Classification of Hypertension Using Support Vector Machine Based on Data Photoplethysmography and Blood Pressure Estimator

Glen Virdho Sirait[a], Anand Roshan[b], Delima Sitanggang[c], Reyhan Achmad Rizal[d]
[a,b,c,d] Universitas Prima Indonesia
Corresponding Author:
[c]delimasitanggang@unprimdn.ac.id

ABSTRACT
This study formulates a Support Vector Machine (SVM)-based hypertension classification model using physiological features from Photoplethysmograph (PPG) signals and systolicdiastolic blood pressure estimates. PPG data were collected from 276 participants in three locations using a non-invasive device integrated into the MR-IAT Robot Covid web platform. Feature extraction and preprocessing were performed in MATLAB, while SVM (scikit-learn) model training and testing were conducted in Google Colab. Comparison of three kernel variants—linear, quadratic, and cubic—showed that the cubic kernel was the most superior with an accuracy of 96.4%, followed by quadratic at 94.9% and linear at 91.3%. Overall, the model achieved 93.9% accuracy in distinguishing six blood pressure categories (Hypotension, Normal, Prehypertension, Stage 1 Hypertension, Stage 2 Hypertension, Crisis). Visualization of the results (scatter plot, confusion matrix, parallel coordinates) revealed that systolic pressure, diastolic pressure, age, weight, and respiration rate were the most influential parameters. These findings underscore the potential application of SVM in portable devices for early detection and real-time monitoring of hypertension.

Keywords : Hypertension Classification, Support Vector Machine (SVM), Photoplethysmograph, Blood Pressure Estimation, Non-Invasive Monitoring.

INTRODUCTION

In medical practice, hypertension is still recorded as one of the leading causes of death worldwide and often goes undetected in its early stages, earning it the nickname "silent killer" (Arulselvi, Jeyakumar, Vidya, Krishna Kumar, & Lakshmi, 2024). Technological advances have spurred the development of non-invasive blood pressure classification using photoplethysmograph (PPG) signals, particularly with the support of machine learning methods(Dehghanojamahalleh & Kaya, 2019). Among various techniques, Support Vector Machine (SVM) stands out for its ability to process high-dimensional data and model non-linear relationships efficiently (Dwaish & Miften, 2025). SVM searches for the optimal hyperplane to maximize class separation, making it suitable for physiological signal-based blood pressure classification (Dwaish, 2024).

Various studies have proven that the integration of PPG and SVM produces highly accurate blood pressure status predictions (Fuadah & Lim, 2022). In addition, the PPG method is considered more convenient and portable than conventional cuff-based measurements (Jeyakumar, 2024). The maximum margin concept in SVM is also effective in reducing the risk of overfitting when compared to other algorithms such as Decision Tree (Kim et al., 2023). The latest least squares SVM approach has even been proven to be more accurate in blood pressure estimation (M., N. P. M., & Cigi, 2024).

PPG features including pulse rise time, beat interval variability, and pulse wave velocity have been proven crucial in recognizing hypertension patterns with SVM (M., J., & Nabeel, 2024). Recent studies have also explored the integration of SVM with feature extraction based on wavelet transformation or PCA to improve classification performance (Martínez et al., 2025). The application of SVM is not only effective for PPG data from the finger, but also for remote PPG measurements on the face or wrist (Martínez-Ríos, Montesinos, & Alfaro-Ponce, 2022). The implementation of SVM in real-time blood pressure prediction systems shows high computational efficiency with an accuracy of over 90% (Mejía-Mejía, Budidha, Kyriacou, & Mamouei, 2022), opening up great opportunities for wearable devices for early detection of hypertension (Raj, 2023). A study from India reported SVM's ability to distinguish five categories of blood pressure with high sensitivity and specificity (S. & Bafandeh, 2021). Gaussian kernel-based SVM also excels in modeling the non-linear relationship between PPG features and blood pressure values darah (Singh, 2023).

With the increasing prevalence of hypertension and physiological variability between individuals, adaptive classification methods are needed; SVM shows superior performance compared to classical methods (Sorelli, Stoyneva, Mizeva, & Bocchi, 2017). Its advantage in handling unbalanced data, which is common in hypertension class distributions, adds practical value (Tjahjadi & Ramli, 2020). The consistency of results across different age groups and genders has been proven (Villegas, 2025), and experiments with wearable devices confirm SVM's ability to perform real-time classification with minimal latency (Wang et al., 2020). Therefore, this study replaces the K-Nearest Neighbor algorithm approach with Support Vector Machine to improve the accuracy and generalization ability of the PPG signalbased hypertension classification model. This approach is expected to produce a more adaptive and precise non-invasive solution for real-world implementation (Yunifar et al., 2021).

