

Hypertension Screening via Awake-Sleep Differences in Photoplethysmogram Signals

Jingyuan Hong¹, Manasi Nandi², Yali Zheng³, Jordi Alastruey¹

¹ School of Imaging Sciences and Biomedical Engineering, King's College London, London,
United Kingdom

² School of Cancer and Pharmaceutical Science, King's College London, London, United
Kingdom

³ College of Health Science and Environmental Engineering, Shenzhen Technology
University, Shenzhen, China

Running title: Ambulatory PPG for Hypertension Screening

Correspondence:

Dr Jordi Alastruey, School of Imaging Sciences and Biomedical Engineering, King's College London,
St. Thomas' Hospital, London SE1 7EU, United Kingdom
Tel: + 44 020 7848 9556, E-mail: jordi.alastruey-arimon@kcl.ac.uk

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

18 **Novelty and Relevance**

19 **What Is New?**

- 20 • This study introduces a novel machine learning–based hypertension screening method using
- 21 awake-sleep differences in ambulatory photoplethysmogram (PPG) signals, validated on both
- 22 internal and external datasets with a novel p-value–based feature selection strategy.

23 **What Is Relevant?**

- 24 • The proposed approach enables non-invasive, cuff-free hypertension detection using wearable
- 25 PPG data, offering improved generalisability and real-world applicability over traditional
- 26 models.

27 **Clinical/Pathophysiological Implications?**

- 28 • The ability to detect altered nocturnal cardiovascular patterns from PPG may support early
- 29 identification of unmanaged hypertension and promote more accessible long-term monitoring.

30

Abstract

Background: Hypertension is a major risk factor for cardiovascular diseases. This study proposes a novel hypertension screening framework based on awake-sleep differences in photoplethysmography (PPG) indices, using machine learning. We hypothesised that normotensive individuals exhibit greater PPG variation between awake and sleep states than unmanaged hypertensive individuals.

Methods: The Aurora-BP dataset (n=180; 138 normotensive, 42 hypertensive) was used for model development, with 18 subjects reserved for internal testing. External validation was performed using the independent CUHK-BP dataset (n=26; 11 normotensive, 15 hypertensive). Twenty PPG-based indices were extracted, and subject-level p-values from Mann-Whitney U-tests comparing awake and sleep periods were used as model features. Discretised p-values served as inputs for four machine learning models.

Results: The Support Vector Machine (SVM) achieved the highest performance, with $81.1 \pm 8.4\%$ accuracy and $82.8 \pm 8.1\%$ F1-score on the internal test set using all indices. On the external test set, the SVM using only temporal indices achieved 84.6% accuracy and 86.7% F1-score. Temporal indices, especially those linked to the dicrotic notch, showed strong generalisability across datasets.

Conclusion: The study demonstrates the feasibility of awake-sleep PPG analysis for hypertension screening, highlighting the potential of wearable PPG devices for ambulatory monitoring.

Key words: hypertension, photoplethysmogram, ambulatory monitoring, machine learning

Introduction

Hypertension is a leading modifiable risk factor for cardiovascular mortality worldwide [1]. The standard method for diagnosing hypertension relies on absolute blood pressure (BP) values measured using an upper-arm cuff device, which operates via an inflatable cuff and an oscillometric algorithm to detect brachial BP readings [2]. Diagnosis strategies range from repeated office-visit readings to out-of-office monitoring and ambulatory measurements, depending on clinical guidelines and physician recommendations [2], [3]. Regular BP monitoring in daily life is also recommended for early detection and effective management of hypertension [2], [4]. However, current self-monitoring techniques, primarily cuff-based, can be uncomfortable and difficult to maintain over time, especially for frequent or ambulatory measurements [5], [6]. This highlights the need for a precise, user-friendly, and unobtrusive method for at-home hypertension screening and management.

Photoplethysmography (PPG), which measures peripheral blood volume changes during the cardiac cycle, is widely used in wearable devices [7]. PPG signals have been extensively studied as indicators of cardiovascular health, with applications ranging from atrial fibrillation detection to real-time BP estimation [8], [9]. Previous studies have used machine learning algorithms to explore the relationship between cardiovascular function and various PPG-based indices, including morphological, temporal, and derivative features [10], [11]. Despite its potential, PPG faces several challenges: the signal is sensitive to motion artefacts and contact pressure, and its stability can be affected by variations in amplitude and waveform shape across devices, populations, and environmental conditions [12], [13], [14]. As a result, many previous studies have been limited to single datasets, restricting the generalisability of their findings across diverse populations and settings [9], [15], [16], [17].

Ambulatory BP monitoring has been recognised in recent clinical guidelines as a more accurate method than office-based measurement. It provides comprehensive readings, reduces the risk of white coat hypertension, and minimally disrupts daily activities [2]. Crucially, it enables the detection of nocturnal BP patterns, which are key indicators of cardiovascular risk. Normotensive individuals typically exhibit a nocturnal ‘sleep-dip’ in BP, while hypertensive individuals may show altered patterns such as non-dipping, extreme dipping, or reverse dipping [18]. Nighttime readings are often more reliable due to reduced motion and temperature fluctuations [19]. Similarly, ambulatory PPG monitoring has the potential to capture awake-sleep variations in cardiovascular function [20], [21]. Continuous 24-hour PPG monitoring could detect these variations using a single wearable device, reducing the impact of measurement errors and inconsistencies across subjects, devices, and measurement conditions.

While ambulatory BP monitoring is well-established in current clinical guidelines [2], [3], [22], there remains a gap in methods that leverage less obtrusive physiological signals (i.e., PPG) for robust hypertension screening. Previous studies have focused on short-term or single PPG measurements within isolated datasets, limiting their clinical reliability and applicability. Moreover, fewer studies have examined whether diurnal variations in PPG indices can serve as indicators of cardiovascular health.

This study introduces a novel machine learning-based method for hypertension screening using ambulatory PPG measurements. We hypothesised that normotensive individuals would exhibit greater awake-sleep differences in PPG features, while individuals with unmanaged hypertension would show diminished or absent differences. To test this, we used two independent datasets: the Aurora-BP dataset (n=180) for model development and internal validation, and the CUHK-BP dataset (n=26) for external testing to assess generalisability.

Methodology

Dataset Description

The Aurora-BP dataset is an open-access resource containing data from 1,125 participants, collected to investigate cuff-based brachial BP readings, including systolic BP (SBP) and diastolic BP (DBP), along with Lead-I electrocardiogram (ECG), wrist PPG, and radial tonometric pressure waveforms [24]. Data were collected in two settings: laboratory (642 participants) and ambulatory (483 participants). PPG signals were recorded from the left wrist using an optical sensor (MAX30101, Analog Devices), while brachial BP was measured on the right arm using an ambulatory BP monitor (OnTrak 90227, Spacelabs Healthcare) (Figure 1a). During ambulatory monitoring, BP and PPG data were collected twice per hour during the daytime (8 a.m. to 8 p.m.) and once per hour at night (8 p.m. to 8 a.m.), producing 15-second PPG segments paired with single SBP and DBP measurements (Figures 1b and 1c) [24].

From the ambulatory subset, 240 participants were included after preprocessing. This involved excluding segments with abnormal BP values ($\text{DBP} < 30 \text{ mmHg}$) [25] and poor-quality PPG signals (quality score < 0.8) [24], reducing the sample from 483 to 441. Participants with fewer than three valid daytime and three valid nighttime measurements were further excluded, yielding a final cohort of 240. These participants were classified as normotensive, managed hypertensive, or unmanaged hypertensive (Table 1). Hypertensive status was self-reported based on the 2017 AHA guidelines [22], and classification labels were prelabelled in the dataset. Only the normotensive and unmanaged hypertensive participants were used for model training, testing, and evaluation as the internal dataset.

