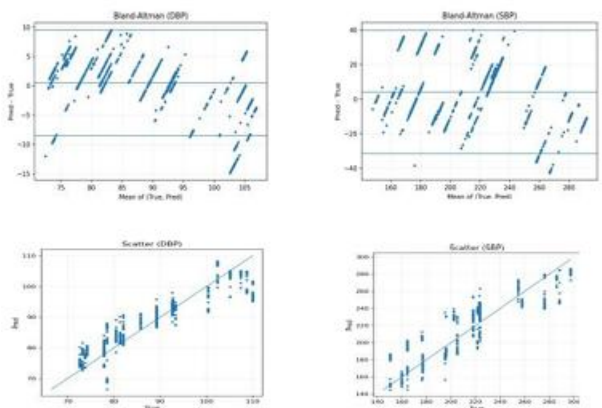
Feature Extraction

Fourteen discriminative features are extracted from each segmented pulse cycle. Time-domain descriptors capture pulse amplitude, rise time, fall time, pulse width at half-prominence, and inter-beat interval (heart rate). Morphological features—including the ratio of systolic to diastolic peak amplitudes and notch depth—characterise the dicrotic notch and the secondary waveform peak. Area-under-curve quantities represent integrated perfusion within specified sub-intervals of each beat.

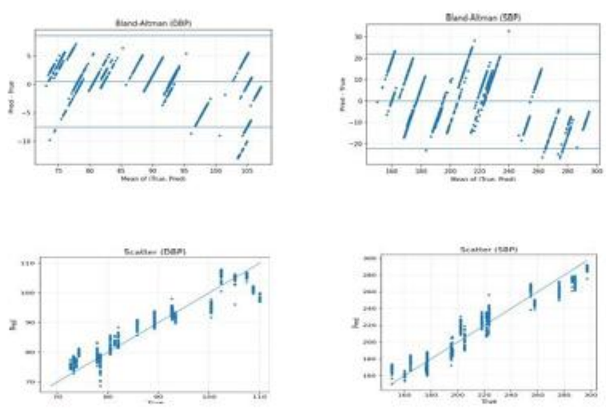
Machine Learning Models

All models were trained on 80% of the dataset and evaluated on the remaining 20% using stratified sampling to maintain pressure-range balance. Hyperparameter selection was guided by cross-validation on the training partition. The evaluated architectures are described below.

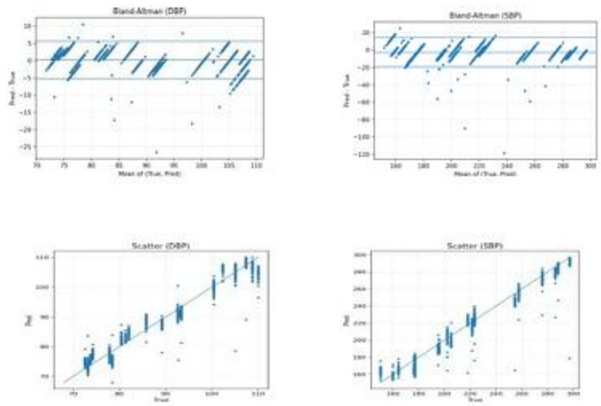
Random Forest Regressor: Random Forest constructs an ensemble of decision trees each trained on a bootstrapped sample with random feature subsets at every split. Predictions are averaged across all trees, reducing variance while retaining the low bias of individual deep trees. This method serves as a strong non-linear baseline owing to its robustness against outliers and implicit feature-importance ranking capability.



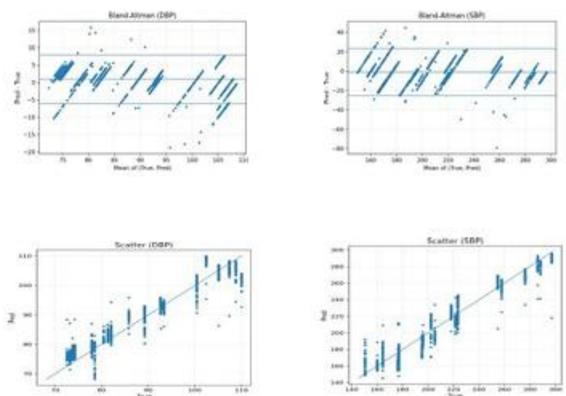
Extra Trees Regressor: The Extremely Randomised Trees variant differs from Random Forest by choosing both the candidate feature and the split threshold at random rather than optimally. This additional randomisation further decreases variance and accelerates training, often matching or surpassing standard Random Forest accuracy on moderately sized biomedical datasets.