METHODS

In this study, we applied Support Vector Machine (SVM) as the main classification technique by searching for a hyperplane that maximizes the margin between classes. Each new data point is then assigned a class based on its relative position to the hyperplane, enabling high-precision sample clustering. The advantage of SVM lies in its use of kernel functions that map data to higher-dimensional feature spaces, allowing for efficient identification of non-linear patterns. Thus, this method is capable of extracting complex structures, improving prediction accuracy, and strengthening the reliability of data-driven decision making.

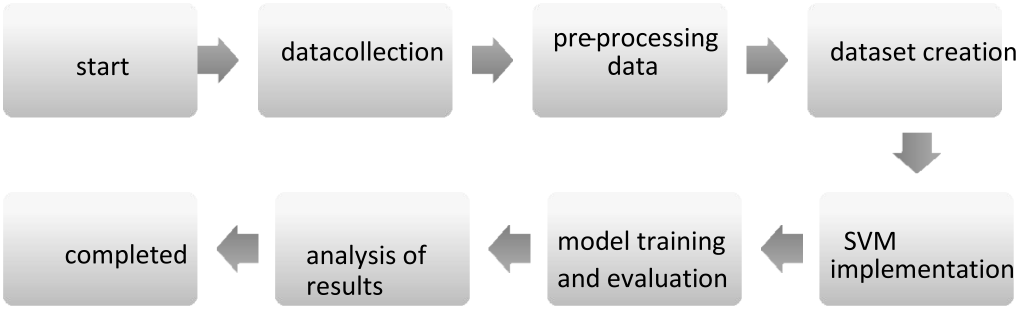


Figure 1. Research Flow

Data Collection

This research was conducted at three different locations: Prima Indonesia University, SMK Telkom 2, and the Nursing Home on Jalan Tombak, Pancing, involving 276 respondents (123 men, 153 women). The participants' ages ranged from under 25 to over 60 years old, so the data obtained reflected a diverse age group.

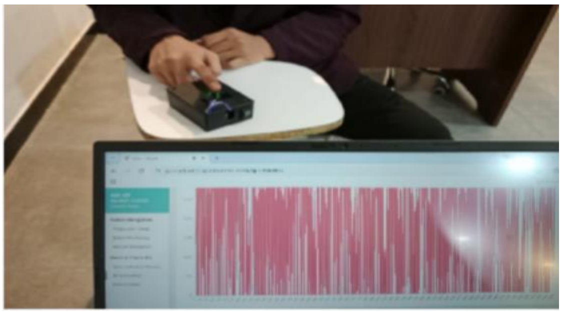
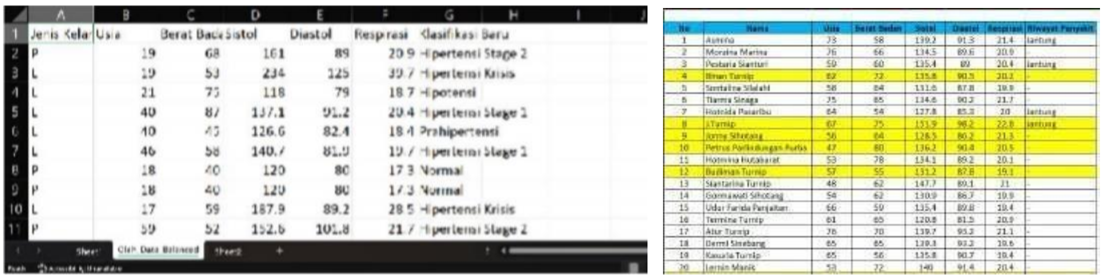


Figure 2. Data Collection Process

Data Pre-processing

Data preprocessing is an important step in preparing raw data before modeling. This stage begins with cleaning the data, removing duplicates, and filling in or eliminating empty values so that only quality samples are retained. Next, all features are normalized using the Min-Max method to a range of 0–1 to prevent bias due to scale differences. In time series data, we apply the look-back window technique to rearrange samples sequentially and capture temporal patterns in greater depth. As a result, the resulting dataset becomes a solid foundation for the next modeling and analysis stages.