The CUHK-BP dataset, collected by the Chinese University of Hong Kong and Prince of Wales Hospital [20], includes data from 26 participants who underwent 24-hour ambulatory monitoring using a commercial BP monitor (Oscar2, SunTech Medical) and a custom armband equipped with a near-infrared PPG sensor and ECG electrodes (two e-textile and one Ag-Ag/Cl). PPG signals were recorded from the left upper arm, and BP was measured on the right arm (Figure 1a). Data were collected twice per hour, yielding 60-second PPG segments with corresponding single SBP and DBP readings (Figure 1b and 1c). Demographic information is provided in Table 2. Participants were diagnosed as normotensive or unmanaged hypertensive according to the 2010 Hong Kong Reference Framework (HKRF) [26], with hospital-confirmed diagnoses. This dataset was used solely for external testing and evaluation of model generalisability.

Signal Processing

The 60-second PPG segments in the CUHK-BP dataset were trimmed to 25 seconds by extracting the middle portion to remove unstable waveforms and noise typically found at the segment boundaries. Raw PPG segments from both datasets were then filtered using a 4th-order high-pass Butterworth filter (cut-off: 0.25 Hz) to eliminate low-frequency noise and baseline wander, followed by an equiripple low-pass finite impulse filter with 10 Hz passband, 12 Hz stopband, 1 dB passband ripple, and 60 dB stopband attenuation to suppress high-frequency noise (Figure 2a). Fiducial points were extracted from each beat within the filtered PPG segments using the PulseAnalyse algorithm [27] (Figure 2b). From these fiducials, 20 indices were computed as described in Supplementary Table S1. The average value of each index across all beats in a segment was then calculated for subsequent analysis.

Awake and Sleep Data Selection

The Aurora-BP dataset included step counts from the 30 minutes preceding each PPG segment recording. Sleep onset was defined as the first occurrence of zero steps after 8 p.m. (indicating sustained inactivity), and wake time as the first non-zero step count thereafter. These criteria

delineated awake and sleep periods. In the CUHK-BP dataset, participant self-reported their sleep times, allowing direct separation of awake and sleep data.

Statistical Analysis

Awake and sleep measurements were compared at both the cohort and individual levels.

At the cohort level, paired t-tests were used to assess awake-sleep differences in SBP, DBP, mean BP (MBP = (SBP+2*DBP)/3) [28], and pulse pressure (PP=SBP-DBP) across the three Aurora-BP cohorts and the two CUHK-BP cohorts. The difference between awake and sleep periods for all 20 PPG indices were also analysed using paired t-tests to assess statistical significance. Kruskal–Wallis tests were used to compare awake- and sleep-period SBP, DBP, MBP, PP, and continuous demographic variables among the three Aurora-BP cohorts, while Mann–Whitney U-tests were used for the same comparisons between the two CUHK-BP cohorts. Categorical variables were compared using chi-squared tests.

At the individual level, circadian SBP patterns were calculated as the percentage reduction from mean awake to mean sleep SBP and classified as dipper, extreme dipper, non-dipper, and reverse dippers [29]. Awake-sleep differences in the 20 PPG indices were evaluated for normotensive and unmanaged hypertensive subjects in both datasets. Due to the unequal numbers of awake and sleep measurements, Mann-Whitney U-tests were used. The resulting p-values reflected each index's ability to distinguish between awake-sleep states while remaining robust to outlier values (particularly important given the noise sensitivity of PPG signals). To normalise and simplify machine learning model inputs, p-values were discretised into three categories: (i) highly statistically significant ($p < 0.01$) was set to 1, (ii) statistically significant ($0.01 \leq p < 0.05$) was set to 0.5, and (iii) not statistically significance ($p \geq 0.05$) was set to zero. These discretised set of 20 p-values were used as input features for machine learning models.

The frequency of significance labels was analysed across individuals within and between cohorts in the internal dataset to assess intra-cohort variability and inter-cohort separability.

Machine Learning Models

Four binary classifiers—Logistic Regression (LR), K-Nearest Neighbours (KNN), Support Vector Machines (SVM), and eXtreme Gradient Boosting (XGBoost)—were used to classify subjects as normotensive or hypertensive. These models have previously been applied in hypertension screening studies [30], [31], [32].

The same processing pipeline, illustrated in Figure 3 was applied to each model. A randomly selected subset of 18 subjects (9 normotensive, 9 unmanaged hypertensive; 10% of Aurora-BP) was held out as an independent internal test set. The remaining 162 normotensive and unmanaged hypertensive subjects were used for training and validation. To address class imbalance (129 normotensive vs. 33 hypertensive), the Synthetic Minority Oversampling TEchnique (SMOTE) [33] was applied using five nearest neighbours to synthetically balance the dataset (157 normotensive, 159 hypertensive). The internal test set and the external CUHK-BP dataset were used for model evaluation.

All models were trained and optimised on discretised p-values from 20 PPG indices (11 temporal, 9 morphological) using the hyperparameters listed in Supplementary Table S2. A 5-fold cross-validation was used during training. Model performance was evaluated using accuracy and F1-score on both internal and external test sets. Internal testing included five random splits for robustness. For external testing, the model was trained on the entire internal dataset.

The 20 PPG-based indices were grouped into three categories: complete, temporal, and morphological. The complete group included all 20 indices listed in Table S1. The temporal group comprised 11 indices derived from time-fiducial points of the PPG waveform, while the morphological group included the remaining 9 indices extracted from morphological fiducial points.

The best-performing model was further analysed using SHapley Additive exPlanations (SHAP) [34] to evaluate the contribution of each index to classification decisions.

Baseline Age-Only Model

Given that age is a well-known risk factor for hypertensive [35], a baseline model was developed for the best-performing binary classifier using age as the sole input feature. Due to the statistically significant age difference between normotensive and unmanaged hypertensive subjects in the CUHK-BP dataset (Table 2), this baseline model was trained and evaluated only on the internal Aurora-BP dataset to avoid confounding.

All signal processing, index extraction, and model development were implemented in MATLAB (version 2023a, MathWorks). A pseudocode representation of this process is provided in Figure S1.

Results

Awake and Sleep Differences

In the internal dataset, all three cohorts (normotensive, managed hypertensive, and unmanaged hypertensive) exhibited statistically significant reductions in cuff-based SBP, DBP, MBP and PP during sleep (Figure 4b). SBP, DBP and MBP followed similar circadian trends, reaching their lowest values around 2:30 a.m. (Figure 4a). Although PP changes were less visually pronounced (Figure 4a), they remained statistically significant (Figure 4b). The distribution of circadian SBP patterns (dipper, extreme-dipper, non-dipper, and reverse dipper) was similar across the three cohorts (Supplementary Table S3). However, comparisons across cohorts revealed significant differences in SBP, DBP, MBP and PP between the normotensive and unmanaged hypertensive groups during both awake and sleep periods (Supplementary Figure S3a).

In the external dataset, DBP and MBP also decreased significantly during sleep, while SBP showed a significant reduction only in the unmanaged hypertensive cohort (Supplementary Figure S3). PP showed no significant change. These findings align with SBP patterns distributions (Table S3), where the normotensive cohort showed a low proportion of dippers and reverse dippers (9.1% each) and a high proportion of non-dippers (81.8%), whereas the unmanaged hypertensive cohort had more dippers (33.3%) and reverse dippers (20.0%). All measures except for SBP and PP during sleep showed significant differences between normotensive and unmanaged hypertensive groups (Supplementary Figure S3b).

PPG waveform comparisons in Figure 5 further illustrate typical awake-sleep differences. In normotensive individuals, the awake-period waveform exhibited a clear diastolic notch and diastolic peak, which were less prominent during sleep and in hypertensive subjects. Multiple PPG-derived indices also showed significant awake-sleep differences, particularly in the normotensive group (Table 3). Three temporal indices (t_{sys} , t_{dia} , t_{p1-dia}) and three morphological indices (RI , $radio_{p2-p1}$, and $A2$) were highly significant in normotensives ($p < 0.01$), but not in hypertensives ($p \geq 0.05$). Four additional temporal indices (Δt , SI , $prop_s$, t_{radio}) and two morphological indices (RI_{p2} and $ratio_{p2-p1}$) show moderate significance in normotensives ($0.01 \leq p < 0.05$), but again were not significant in hypertensives. The managed hypertensive cohort showed a p-value distribution more closely aligned with the normotensive group, differing in only five indices.