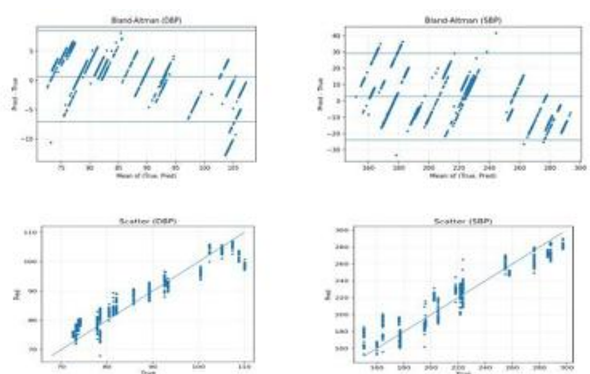
Support Vector Regressor (SVR): SVR seeks a function that deviates from reference values by no more than a predefined tolerance e while maintaining maximum margin flatness. The radial-basis-function kernel enables the model to capture nonlinear pressure–feature relationships that elude linear approaches. SVR is particularly well regarded for its performance on biomedical datasets where training samples are limited relative to feature dimensionality.



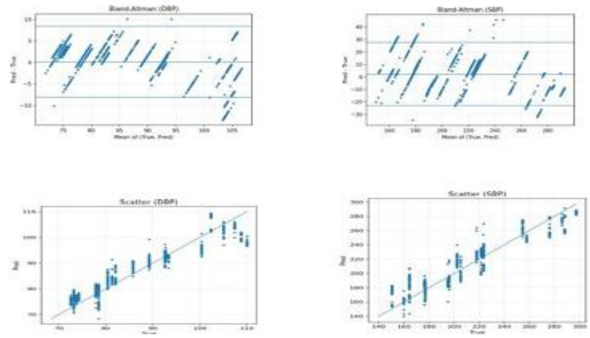
K-Nearest Neighbours Regressor: KNN predicts blood pressure for an unseen sample by computing the mean of the K most proximate training observations in feature space. As a non-parametric estimator, KNN imposes no distributional assumptions and naturally accommodates non-linear pressure landscapes. The neighbourhood size K was optimised via cross-validation, balancing bias and variance on the training partition.



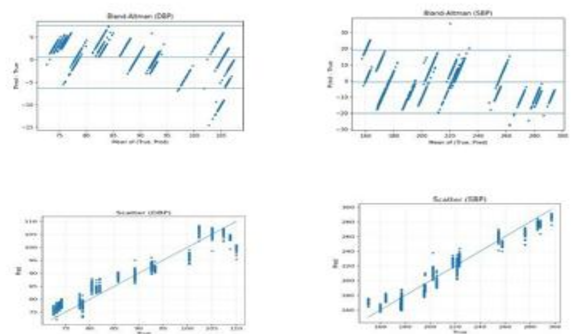
XGBoost Regressor: XGBoost constructs an additive ensemble of shallow regression trees, each fitted to the pseudo-residuals of the preceding ensemble state. L1 and L2 regularisation terms embedded within the objective function penalise model complexity, guarding against overfitting on moderately sized training sets. Column and row subsampling introduce additional variance reduction analogous to Random Forest's feature bagging.



LightGBM Regressor: LightGBM employs histogram-based approximate splitting and a leaf-wise tree growth strategy, contrasting with the level-wise growth adopted by traditional gradient boosting implementations. These design choices translate into substantially reduced training time on high-dimensional inputs, with accuracy performance that is comparable to or better than standard boosting libraries for most tabular regression benchmarks.