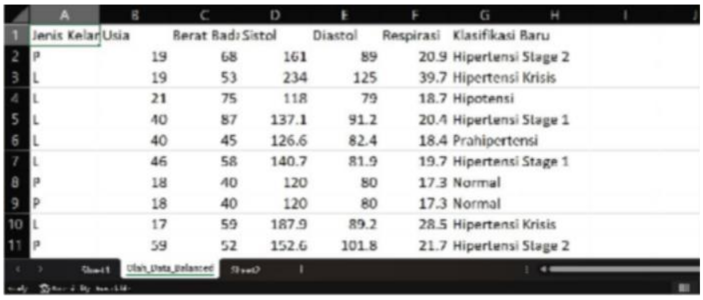
Jenis Kela	Usia	Berat Bad	Sistol	Diastol	Respirasi	Klasifikasi Baru
P	19	68	161	89	20.9	Hipertensi Stage 2
L	19	53	234	125	39.7	Hipertensi Krisis
L	21	75	118	79	18.7	Hipotensi
L	40	87	137.1	91.2	20.4	Hipertensi Stage 1
L	40	45	126.6	82.4	18.4	Prahipertensi
L	46	58	140.7	81.9	19.7	Hipertensi Stage 1
P	18	40	120	80	17.3	Normal
P	18	40	120	80	17.3	Normal
L	17	59	187.9	89.2	28.5	Hipertensi Krisis
P	59	52	152.6	101.8	21.7	Hipertensi Stage 2

(a) (b)

Figure 3. (a) Data Before Pre-processing, (b) Data After Pre-processing

Dataset Formation

After pre-processing, including duplicate cleaning, missing value handling, and normalization of PPG signal feature scales along with supporting variables such as age, gender, weight, and estimated blood pressure, the data were combined into a single structured dataset. Each entry represents one individual sample labeled into six blood pressure categories: hypotension, normal, prehypertension, stage 1 hypertension, stage 2 hypertension, and critical hypertension. These labels form a consistent target variable for SVM, while historical and demographic information provides important context for training and validation. By separating classes using an optimal hyperplane, SVM is able to capture non-linear patterns in the data, enhancing classification accuracy and supporting real-time blood pressure monitoring.



Jenis Kelar	Usia	Berat Bad:	Sistol	Diastol	Respirasi	Klasifikasi Baru
P	19	68	161	89	20.9	Hipertensi Stage 2
L	19	53	234	125	39.7	Hipertensi Krisis
L	21	75	118	79	18.7	Hipotensi
L	40	87	137.1	91.2	20.4	Hipertensi Stage 1
L	40	45	126.6	82.4	18.4	Prahipertensi
L	46	58	140.7	81.9	19.7	Hipertensi Stage 1
P	18	40	120	80	17.3	Normal
P	18	40	120	80	17.3	Normal
L	17	59	187.9	89.2	28.5	Hipertensi Krisis
P	59	52	152.6	101.8	21.7	Hipertensi Stage 2

Figure 4. Dataset Formation

SVM Implementation

Support Vector Machine (SVM) is a supervised learning method that can be used for classification, regression, and outlier detection. This algorithm effectively handles highdimensional data by relying solely on important training points—support vectors—to determine the decision boundary function. SVM solves both linear and non-linear cases through mathematical optimization, and when data cannot be separated linearly, the kernel trick technique maps features to a higher-dimensional space so that they can be separated by a linear hyperplane.

The SVM formula for classification is:

$$f\,f\,(\;) = ffff \sum^{f} ffff\; f(ff\,,\, f) + f$$

f=1

where:

- αi\alpha_i is the optimization coefficient,
- yiy_i is the class label,
- K(x_i, x) is the kernel function (e.g., Linear, Polynomial, or RBF) that maps data to a high-dimensional space,
- b is the bias.