Although most indices were not statistically significant across all subjects (Figure 6), normotensive individuals showed a higher proportion of significant indices than unmanaged hypertensives (8.3% vs. 3.2%). Notably, several indices (CT/h , t_{sys} , t_{dia} , t_{p1-dia} , AI , AP , RI_{p1} , $ratio_{p2-p1}$ and $A1$) were highly significant ($p < 0.01$) in over 10% of normotensive subjects.

Hypertension Screening Performance

The SVM model achieved the best performance on both datasets, although with different sets of indices (Table 4). On the internal dataset, it attained the highest accuracy ($81.1\% \pm 8.4\%$) and F1-score ($82.8\% \pm 8.1\%$) using all 20 indices. On the external dataset, it reached an accuracy of 84.6% and an F1-score of 86.7% using only temporal indices. LR and XGBoost also performed well externally with temporal indices, each achieving accuracy and F1-scores above 70% and 80%, respectively. In contrast, KNN consistently showed the weakest performance, with internal scores below 63% and external scores below 45%. Overall, the complete index set was most effective internally, while temporal indices yielded better generalisability externally.

In the best-performing SVM model, internal testing showed a higher true negative rate (0.91) than true positive rate (0.71), with a false negative rate of 0.29 and a false positive rate of 0.09. Performance on the external dataset was more balanced across all classification metrics (Figure 7).

SHAP analysis identified four morphological and six temporal indices among the top 10 most important features for hypertension classification in the internal test using all 20 indices (Figure 8a). In the external test, the most important indices were t_{dia} , t_{p1-dia} , t_{p2-dia} , t_{sys} , and t_{ratio} ; all related to the dicrotic notch (Figure 8b).

The baseline SVM model using age alone achieved $67.8\% \pm 12.7\%$ accuracy and a $70.5\% \pm 10.7\%$ F1-score on the internal dataset, indicating that age was a moderate predictor but less effective than PPG-based indices.

Compared to previous studies, our p-value-based SVM approach achieved competitive accuracy and F1-scores on both internal and external datasets (Table 5), outperforming methods in [36] and [30], and approaching those in [31], [36] and [37], which lacked external validation.

Discussion

This study demonstrated the feasibility of hypertension screening using ambulatory PPG measurements through a novel machine learning framework that leverages awake-sleep differences in PPG morphology. Analysis of two independent datasets revealed distinct circadian variations in PPG-derived indices between normotensive and unmanaged hypertensive individuals. The support vector machine model achieved the highest classification performance, with approximately 85% accuracy and F1-score on an external dataset unseen during training. Temporal indices proved particularly robust across datasets, highlighting their generalisability. Our novel p-value-based approach matched or outperformed previous methods and, notably, included external validation; highlighting its potential for non-invasive, continuous hypertension screening in ambulatory settings.

Previous studies have shown that normotensive individuals typically exhibit a nocturnal BP dip, while hypertensive individuals may display altered patterns [18]. Our internal data confirmed this dip in normotensive participants but showed similar circadian trends, with no substantially altered patterns, in both hypertensive groups. Despite comparable BP patterns across cohorts, 13 of the 20 PPG indices exhibited highly significant awake-sleep differences in the normotensive cohort, compared to only

four in the unmanaged hypertensive group. This suggests that PPG captures distinct physiological responses to sleep not reflected by BP alone. The managed hypertensive cohort showed PPG patterns more closely aligned with the normotensive group, indicating that treatment may partially restore normal cardiovascular function [39]. Notably, although BP-based sleep-awake differences were less pronounced in the external dataset, PPG-based indices still effectively distinguished normotensive from unmanaged hypertensive individuals, underscoring the sensitivity of our novel p-value-based approach to subtle sleep-related cardiovascular changes.

At the individual level, both normotensive and hypertensive groups exhibited many non-significant awake-sleep differences. However, normotensives had a higher frequency of highly significant changes, reflecting greater physiological variability and supporting the use of machine learning to leverage these differences for classification.

Temporal indices yielded the strongest external performance, suggesting they capture hypertension-related features that are more generalisable than morphological indices, which may be more dataset-specific. These findings are consistent with previous work demonstrating the robustness of temporal features across PPG sensors and anatomical sites [11].

Unlike other studies that included demographic variables such as age [40], we excluded them to avoid confounding, as normotensive and hypertensive cohorts in the external dataset differed significantly in age. This allowed for a focused assessment of PPG-derived features. While age was a strong predictor of hypertension in our internal dataset, our PPG-based model substantially outperformed an age-only baseline.

Among the tested models, SVM consistently outperformed others across datasets. KNN, while moderately effective internally, performed poorly externally, likely due to its reliance on distance-based similarity metrics, which are sensitive to data distribution shifts. In contrast, SVM, XGBoost and LR, which learn underlying patterns, maintained stronger cross-dataset performance. SVM's advantage may stem from its focus on critical samples near the decision boundary rather than fitting the entire dataset to find optimal decision boundaries [41], making it more robust in small, imbalanced datasets. The internal dataset's imbalance led SVM to favour normotensive predictions during internal testing, while SVM applied to the external dataset using temporal indices produced more balanced classification results.

SHAP analysis identified the most influential features for classification. In the internal dataset, RI_{p1} and $A2$ (both morphological indices) ranked highest, consistent with their known association with arterial stiffness and vascular ageing; factors implicated in disrupted diurnal BP variation in hypertension [42]. These differences in diurnal index variation reflect underlying vascular pathophysiology. Temporal indices were more influential in the external dataset. A previous ICU-based study using 24-hour measurements found that the circadian rhythm of t_{sys} , t_{ratio} , and CT correlated more with SBP than heart rate, while Δt showed no such correlation [43]. Our findings aligned with these results, except for CT which showed moderate importance across both groups. Notably, the top five ranked indices (t_{dia} , t_{p1-dia} , t_{p2-dia} , t_{sys} , and t_{ratio}) are all related to the dicrotic notch of the PPG waveform, which marks aortic valve closure in the cardiac cycle. Sleep typically lowers heart rate and delays valve closure in normotensives, a pattern often disrupted in hypertension [44], [45], [46]. These disruptions may explain the higher importance of dicrotic-notch-related indices in our SHAP analysis and their strong generalisability.

This study has several limitations. First, both datasets were relatively small. Despite this, our method achieved performance comparable to or better than previous work using larger datasets, many of which lacked external validation. Second, we focused on binary classification, whereas clinical guidelines define multiple hypertension stages [2], [3], [22]. Third, we excluded managed hypertensive subjects from classification due to their absence in the external dataset. Future research

should consider multiple hypertension stages and expand to larger, more diverse ambulatory datasets to further enhance the generalisability, reliability and clinical relevance of this approach.

Conclusion

This study introduced a novel, machine learning-based hypertension screening method that leverages awake-sleep differences in PPG indices. This approach effectively distinguished normotensive from hypertensive individuals and demonstrated strong generalisability on an independent dataset. Temporal indices, particularly those related to the dicrotic notch, proved more robust across datasets than morphological indices. By using a p-value-based approach, our method overcomes common challenges such as PPG signal instability and noise sensitivity, thereby expanding the clinical utility of PPG. To our knowledge, this is the first study to exploit awake-sleep PPG variations for hypertension assessment, highlighting the potential of wearable devices for non-invasive, ambulatory monitoring in real-world settings.

Sources of Funding

This work was supported by the British Heart Foundation [PG/17/50/32903] and the Engineering and Physical Sciences Research Council Doctoral Training Partnership Grant [EP/T517963/1].

