iHyptn: Predicting Hypertension using PPG signal for Cardiovascular Disease with Machine Learning Models

Nidhi Sinha[1] and Amit Joshi[2]

Abstract— **Cardiovascular diseases (CVDs) are a key global health concern, accounting for a significant proportion of deaths worldwide. Early detection and precise diagnosis are required for effective CVD treatment and control. One non-invasive method that has shown promise for detecting CVDs through hypertension is the analysis of photoplethysmography(PPG) signals. It measures the difference in amount of blood and oxygen saturation in the underlying venule. PPG signals are non-invasive, cost-effective and easy to acquire, making them an ideal candidate for hypertension screening. This study proposes intelligent machine learning models to detect hypertension with the PPG signal's frequency-domain and time-domain features. The performance of the classification models are evaluated using various performance metrics such as accuracy, specificity, recall and F1-score. The results show that PPG signals can be used to accurately detect hypertension and Possibly facilitate in the early diagnosis of cardiovascular diseases. We saw that random forest classification had the highest accuracy of 93% for detecting hypertension.**

I. Introduction

Hypertension is a chronic disease where the pressure of blood increases in blood vessels. Hypertension is considered to be major public health concern in terms of its identification, treatment, and monitoring [1]. Blood pressure measurement is the main clinical indicator for recognising hypertension. It is essential to predict it early and take necessary actions to mitigate the risk of heart failure. The currently accepted non-invasive blood pressure techniques are automated and auscultatory measurements utilizing a cuff. Although arterial catheterization is a common method for continuously monitoring blood pressure, it is invasive, costly, and uncomfortable and is frequently used in intensive care facilities. Chronic elevation of blood pressure can cause damage to the walls of arteries, leading to the development of atherosclerosis. Hypertension can also lead to an enlargement of the heart, which can reduce the heart's ability to circulate blood efficiently, resulting to heart failure [2]. Several studies have found associate cardiovascular disease with high blood pressure values [3].

Hypertension is a major risk factor for cardiovascular diseases (CVD), the primary cause of mortality worldwide [4]. The relationship between hypertension and CVD is well-established. Hypertension is a significant predictor of CVD morbidity and mortality, and reducing blood pressure can reduce the incidence of CVD events by up to 40% [5]. Even mild hypertension increases the risk of CVD. American College of Cardiology found that hypertension is a leading cause of premature death from CVD and even mild hypertension i.e., 130-139 mm Hg of systolic or 80-89 mm Hg of diastolic blood pressure increases the risk of CVD. Prevention and management of hypertension can reduce the risk of CVD. The American Heart Association recommends that individuals with hypertension should monitor their blood pressure regularly, make lifestyle changes (such as reducing salt intake and increasing physical activity), and taking medication as prescribed by their healthcare provider to reduce the risk of CVD [6]. Cardiovascular diseases (CVDs) pose a significant global health challenge [7]. There are 17.9 million deaths from only cardiovascular diseases (CVDs) each year, and it is the biggest cause of mortality worldwide as of now [8]. As per the Medical Research and Registrar General of India, India bears almost 60% of the world's burden of cardiac disease. Over three-quarters of deaths due to cardiovascular devices occur in middle and low income countries and are responsible for 38 percent of the 17 million premature demises (past the age 70 years) from non-communicable diseases [9]. High blood pressure is linked to around 54% of strokes and 47% of coronary heart disease.

Early diagnosis and effective management is essential in preventing complications associated with CVDs. Compared to other cardiovascular disease (CVD) detection methods, analyzing PPG signals has several advantages. PPG signals are obtained non-invasively and continuously with a low-cost device and can provide information on arterial oxygen saturation and blood flow. PPG signals also offer real-time feedback during physical activity, making them useful in detecting

*This work was not supported by any organization

[1]Nidhi Sinha is with Research Scholar at Electronics and Communication Engineering, Malaviya National Institute of Technology, Jaipur, Rajasthan (India) 2018rec9033@mnit.ac.in

[2]Amit Joshi with the Faculty of Department of Electronics and Communication Engineering, Malaviya National Institute of Technology, Jaipur, Rajasthan (India) amjoshi.ece@mnit.ac.in

exercise-induced cardiac abnormalities. Overall, PPG signals have the potential to provide accurate and reliable methods for early detection, diagnosis, and management of CVDs, particularly in resource-limited settings. The paper are organized as follows: Section II presents related research work for PPG-based blood pressure monitoring. The proposed methodology of intelligent hypertension prediction (iHypth) is presented in Section III. The discussion about results are carried out at section IV and the final conclusion has been defined in section V.