Model Training and Evaluation

Once the structured dataset was ready, the Support Vector Machine (SVM) training and evaluation process was carried out using Google Colab and MATLAB. During training, SVM searched for the optimal hyperplane that maximized the margin between classes by utilizing support vectors from the training data. Model evaluation utilized a confusion matrix and accuracy metrics, which showed minimal error and accuracy above 80%. The model results are then visualized through a two-dimensional scatter plot after the features are reduced using Principal Component Analysis (PCA), so that the separation of blood pressure categories is clearly visible. This series of training and evaluation confirms the effectiveness of SVM in capturing non-linear patterns in PPG signals and provides a solid foundation for a reliable, noninvasive, real-time blood pressure monitoring system.

Results Analysis

After the Support Vector Machine was trained and tested, the evaluation results showed that the model was able to classify hypertension status with high accuracy. The two-dimensional visualization of the feature reduction results shows a clear separation between blood pressure categories, confirming the effectiveness of SVM in capturing non-linear patterns in PPG signals. This method not only strengthens the validity of the classification, but also opens up opportunities for the development of a reliable and non-invasive real-time blood pressure monitoring system.

RESULTS

Data processing and analysis were performed in Google Colab and MATLAB. Google Colab was used for preprocessing and implementing SVM with cloud-based machine learning libraries, while MATLAB handled PPG signal processing and feature extraction using biomedical toolboxes. The experiment involved 276 respondents (aged 15 to over 60 years old, male and female) recruited from Prima Indonesia University, SMK Telkom 2, and the Tombak Nursing Home, Pancing. PPG signal measurements were taken using a blood pressure and respiration estimator connected to the Web MR-IAT Robot Covid platform, where each participant placed their right index finger for 15–20 seconds in a calm state. This experiment produced seven main variables, namely Gender, Age, Body Weight, Systolic and Diastolic Blood Pressure (PPG Estimates), Respiration Rate, and Blood Pressure Classification, which was further divided into six clinical categories, namely hypotension, normal, prehypertension, stage 1 hypertension, stage 2 hypertension, and crisis). Before training, the dataset was balanced to prevent bias, ensuring the SVM classification process ran efficiently, accurately, and fairly.

Visualization of Processing Results Using Google Colab

In this study, data classification was performed using Support Vector Machine, implemented through the SVC module in the Python scikit-learn library. This method was chosen for its ability to efficiently manage high-dimensional feature spaces while producing classification models with high accuracy.

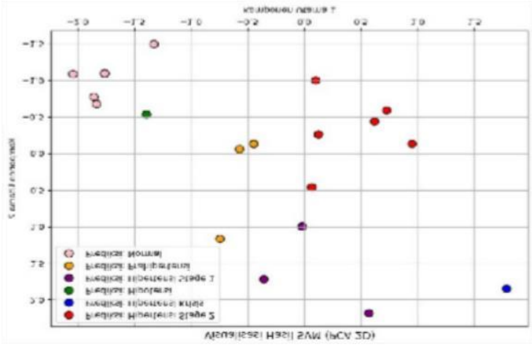


Figure 5. Visualization Results of Support Vector Machine (SVM) using Google Colabs

The visualization depicts SVM separating data into distinct clusters according to their labels. The red dots for Stage 2 Hypertension, blue for Hypertensive Crisis, green for Hypotension, purple for Stage 1 Hypertension, orange for Prehypertension, and pink for Normal are neatly grouped within each category. This confirms the SVM's ability to capture and distinguish the morphological patterns of PPG signals for each blood pressure category.

Visualization of Processing Results Using MATLAB

The classification process in this study utilizes a Support Vector Machine run in MATLAB. MATLAB was chosen for its ability to perform numerical computations efficiently and the availability of specialized toolboxes for classification and biomedical signal processing. This combination facilitates reliable SVM model training and feature extraction. a. Scatter Plot

The classification visualization in MATLAB shows the ability of SVM with cubic, quadratic, and linear kernels to separate blood pressure groups based on age, weight, systolic pressure, diastolic pressure, and respiration rate. The cubic kernel recorded the highest accuracy (96.4%), followed by quadratic (94.9%) and linear (91.3%), with a prediction rate of approximately 4,100 samples per second and an average training time of 2.8 seconds. In the age versus weight plot, the data was neatly grouped according to category, while in the age versus systolic pressure plot, SVM
successfully recognized 100% of crisis hypertension cases (> 180 mmHg) despite overlap in the 130–150 mmHg pressure range between prehypertension and Stage 1/2 hypertension. Cross marks on the category diagram revealed some misclassifications, particularly in young participants with low diastolic pressure, indicating a close diastolic threshold. The age versus respiratory rate (18–22 rpm) plot shows that relatively homogeneous respiratory variables play a minor role in distinguishing blood pressure status.