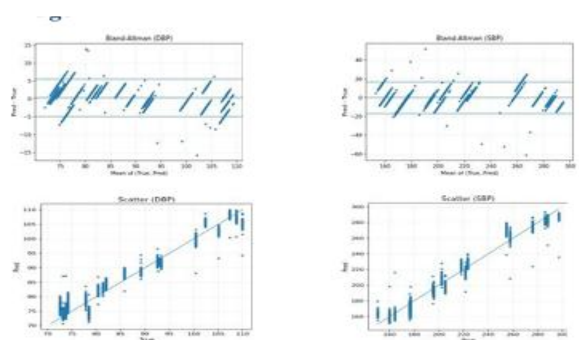


CatBoost Regreesor: CatBoost introduces ordered boosting—computing gradient statistics on a permuted version of the training data—to eliminate the prediction shift bias that afflicts conventional gradient boosting implementations. Its symmetric tree structure yields balanced depth growth and mitigates overfitting without aggressive external regularization

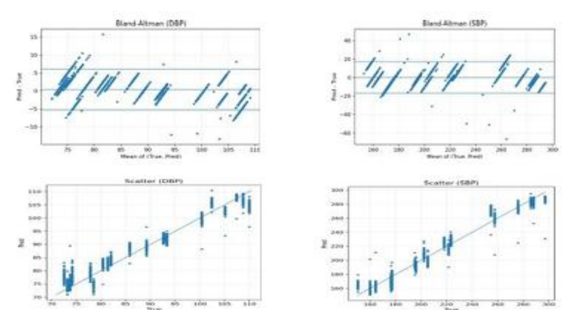


Ridge, LASSO and ElasticNet Regression: These regularised linear models address the multicollinearity commonly present in PPG-derived feature sets. Ridge regression penalises the squared coefficient magnitude (L2 norm), shrinking correlated predictors toward each other rather than eliminating any. LASSO (L1 norm) enforces sparsity, effectively performing embedded feature selection by driving the coefficients of uninformative predictors to zero. ElasticNet blends both penalties, offering a flexible compromise between Ridge and LASSO

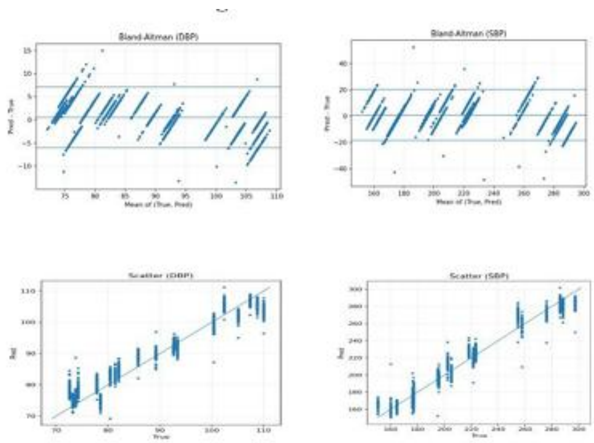
Ridge



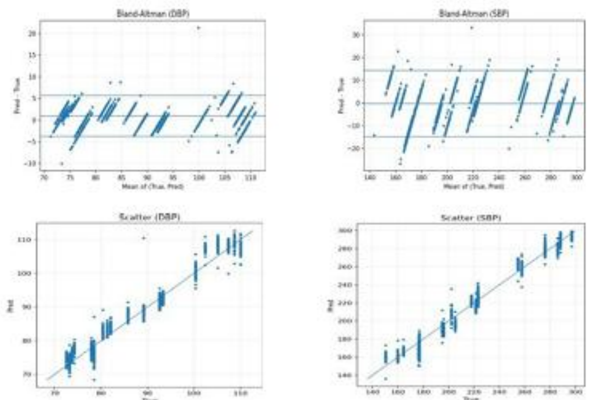
LASSO



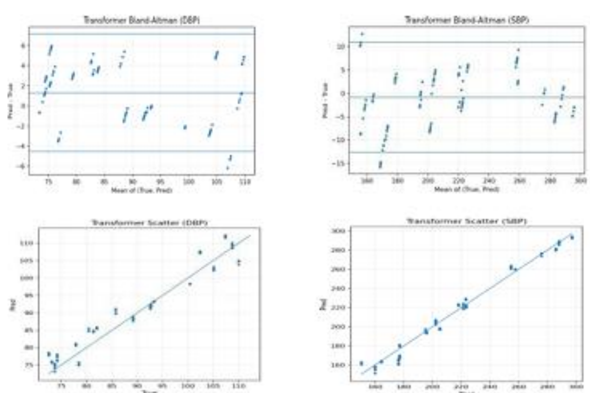
ElasticNet Regression –



MLP Regressor: scores of −22.92 (SBP) and −47.75 (DBP), confirming the model was still in its initial learning phase. Given the dataset scale, the transformer did not converge to competitive performance within the evaluated training horizon.