Disclosures

None

References

- [1] K. T. Mills, A. Stefanescu, and J. He, 'The global epidemiology of hypertension', *Nat. Rev. Nephrol.*, vol. 16, no. 4, pp. 223–237, Apr. 2020, doi: 10.1038/s41581-019-0244-2.
- [2] J. W. McEvoy *et al.*, '2024 ESC Guidelines for the management of elevated blood pressure and hypertension: Developed by the task force on the management of elevated blood pressure and hypertension of the European Society of Cardiology (ESC) and endorsed by the European Society of Endocrinology (ESE) and the European Stroke Organisation (ESO)', *Eur. Heart J.*, vol. 45, no. 38, pp. 3912–4018, Oct. 2024, doi: 10.1093/eurheartj/ehae178.
- [3] NICE guideline, 'Hypertension in adults: diagnosis and management (NG136)', *Natl. Inst. Health Care Excell.*, Aug. 2019, [Online]. Available: <https://www.nice.org.uk/guidance/ng136>
- [4] M.-Y. Rhee, M. Munakata, D.-Y. Nah, J. S. Kim, and H.-Y. Kim, 'Home blood pressure measurement for hypertension management in the real world: Do not just measure, but share with your physician', *Front. Cardiovasc. Med.*, vol. 10, Jan. 2023, doi: 10.3389/fcvm.2023.1103216.
- [5] M. Proença *et al.*, 'Method-comparison study between a watch-like sensor and a cuff-based device for 24-h ambulatory blood pressure monitoring', *Sci. Rep.*, vol. 13, no. 1, p. 6149, Apr. 2023, doi: 10.1038/s41598-023-33205-z.
- [6] K. L. Tucker *et al.*, 'Self-monitoring of blood pressure in hypertension: A systematic review and individual patient data meta-analysis', *PLOS Med.*, vol. 14, no. 9, p. e1002389, Sep. 2017, doi: 10.1371/journal.pmed.1002389.
- [7] P. H. Charlton, P. A. Kyriaco, J. Mant, V. Marozas, P. Chowienzyk, and J. Alastruey, 'Wearable photoplethysmography for cardiovascular monitoring', *Proc. IEEE*, vol. 110, no. 3, pp. 355–381, Mar. 2022, doi: 10.1109/JPROC.2022.3149785.
- [8] S. Gill *et al.*, 'Smartphone detection of atrial fibrillation using photoplethysmography: a systematic review and meta-analysis', *Heart*, vol. 108, no. 20, pp. 1600–1607, Oct. 2022, doi: 10.1136/heartjnl-2021-320417.
- [9] Y. Huang, Y. He, Z. Song, K. Gao, and Y. Zheng, 'Validation of deep learning models for cuffless blood pressure estimation on a large benchmarking dataset', *Connect. Health Telemed.*, vol. 3, no. 1, Mar. 2024, doi: 10.20517/chatmed.2023.23.
- [10] E. Finnegan, S. Davidson, M. Harford, P. Watkinson, L. Tarassenko, and M. Villarroel, 'Features from the photoplethysmogram and the electrocardiogram for estimating changes in blood pressure', *Sci. Rep.*, vol. 13, no. 1, p. 986, Jan. 2023, doi: 10.1038/s41598-022-27170-2.
- [11] D. Djeldjli, F. Bousefsaf, C. Maaoui, F. Bereksi-Reguig, and A. Pruski, 'Remote estimation of pulse wave features related to arterial stiffness and blood pressure using a camera', *Biomed. Signal Process. Control*, vol. 64, p. 102242, Feb. 2021, doi: 10.1016/j.bspc.2020.102242.
- [12] P. H. Charlton *et al.*, 'The 2023 wearable photoplethysmography roadmap', *Physiol. Meas.*, vol. 44, no. 11, p. 111001, Nov. 2023, doi: 10.1088/1361-6579/acead2.

- [13]J.-P. Sirkkä, T. Panula, and M. Kaisti, ‘Investigating the impact of contact pressure on photoplethysmograms’, *Biomed. Eng. Adv.*, vol. 7, p. 100123, Jun. 2024, doi: 10.1016/j.bea.2024.100123.
- [14]J. Fine *et al.*, ‘Sources of Inaccuracy in Photoplethysmography for Continuous Cardiovascular Monitoring’, *Biosensors*, vol. 11, no. 4, p. 126, Apr. 2021, doi: 10.3390/bios11040126.
- [15]M. Elgendi *et al.*, ‘The use of photoplethysmography for assessing hypertension’, *Npj Digit. Med.*, vol. 2, no. 1, p. 60, Dec. 2019, doi: 10.1038/s41746-019-0136-7.
- [16]C. Ma *et al.*, ‘KD-Informer: A cuff-Less continuous blood pressure waveform estimation approach based on single photoplethysmography’, *IEEE J. Biomed. Health Inform.*, vol. 27, no. 5, pp. 2219–2230, May 2023, doi: 10.1109/JBHI.2022.3181328.
- [17]C. Ma *et al.*, ‘PPG-based Continuous BP Waveform Estimation Using Polarized Attention-guided Conditional Adversarial Learning Model’, *IEEE J. Biomed. Health Inform.*, pp. 1–13, 2023, doi: 10.1109/JBHI.2023.3324319.
- [18]R. D. Rudic and D. J. Fulton, ‘Pressed for time: the circadian clock and hypertension’, *J. Appl. Physiol. Bethesda Md 1985*, vol. 107, no. 4, pp. 1328–1338, Oct. 2009, doi: 10.1152/japplphysiol.00661.2009.
- [19]Y. Yano and K. Kario, ‘Nocturnal blood pressure and cardiovascular disease: a review of recent advances’, *Hypertens. Res.*, vol. 35, no. 7, pp. 695–701, Jul. 2012, doi: 10.1038/hr.2012.26.
- [20]Y. Zheng, C. C. Y. Poon, B. P. Yan, and J. Y. W. Lau, ‘Pulse Arrival Time Based Cuff-Less and 24-H Wearable Blood Pressure Monitoring and its Diagnostic Value in Hypertension’, *J. Med. Syst.*, vol. 40, no. 9, p. 195, Sep. 2016, doi: 10.1007/s10916-016-0558-6.
- [21]J. Hong, Y. Zheng, S. Wu, G. Geng, Q. Liu, and C. C. Y. Poon, ‘Characterization of the Vascular System using Overnight Wearable-based Pulse Arrival Time and Ambulatory Blood Pressure: A Pilot Study’, *Comput. Biol. Med.*, p. 104861, Sep. 2021, doi: 10.1016/j.compbimed.2021.104861.
- [22]P. K. Whelton *et al.*, ‘2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines’, *Hypertension*, vol. 71, no. 6, pp. e13–e115, Jun. 2018, doi: 10.1161/HYP.0000000000000065.
- [23]S. C. Millasseau *et al.*, ‘Noninvasive Assessment of the Digital Volume Pulse’, *Hypertension*, vol. 36, no. 6, pp. 952–956, Dec. 2000, doi: 10.1161/01.HYP.36.6.952.
- [24]R. Mieloszyk *et al.*, ‘A comparison of wearable tonometry, photoplethysmography, and electrocardiography for cuffless measurement of blood pressure in an ambulatory setting’, *IEEE J. Biomed. Health Inform.*, vol. 26, no. 7, pp. 2864–2875, Jul. 2022, doi: 10.1109/JBHI.2022.3153259.