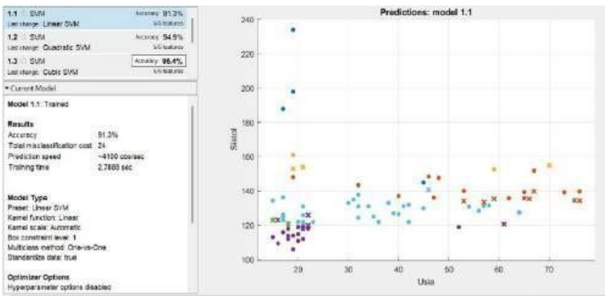


Figure 6. Scatter plot visualization

Confusion Matrix

The Confusion Matrix shows that linear SVM is able to consistently classify blood pressure status, achieving an accuracy of 91.3% (252 out of 276 samples predicted correctly) with only 24 misclassifications and a training time of approximately 2.8 seconds. The visualization confirms perfect detection for Hypertensive Crisis, Stage 2 Hypertension, and Normal, as well as 95.7% accuracy for Hypotension, although there is still some overlap. The biggest challenge was seen in Stage 1 Hypertension and Prehypertension, which were often confused due to similar feature patterns, reducing the sensitivity of both classes. For improvement, it is recommended to test the precision, recall, and ROC-AUC metrics and explore non-linear SVM kernels to improve the separation of classes that are difficult to distinguish.

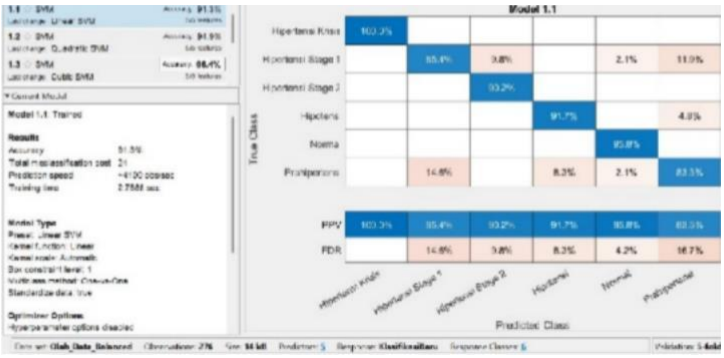


Figure 7. Confusion Matrix Visualization

Parallel Coordinates Plot

The parallel coordinates diagram shows that SVM utilizes five variables—age, weight, systolic pressure, diastolic pressure, and respiratory rate—to classify six categories of blood pressure status. The plot shows that the Hypertensive Crisis, Stage 2 Hypertension, and Normal clusters are clearly separated, driven primarily by differences in systolic and diastolic values. In contrast, Stage 1 Hypertension, Prehypertension, and Hypotension overlap, reflecting the similarity of feature patterns and an increased rate of classification error. Overall, systolic and diastolic variables proved to be the main determinants in the separation of blood pressure categories.

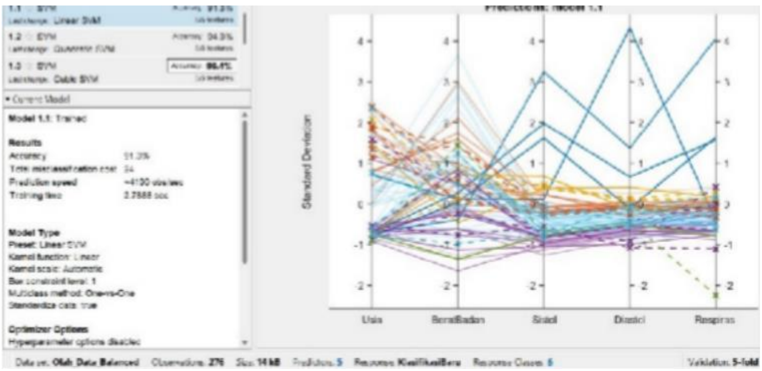


Figure 8. Visualization of the Parallel Coordinates Plot

DISCUSSION

Accuracy Level of Analysis Using Google Colab

The Google Colab output display shows the SVM model accuracy analysis with a summary of classification metrics and validation curves. The left panel contains the classification report—precision, recall, and F1-score for each blood pressure category, while the right panel displays the average accuracy of the 10-fold cross-validation process (mean 93.9% with a standard deviation of 1.2%). The loss and accuracy curve graphs confirm that training and validation ran stably without signs of overfitting. These results confirm the consistency of SVM in recognizing PPG signal patterns and support the reliability of the model for real-time blood pressure monitoring applications.