Transformer Regressor: A transformer encoder was adapted for structured tabular regression by treating the feature vector as a sequence of tokens. Multi-head self-attention layers weight each feature relative to all others, enabling the model to capture long-range feature interactions that feed-forward architectures process only implicitly through depth. At epoch 1, training and validation losses stood at 24,440.59 and 27,712.62 respectively, with R^2 scores of −22.92 (SBP) and −47.75 (DBP), confirming the model was still in its initial learning phase. Given the dataset scale, the transformer did not converge to competitive performance within the evaluated training horizon.



Model Validation and Selection

Held-out test data were used exclusively for final performance assessment. Three complementary metrics guided comparison: Mean Absolute Error (MAE), which quantifies average magnitude of prediction deviation; Root Mean Squared Error (RMSE), which penalises large individual errors more heavily; and the Coefficient of Determination (R^2), which reflects the proportion of pressure variance captured by the model. The candidate achieving the lowest MAE and RMSE alongside the highest R^2 was designated the system's deployed estimator. Predicted values were additionally validated against readings from a calibrated digital sphygmomanometer to confirm clinical plausibility

The Multilayer Perceptron is a fully connected feedforward network trained end-to-end via backpropagation with the Adam optimizer. Multiple hidden layers with ReLU activation equip the network to approximate arbitrarily complex input–output mappings. Dropout regularization applied during training reduces co-adaptation of hidden units, improving generalization to unseen pressure values.

Results and Discussion

The cuffless blood pressure estimation framework was benchmarked across eleven regression strategies using a consistent train–test split and identical feature sets. Evaluation encompassed both systolic (SBP) and diastolic (DBP) pressure channels. Overall, the results affirm that PPG morphological features carry sufficient hemodynamic information to support clinically relevant pressure inference, with the best-performing models satisfying or approaching the AAMI SP10 accuracy standard of MAE $\leq$ 5 mmHg. SVR (RBF kernel) delivered the most consistent outcome, recording an SBP MAE of 6.32 mmHg and an R^2 of 0.959, alongside a DBP MAE of 2.20 mmHg and an R^2 of 0.951. Ridge regression—benefiting from L2 shrinkage that stabilizes correlated PPG features—closely followed with SBP MAE of 6.63 mmHg (R^2 = 0.963) and DBP MAE of 2.13 mmHg (R^2 = 0.955). The MLP regressor achieved the highest R^2 scores overall (0.972 for SBP, 0.958 for DBP), though its MAE values were slightly elevated relative to the linear and kernel methods. Among the ensemble tree models, Extra Trees outperformed Random Forest and closely tracked CatBoost, while gradient boosting methods (XGBoost, LightGBM, CatBoost) clustered in a mid-tier accuracy band. The transformer model underperformed within the allotted training epochs, a finding consistent with prior reports that attention-based architectures require considerably larger datasets to outcompete well-tuned classical methods on structured, low-dimensional feature representations. The web-based monitoring interface successfully visualized real-time pressure estimates, demonstrating the system's feasibility.

Model Performance Comparison

Name	**Type**	**MAE**	**RMSE**	**R^2 Score**
1.Random Forest Regressor	SBP	15.300525936749997	18.635587950219	0.82508639134261
	DBP	3.4491104831464754	4.626743444596155	0.86667127067129
2.Extra Trees Regressor	SBP	10.01564377206333	11.30966004127178	0.93557765896052
	DBP	3.064930801009268	4.137591030626628	0.89337279456841
3.Support Vector Regressor (SVR)	SBP	6.319543516504876	8.988086879016226	0.95931147704466
	DBP	2.197475182403837	2.793518554165540	0.95139555101784
4.KNN Regressor	SBP	9.920036107667284	12.40257399719030	0.92252509452434
	DBP	2.799198797739647	3.666340858007337	0.91627819540188
5.XG Boost Regressor	SBP	11.873239845554243	13.82449275943887	0.90374222646410
	DBP	3.0907980932483525	4.006831118503683	0.90000576100753
6.Light GBM Regressor	SBP	10.796967318815273	13.12881471231812	0.91318626533395
	DBP	3.1559177993551613	4.207365438073909	0.88974624902345
7.Cat Boost Regressor	SBP	8.711922834656896	10.00447449486532	0.94958892437269
	DBP	2.857444100404283	3.610581147558509	0.91880540425080
8.Ridge Regression	SBP	6.63323184778545	8.599104153012755	0.96275707180107