- [25]S. J. Thomas *et al.*, ‘Association of Sleep Characteristics With Nocturnal Hypertension and Nondipping Blood Pressure in the CARDIA Study’, *J. Am. Heart Assoc.*, Apr. 2020, doi: 10.1161/JAHA.119.015062.
- [26]M. Lim, S. C. Ha, K. Luk, W. Yip, C. S. Tsang, and M. C. Wong, ‘Update on the Hong Kong Reference Framework for Hypertension Care for Adults in Primary Care Settings—review of evidence on the definition of high blood pressure and goal of therapy’, *Hong Kong Med. J.*, Jan. 2019, doi: 10.12809/hkmj187701.
- [27]P. H. Charlton, P. Celka, B. Farukh, P. Chowienzyk, and J. Alastruey, ‘Assessing mental stress from the photoplethysmogram: a numerical study’, *Physiol. Meas.*, vol. 39, no. 5, p. 054001, May 2018, doi: 10.1088/1361-6579/aabe6a.
- [28]G. Sainas *et al.*, ‘Mean blood pressure assessment during post-exercise: Result from two different methods of calculation’, *J. Sports Sci. Med.*, vol. 15, no. 3, pp. 424–433, Aug. 2016.
- [29]A. de la Sierra *et al.*, ‘Prevalence and Factors Associated With Circadian Blood Pressure Patterns in Hypertensive Patients’, *Hypertension*, Mar. 2009, doi: 10.1161/HYPERTENSIONAHA.108.124008.
- [30]L. Evdochim, D. Dobrescu, S. Halichidis, L. Dobrescu, and S. Stanciu, ‘Hypertension Detection Based on Photoplethysmography Signal Morphology and Machine Learning Techniques’, *Appl. Sci.*, vol. 12, no. 16, p. 8380, Jan. 2022, doi: 10.3390/app12168380.
- [31]Y. Liang, Z. Chen, R. Ward, and M. Elgendi, ‘Hypertension Assessment via ECG and PPG Signals: An Evaluation Using MIMIC Database’, *Diagnostics*, vol. 8, no. 3, p. 65, Sep. 2018, doi: 10.3390/diagnostics8030065.
- [32]L.-C. Shih *et al.*, ‘Prediction of white-coat hypertension and white-coat uncontrolled hypertension using machine learning algorithm’, *Eur. Heart J. - Digit. Health*, vol. 3, no. 4, pp. 559–569, Dec. 2022, doi: 10.1093/ehjdh/ztac066.
- [33]N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, ‘SMOTE: Synthetic Minority Over-sampling Technique’, *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.
- [34]S. M. Lundberg and S.-I. Lee, ‘A Unified Approach to Interpreting Model Predictions’, in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2017.
- [35]M. Gurven, A. D. Blackwell, D. E. Rodríguez, J. Stieglitz, and H. Kaplan, ‘Does Blood Pressure Inevitably Rise With Age?’, *Hypertension*, vol. 60, no. 1, pp. 25–33, Jul. 2012, doi: 10.1161/HYPERTENSIONAHA.111.189100.
- [36]A. Ghosh, J. M. M. Torres, M. Danieli, and G. Riccardi, ‘Detection of essential hypertension with physiological signals from wearable devices’, in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug. 2015, pp. 8095–8098. doi: 10.1109/EMBC.2015.7320272.

- [37]K. Lan, P. Raknim, W.-F. Kao, and J.-H. Huang, ‘Toward Hypertension Prediction Based on PPG-Derived HRV Signals: a Feasibility Study’, *J. Med. Syst.*, vol. 42, no. 6, p. 103, Apr. 2018, doi: 10.1007/s10916-018-0942-5.
- [38]K. Gong, Y. Chen, X. Song, Z. Fu, and X. Ding, ‘Causal Inference for Hypertension Prediction With Wearable Electrocardiogram and Photoplethysmogram Signals: Feasibility Study’, *JMIR Cardio*, vol. 9, no. 1, p. e60238, Jan. 2025, doi: 10.2196/60238.
- [39]P. Boutouyrie, P. Chowienzyk, J. D. Humphrey, and G. F. Mitchell, ‘Arterial Stiffness and Cardiovascular Risk in Hypertension’, *Circ. Res.*, vol. 128, no. 7, pp. 864–886, Apr. 2021, doi: 10.1161/CIRCRESAHA.121.318061.
- [40]E. Martinez-Ríos, L. Montesinos, and M. Alfaro-Ponce, ‘A machine learning approach for hypertension detection based on photoplethysmography and clinical data’, *Comput. Biol. Med.*, vol. 145, p. 105479, Jun. 2022, doi: 10.1016/j.combiomed.2022.105479.
- [41]A. A. Aksenov *et al.*, ‘Chapter 8 - Volatile Organic Compounds in Human Breath: Biogenic Origin and Point-of-Care Analysis Approaches’, in *Volatile Biomarkers*, A. Amann and D. Smith, Eds., Boston: Elsevier, 2013, pp. 129–154. doi: 10.1016/B978-0-44-462613-4.00008-8.
- [42]J. M. Ahn, ‘New Aging Index Using Signal Features of Both Photoplethysmograms and Acceleration Plethysmograms’, *Healthc. Inform. Res.*, vol. 23, no. 1, pp. 53–59, Jan. 2017, doi: 10.4258/hir.2017.23.1.53.
- [43]E. Finnegan, S. Davidson, M. Harford, J. Jorge, M. Villarroel, and L. Tarassenko, ‘Classifying nocturnal blood pressure patterns using photoplethysmogram features’, in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Jul. 2022, pp. 3401–3404. doi: 10.1109/EMBC48229.2022.9871099.
- [44]T. Penzel, J. W. Kantelhardt, C.-C. Lo, K. Voigt, and C. Vogelmeier, ‘Dynamics of Heart Rate and Sleep Stages in Normals and Patients with Sleep Apnea’, *Neuropsychopharmacology*, vol. 28, no. 1, pp. S48–S53, Jul. 2003, doi: 10.1038/sj.npp.1300146.
- [45]H. Schwimmer, H. M. Stauss, F. Abboud, S. Nishino, E. Mignot, and J. M. Zeitzer, ‘Effects of sleep on the cardiovascular and thermoregulatory systems: a possible role for hypocretins’, *J. Appl. Physiol.*, vol. 109, no. 4, pp. 1053–1063, Oct. 2010, doi: 10.1152/japplphysiol.00516.2010.
- [46]K. Rahimi *et al.*, ‘Elevated blood pressure and risk of aortic valve disease: a cohort analysis of 5.4 million UK adults’, *Eur. Heart J.*, vol. 39, no. 39, pp. 3596–3603, Oct. 2018, doi: 10.1093/eurheartj/ehy486.

Tables

Table 1: Characteristics of Aurora-BP cohorts used in this study.

	Normotensive	Managed hypertensive	Unmanaged hypertensive	<i>p</i> -value
No. of subjects	138	60	42	0.739
PPG segments per subject	17±7	16±7	17±8	
PPG segment length, seconds	15	15	15	<0.01**
Age, years	45±9	49±8	49±9	
Female (%)	70 (51%)	22 (37%)	16 (38%)	0.115
Height, m	1.72±0.10	1.73±0.10	1.71±0.10	0.632
Weight, kg	84.0±21.5	95.3±23.9	95.6±28.3	<0.01**
Static SBP, mmHg	120.9±13.8	128.1±13.1	139.0±15.2	<0.01**
Static DBP, mmHg	74.4±8.5	77.8±9.1	84.8±10.5	<0.01**

Data are shown as value or mean ± SD for continuous variables and count (%) for categorical variables. Overall differences among the three groups were assessed using several statistical tests. *0.01 ≤ *p* < 0.05; ***p*<0.01.

Table 2: Characteristics of CUHK-BP cohorts used in this study.

	Normotensive	Unmanaged hypertensive	<i>p</i> -value
No. of subjects	11	15	0.022*
PPG segments per subject	33±5	28±6	
PPG segment length, seconds	60	60	<0.01**
Age, years	26±4	63±4	
Female (%)	6 (54%)	10 (67%)	0.530
Height, m	1.71±0.13	1.58±0.08	0.025*
Weight, kg	63.5±13.4	73.4±8.2	0.697
Static SBP, mmHg	120.4±11.6	137.9±21.5	0.033*
Static DBP, mmHg	70.4±6.9	73.4±8.4	0.275

Data are shown as value or mean ± SD for continuous variables and count (%) for categorical variables. Differences between the two groups were assessed using several statistical tests. *0.01 ≤ *p* < 0.05; ***p*<0.01.