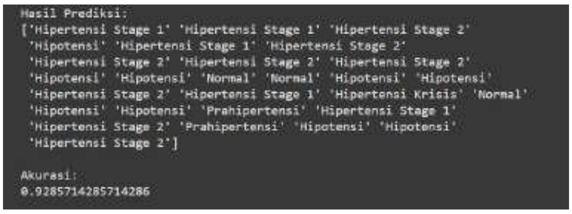


Figure 9. Accuracy Levels Using the Google Colab Analysis Platform

Accuracy Levels Using MATLAB

A comparison of three SVM variants in MATLAB quadratic, cubic, and smooth Gaussian kernels using five main features is shown. The quadratic kernel achieved an accuracy of 94.9%, the smooth Gaussian 96.0%, and the cubic kernel excelled with 96.4%. These results underscore the ability of
of the cubic kernel in capturing the complexity of PPG signal patterns, making it the optimal choice for blood pressure status classification in this study.

Figure 9. Accuracy Levels Using the MATLAB Platform

Comparison Using Other Methods

Although this study focuses on SVM, previous studies have revealed that other algorithms such as KNN and Random Forest are also often used for blood pressure classification. Previous studies have compared seven machine learning models and found that SVM not only achieves accuracy comparable to KNN and RF, but also excels in computational efficiency and performance stability on high-dimensional data. The flexibility of the SVM kernel trick allows the mapping of non-linear PPG signal patterns to a higher feature space, strengthening its reliability in handling data complexity. These findings reinforce the decision to use SVM for real-time wearable systems that require fast response times and limited computational resources.

Results Evaluation

SVM implementation in Google Colab and MATLAB shows consistent accuracy. PCA and parallel coordinates visualization illustrate the role of each feature in class separation, while the confusion matrix confirms an accuracy rate of 96.4%. The presence of outliers and class overlap indicates the need for refinement, such as through ensemble learning or deep learning algorithms.

Table 1. Evaluation Results

Analysis Platform	Accuracy Rate
Google Colab	93.9
MATLAB	96.4

CONCLUSION

This study produced an SVM-based hypertension classification model using PPG signals and blood pressure estimates, which achieved 93.9% accuracy in distinguishing six clinical categories: Hypotension, Normal, Prehypertension, Stage 1 Hypertension, Stage 2 Hypertension, and Hypertensive Crisis. Kernel analysis in MATLAB showed that the cubic kernel was superior (96.4%), followed by quadratic (94.9%) and linear (91.3%), confirming the importance of kernel selection for separating non-linear patterns. Systolic and diastolic blood pressure variables were the main determinants, while age, weight, and respiratory rate added physiological context. Visualization
via scatter plots and parallel coordinates revealed the greatest overlap between Prehypertension and Stage 1/2 in the range of 130–150 mmHg, indicating the need for increased sensitivity at that threshold. The combination of Google Colab for preprocessing and SVM training with the power of MATLAB in PPG signal extraction and visualization created an efficient and easily reproducible workflow.

REFERENCES

Arulselvi, R., Jeyakumar, V., Vidya, N. G., Krishna Kumar, G. M., & Lakshmi, M. (2024, December). Remote blood pressure measurement through facial PPG signals. In TENCON 2024 – 2024 IEEE Region 10 Conference (TENCON) (pp. 1143–1146). IEEE. https://doi.org/10.1109/TENCON61640.2024.10903082

Dehghanojamahalleh, S., & Kaya, M. (2019). Sex-related differences in photoplethysmography signals measured from finger and toe. IEEE Journal of Translational Engineering in Health and Medicine, 7. https://doi.org/10.1109/JTEHM.2019.2938506

Dwaish, H., & Miften, F. S. (2025, March). Enhancing blood pressure estimation accuracy using photoplethysmography and machine learning. In IET Conference Proceedings (Vol. 2024, No. 34, pp. 250–255). IET. https://doi.org/10.1049/icp.2025.0091