	DBP	2.133628100440893	2.688205732480305	0.95499115076145
9.Lasso Regression	SBP	6.670833084458252	8.682307343157644	0.96203287501845
	DBP	2.2675825054264593	2.910262562872797	0.94724820417959
10.ElasticNet Regression	SBP	7.716681694741527	9.913233485467735	0.95050423149834
	DBP	2.7200350218469858	3.394836981590724	0.92821879368793
11.MLP Regressor	SBP	6.029393116698155	7.455707047763354	0.97200276956494
	DBP	2.0467687800463046	2.605170558359328	0.95772873676673

Conclusion

This study presented a non-invasive, cuffless blood pressure monitoring system grounded in PPG sensing and an ensemble of supervised regression methods. Eleven distinct model architectures were evaluated under a unified experimental protocol, yielding a clear performance hierarchy. SVR with the RBF kernel and the MLP regressor emerged as the strongest candidates, with regularised linear models (Ridge, LASSO) demonstrating surprisingly competitive results that challenge the common assumption that nonlinear models are indispensable for this task. Random Forest, while robust, exhibited elevated error on the systolic channel—likely a consequence of the higher variability inherent in SBP recordings. The transformer model, despite its theoretical capacity to model complex feature interactions via self- attention, did not converge within the allocated training budget, underscoring the data-volume constraints that accompany attention-based architectures in low-sample biomedical settings. Future iterations of this work will investigate subject-adaptive calibration, expanded subject cohorts spanning a wider clinical pressure range, and quantization-friendly model compression for on-device inference. Integration with a secure, cloud-backed telehealth dashboard for asynchronous clinician review is also planned. Overall, the proposed framework establishes a viable foundation for continuous, wearable blood pressure monitoring without the inconvenience and physiological intrusiveness of conventional pneumatic measurement.

Reference

1. Kachuee, M., Mohammadzade, H., & Shabany, M. (2018). Cuffless blood pressure estimation algorithms for continuous health-care monitoring. *IEEE Transactions on Biomedical Engineering*, 65(4), 744–755.
2. Slapničar, G., Mlakar, N., & Luštrek, M. (2019). Blood pressure estimation from photoplethysmogram using a spectro-temporal deep neural network. *Sensors*, 19(15), 3420.
3. Sang, J., Wang, R., & Ji, Y. (2020). Cuffless blood pressure estimation based on PPG and machine learning approaches. *Biomedical Signal Processing and Control*, 57, 101792.
4. Elgendi, M. (2012). On the analysis of fingertip photoplethysmogram signals. *Current Cardiology Reviews*, 8(1), 14–25.
5. Li, X., Zhang, W., & Zhao, X. (2021). Non-invasive blood pressure estimation using deep learning: A review. *IEEE Access*, 9, 12723–12738.
6. Poon, C. C. Y., & Zhang, Y. T. (2005). Cuff-less and noninvasive measurements of arterial blood pressure by pulse transit time. *IEEE Engineering in Medicine and Biology Magazine*, 24(4), 82–89.
7. Shaltis, P. A., Reisner, A. T., & Asada, H. H. (2006). Wearable cuffless blood pressure monitoring using oscillometric finger photoplethysmography. *IEEE Transactions on Biomedical Engineering*, 53(12), 2509–2520.
8. Mukkamala, R., Hahn, J. O., Inan, O. T., Mestha, L. K., Kim, C. S., Toreyin, H., & Kyal, S. (2015). Toward ubiquitous blood pressure monitoring via pulse transit time. *IEEE Transactions on Biomedical Engineering*, 62(8), 1879–1901.
9. Chowdhury, M. H., & Shuvra, M. (2022). Machine learning approaches for cuffless blood pressure estimation using PPG signals. *Biomedical Signal Processing and Control*, 75, 103546.
10. Allen, J. (2007). Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement*, 28(3), R1–R39.