Table 3: Statistical comparison of awake and sleep PPG indices across the three Aurora-BP cohorts.

Index	Normotensive	Managed hypertensive	Unmanaged hypertensive
Δt	*	-	-
CT	**	**	**
SI	*	-	-
CT/h	**	**	**
$prop_s$	*	*	-
t_{sys}	**	**	-
t_{dia}	**	**	-
t_{ratio}	*	*	-
$prop_{\Delta t}$	**	**	*
t_{p_1-dia}	**	**	-
t_{p_2-dia}	-	-	-
AI	**	**	**
AP	**	**	**
RI	**	**	-
RI_{p_1}	**	*	*
RI_{p_2}	*	*	-
$ratio_{p_2-p_1}$	**	-	-
$A1$	**	**	*
$A2$	**	**	-
IPA	-	*	-

A dash (-) indicates no statistical significance; * denotes statistical significance ($0.01 \leq p < 0.05$); ** indicates high statistical significance ($p < 0.01$).

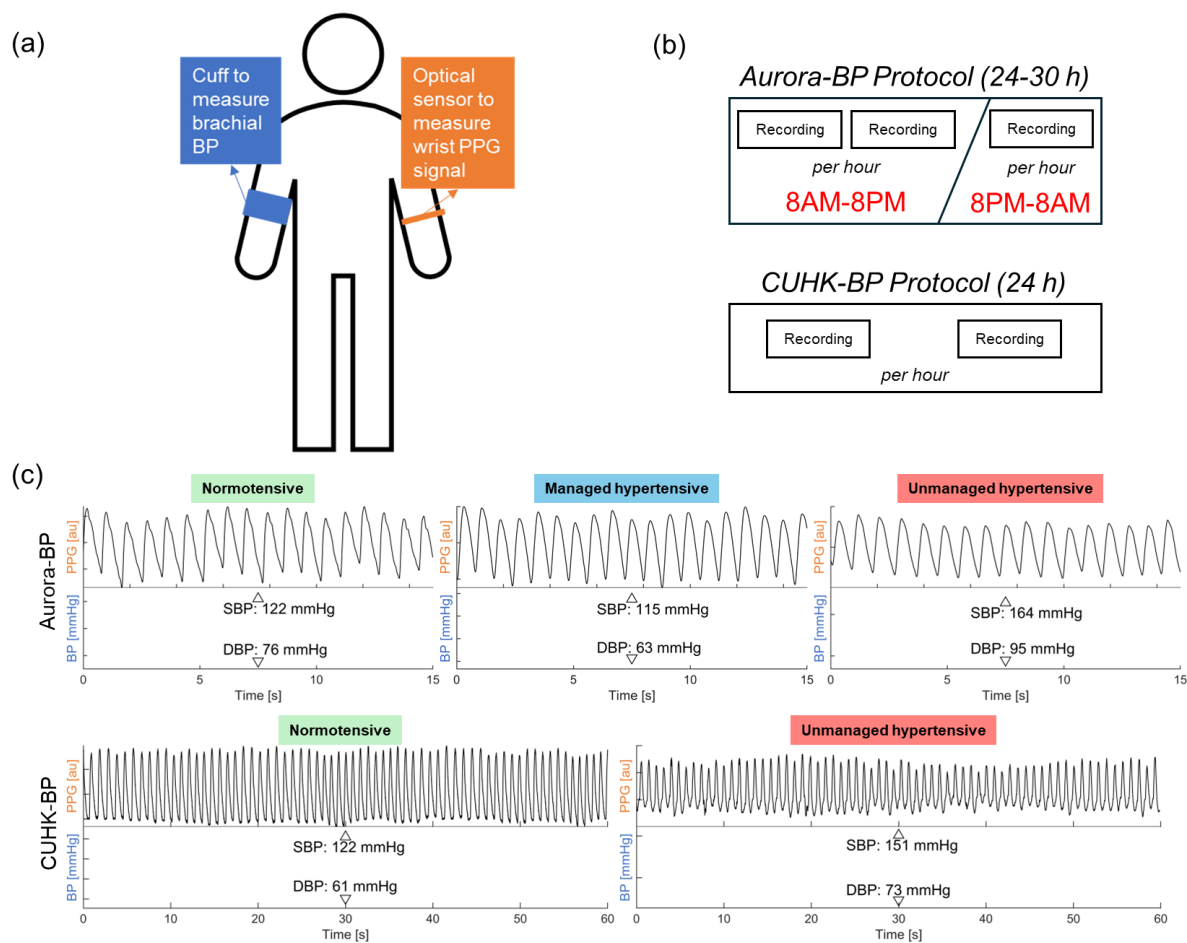
Table 4: Classification performance of the four models on internal and external test datasets using different categories of PPG-based indices.

	Index category	Logistic Regression		K-Nearest Neighbours		Support Vector Machines		eXtreme Gradient Boosting	
		Accuracy	F1-score	Accuracy	F1-score	Accuracy	F1-score	Accuracy	F1-score
Internal dataset	Complete	65.6% (4.7%)	71.1% (2.0%)	63.3% (5.0%)	61.2% (4.2%)	81.1% (8.4%)	82.8% (8.1%)	81.1% (6.3%)	82.1% (6.2%)
	Temporal	65.6% (6.1%)	69.0% (6.8%)	53.3% (3.0%)	29.6% (5.5%)	72.2% (8.8%)	73.8% (9.8%)	71.1% (4.7%)	71.2% (7.1%)
	Morphological	62.2% (8.2%)	70.0% (5.4 %)	48.9% (8.2%)	23.3% (11.4%)	65.6% (8.2%)	69.2% (7.5%)	72.2% (11.8%)	77.6% (8.0%)
External dataset	Complete	69.2%	69.2%	42.3%	44.4%	57.7%	52.2%	61.5%	61.5%
	Temporal	73.1%	75.9%	50.0%	38.1%	84.6%	86.7%	80.8%	82.8%
	Morphological	73.1%	74.1%	38.5%	27.3%	61.5%	58.3%	61.5%	58.3%

Table 5: Comparison of the proposed method with previous studies. Each study is characterised by the input type (1st column), training dataset and sample size (2nd column), test dataset and sample size (3rd column), machine learning model used (4th column), accuracy (5th column), and F1-score (6th column).

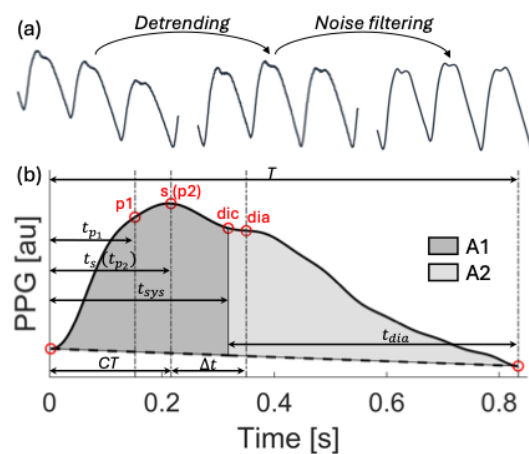
Input	Training	Test	Model	Accuracy	F1-score
17 PPG indices [39]	16 subjects (private dataset)	1 subject (private dataset)	AdaBoost		63%
10 PPG indices [31]	80 MIMIC subjects	80 MIMIC subjects	KNN		86.9%
4 PPG indices [30]	140 MIMIC III recordings	140 MIMIC III recordings	SVM	72.9%	77.0%
A PPG index [36]	30 subjects (private dataset)	13 subjects (private dataset)	Multiple instance learning	85.5%	89.2%
24 PPG indices [37]	405 Aurora-BP subjects	405 Aurora-BP subjects	Causality graphs + LR	89.0%	87.0%
20 complete PPG indices (this study)	162 Aurora-BP subjects	18 Aurora-BP subjects	Circadian statistical analysis + SVM	81.1% (8.4%)	82.8% (8.1%)
11 temporal PPG indices (this study)	180 Aurora-BP subjects	26 CUHK-BP subjects	Circadian statistical analysis + SVM	84.6%	86.7%

523 Figures



524

525 Figure 1: Ambulatory photoplethysmography (PPG) segments and blood pressure (BP) readings from
 526 the Aurora-BP and CUHK-BP datasets used in this study. (a) Data collection sites; (b) Measurement
 527 protocol; (c) Example PPG segments and corresponding BP readings from representative
 528 normotensive, managed hypertensive, and unmanaged hypertensive subjects in the Aurora-BP dataset
 529 (top) and CUHK-BP dataset (bottom).