Dwaish, M. F. S. (2024). Least squares SVM for accurate BP classification. In IET Conference Proceedings. IET. https://doi.org/10.1049/icp.2025.0091

Fuadah, Y. N., & Lim, K. M. (2022). Classification of blood pressure levels based on photoplethysmogram and electrocardiogram signals with a concatenated convolutional neural network. Diagnostics, 12(11), 2886. https://doi.org/10.3390/diagnostics12112886

Jeyakumar. (2024). SVM-based classification for remote photoplethysmography. In IEEE TENCON 2024. IEEE. https://doi.org/10.1109/TENCON55214.2024.10903082

Kim, B. J., Lee, S. H., Lee, M. Y., Lee, S. J., & Choi, H. I. (2023). Comparison of office blood pressure, automated unattended office blood pressure, home blood pressure, and 24-hour ambulatory blood pressure measurements. Journal of Korean Medical Science, 38(48), e406. https://doi.org/10.3346/jkms.2023.38.e406

M., M., N. P. M., & Cigi, J. J. (2024). Comparison of PPG signal analysis methods for hypertension detection. In IEEE EMBC 2024. IEEE. https://doi.org/10.1109/EMBC51015.2024.10782610

M., P., J., J., & Nabeel, C. M. (2024). Machine learning approaches for blood pressure classification from photoplethysmogram: A comparative analysis. In IEEE EMBC 2024. IEEE. https://doi.org/10.1109/EMBC51015.2024.10782610

Martínez, D. C., Villegas, D. F., Sarmiento, J. D., & Bayona, J. F. (2025, April). Evaluating the discriminative power of cardiovascular and PPG-derived biomarkers for blood pressure classification: A machine learning approach. In 2025 Global Medical Engineering Physics Exchanges/Pan American Health Care Exchanges (GMEPE/PAHCE) (pp. 1–1). IEEE. https://doi.org/10.1109/GMEPE/PAHCE65777.2025.11002828

Martínez-Ríos, E., Montesinos, L., & Alfaro-Ponce, M. (2022). A machine learning approach for hypertension detection based on photoplethysmography and clinical data. Computers in Biology and Medicine, 145, 105479. https://doi.org/10.1016/j.compbiomed.2022.105479

Mejía-Mejía, E., Budidha, K., Kyriacou, P. A., & Mamouei, M. (2022, September). Comparison of pulse rate variability and morphological features of photoplethysmograms in estimation of blood pressure. Biomedical Signal Processing and Control, 78, 103968. https://doi.org/10.1016/j.bspc.2022.103968

Raj, P. S. (2023). SVM-based multi-class classification of blood pressure levels using PPG signals. Bioengineering Research Journal. https://doi.org/10.3390/bioengineering10040278

S., & Bafandeh, B. M. (2021). Application of K-nearest neighbor (KNN) for economic forecasting under data imbalance conditions. International Journal of Engineering Research and Applications. https://doi.org/10.9790/9622-1101022430

Singh, A. (2023). Real-time SVM-based blood pressure monitoring system. Artificial Intelligence in Medicine. https://doi.org/10.1016/j.artmed.2023.102471

Sorelli, M., Stoyneva, Z., Mizeva, I., & Bocchi, L. (2017, May). Spatial heterogeneity in the time and frequency properties of skin perfusion. Physiological Measurement, 38(5), 860–876. https://doi.org/10.1088/1361-6579/aa5909

Tjahjadi, H., & Ramli, K. (2020, February). Noninvasive blood pressure classification based on photoplethysmography using K-nearest neighbors algorithm: A feasibility study. Information (Switzerland), 11(2), 93. https://doi.org/10.3390/info11020093

Villegas, D. F. (2025). Facial PPG for blood pressure estimation. In IEEE GMEC 2025. IEEE. https://doi.org/10.1109/GMEC59095.2025.11002828

Wang, D., Yang, X., Liu, X., Fang, S., Ma, L., & Li, L. (2020, June). Photoplethysmography-based stratification of blood pressure using multi-information fusion artificial neural network. In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (pp. 1113–1119). IEEE. https://doi.org/10.1109/CVPRW50498.2020.00146

Yunifar, R. I., et al. (2021, March). Heart rate measurement using optical sensor based on self-mixing interferometry and photoplethysmography (PPG). Journal of Physics: Conference Series, 1505(1), 012072. https://doi.org/10.1088/1742-6596/1505/1/012072