530

531 Figure 2: (a) Signal processing steps for PPG waveform pre-processing; (b) Illustration of the PPG
 532 fiducial points used to compute the PPG-based indices listed in Supplementary Table S1. All labels
 533 are defined in Table S1.

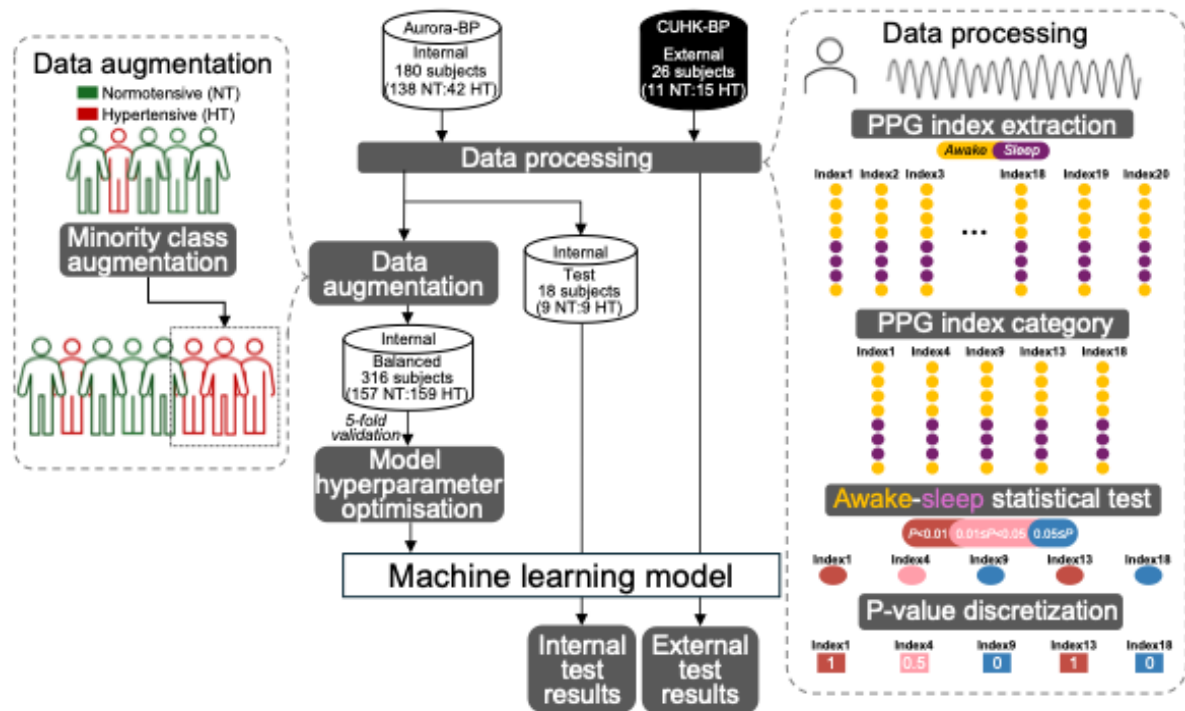


Figure 3: Flowchart of the model training and testing pipeline, including data preprocessing, feature extraction, class balancing, model training, hyperparameter optimisation, and evaluation on internal and external datasets.

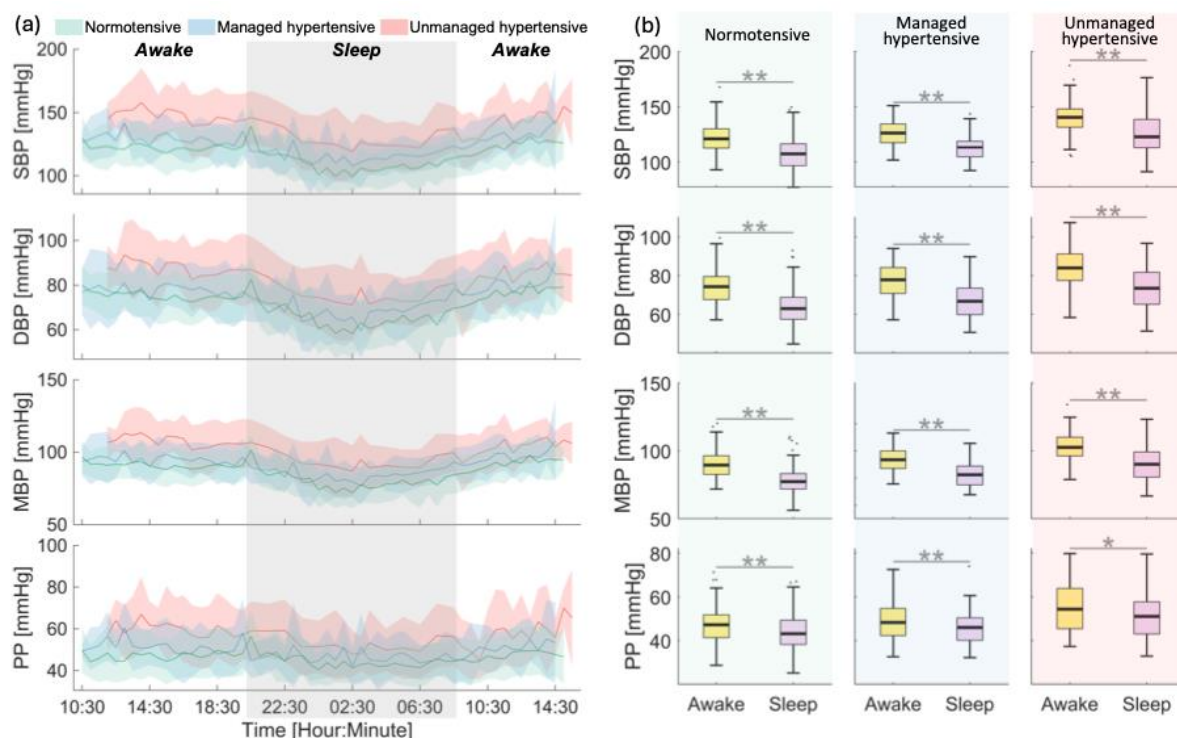


Figure 4: Blood pressure variations and statistical comparisons between awake and sleep periods across normotensive, managed hypertensive, and unmanaged hypertensive cohorts in the Aurora-BP dataset. (a) Trends in SBP, DBP, MBP, and PP across the full 24-hour ambulatory measurement period. Solid lines represent cohort means and shaded areas indicate standard deviations. (b) Statistical comparisons of SBP, DBP, MBP and PP between awake and sleep periods for each cohort.

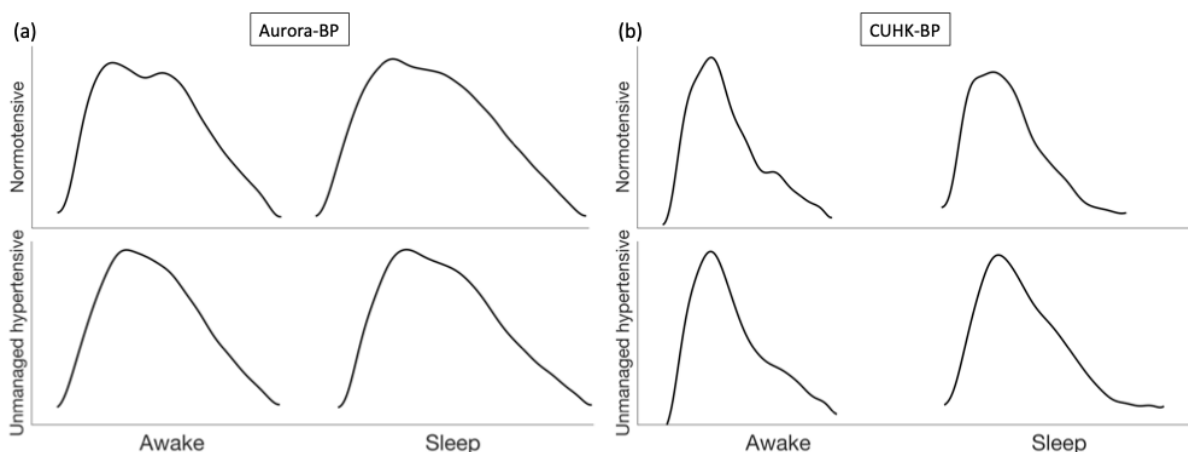


Figure 5: Example PPG waveforms recorded during awake and sleep periods in normotensive (top) and unmanaged hypertensive (bottom) individuals from the (a) Aurora-BP and (b) CUHK-BP datasets.

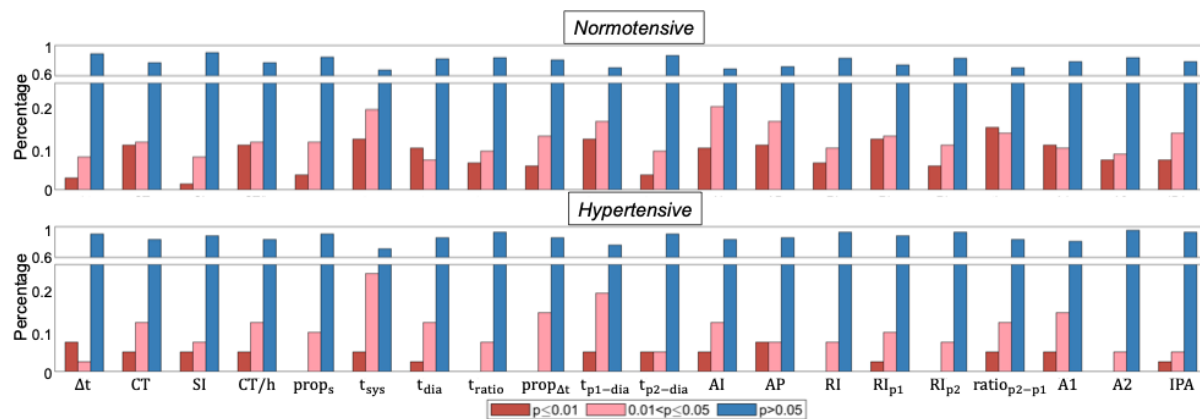


Figure 6: Frequency distribution of p-value ranges for all 20 PPG-based indices in normotensive (top) and unmanaged hypertensive (bottom) individuals from the internal Aurora-BP dataset. Highly significant values ($p < 0.01$) are shown in red, significant values ($0.01 \leq p < 0.05$) in pink, and non-significant values ($p \geq 0.05$) in blue.

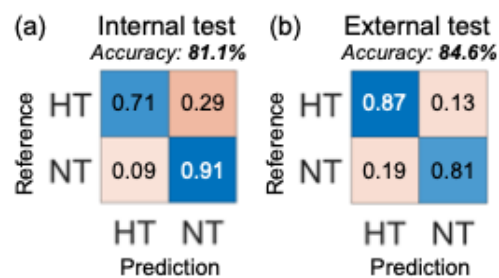


Figure 7: Best classification results of the Support Vector Machines model on (a) the internal Aurora-BP test dataset and (b) the external CUHK-BP test dataset. HT: hypertensive; NT: normotensive.

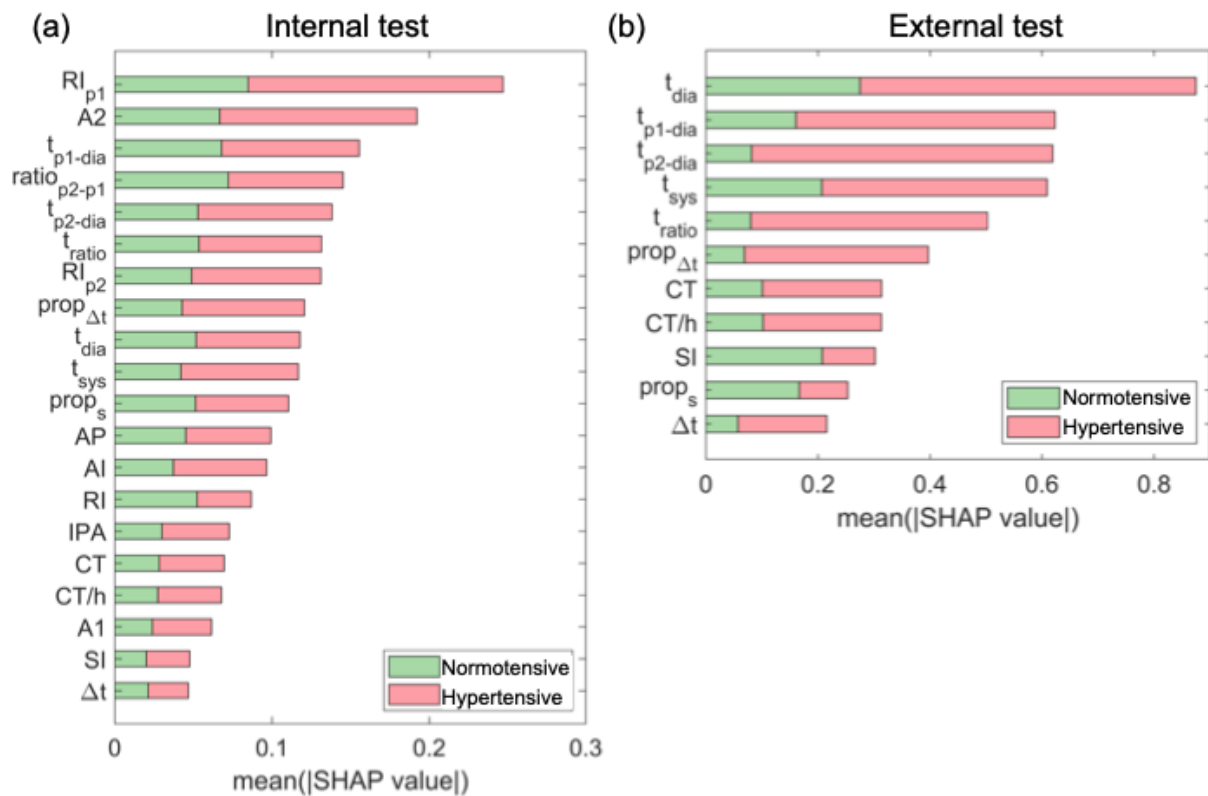


Figure 8: SHapley Additive exPlanations (SHAP) values computed using the Support Vector Machines model for (a) the complete set of PPG-based indices on the internal (Aurora-BP) test dataset and (b) the temporal indices on the external (CUHK-BP) test dataset. Green bars represent the mean absolute SHAP values contributing to normotensive predictions; red bars represent those contributing to hypertensive predictions.