II. Literature survey

There were several notable contributions have been presented in the literature in the past few years. Some of them are summarised in this section. Evdochim et-al, [10] proposed a signal morphology classification technique to have blood pressure estimation with PPG signal. Blood pressure was considered a significant metric as a predictor of cardiovascular characteristics and health-related events. They developed a hypertension detection tool, the ANC Test^{TM}, to analyse the PPG and integration of the medical examination using machine learning (ML) approaches.

Using machine learning algorithms, Brophy et al. [11] developed a system that can estimate arterial blood pressure (ABP) from a PPG sensor. They trained the framework with multiple models and data sources to simulate an affordable, massive dispersed shared learning experiment. Their inventive neural network of T2TGAN (time-series to time-series Generative Adversarial Network) was able to generate a continuous ABP with high quality from a PPG signal when calculating average artery pressure using an unexplored, chaotic, solitary dataset.

Martinez-rios et-al, [12] proposed using the wavelet scattering transform to extract features from the PPG data and integrate them with clinical record to detect early stages of hypertension. ML approaches and feature selection utilising the Gini Score were used to analyse clinical data (such as gender, heart rate, age and BMI) related to elevated blood pressure. Multimodal techniques (e.g., Early Fusion and Late Fusion) were investigated using clinical data (e.g., heart rate, BMI, age) for the detection of NT and PHT, and the Wavelet Scattering Transform was used to acquire PPG characteristics. On the basis of ML techniques for recognising NT and PHT classes, NT and PHT classifications were compared using unimodal and multimodal classifiers.

Ahmad et al. [13] designed a noninvasive system for blood pressure measurement with PPG signal using an optical approach. The PPG were recorded and subsequently analysed for over 450 patients with various health conditions. Using multiple machine learning algorithms, 13 elements were estimated from PPG towards systolic and diastolic blood pressure (SBP and DBP). The estimation results were comparable to the AAMI Association for the American National Standards. According to British Hypertension Society (BHS) standards, their DBP was graded A and their SBP was graded B.

Yen et al., [14] employed Photoplethysmography signals to investigate the categorise of hypertensive patients. Four deep-learning models were determined through kernel size, kernel type and layers under lesser PPG training. Signals from PPG were used to train a deep residual network convolutional neural network (ResNetCNN) and a bidirectional long short-term memory (BILSTM) to determine the optimal operational parameters for every set containing 2100 points of data. The paper presents a prediction of hypertension using effective machine learning models using simple features. The methodology also helps to identify the risk of CVD with the aid of PPG measurement.

III. Methodology

The work presents hypertension prediction with the help of data ingestion, data pre-processing, feature extraction, and modelling. The data is ingested as text files and converted into a data frame with Pandas, while the demographic features are consumed from a CSV file. The raw PPG signals are combined with the demographic features on the subject ID. Both the frequency and time domain features are extracted afterwards. Later, we combined all three kinds of features to make the final set of features, and Hypertension was set as the target variable. Then the prepared set of features and labels is applied to Machine Learning based classifiers to predict the different kinds of Hypertension, i.e., Normal, Pre-hypertension, Stage 1 hypertension, and Stage 2 hypertension. The overall workflow is given in Fig. 1.

A. Data-set

The data set used in this work is PPG-BP [15]. It is publicly available data and is downloaded from [16]. It includes a total of 657 PPG signals and corresponding blood pressure measurements from 219 participants. The distribution data is shown in Table I.

TABLE I
Distribution

Class	No. of subjects	No. of Instances
Normal	80	240
Pre-hypertension	85	255
Stage-I Hypertension	34	102
Stage-II Hypertension	20	60
Total	219	657

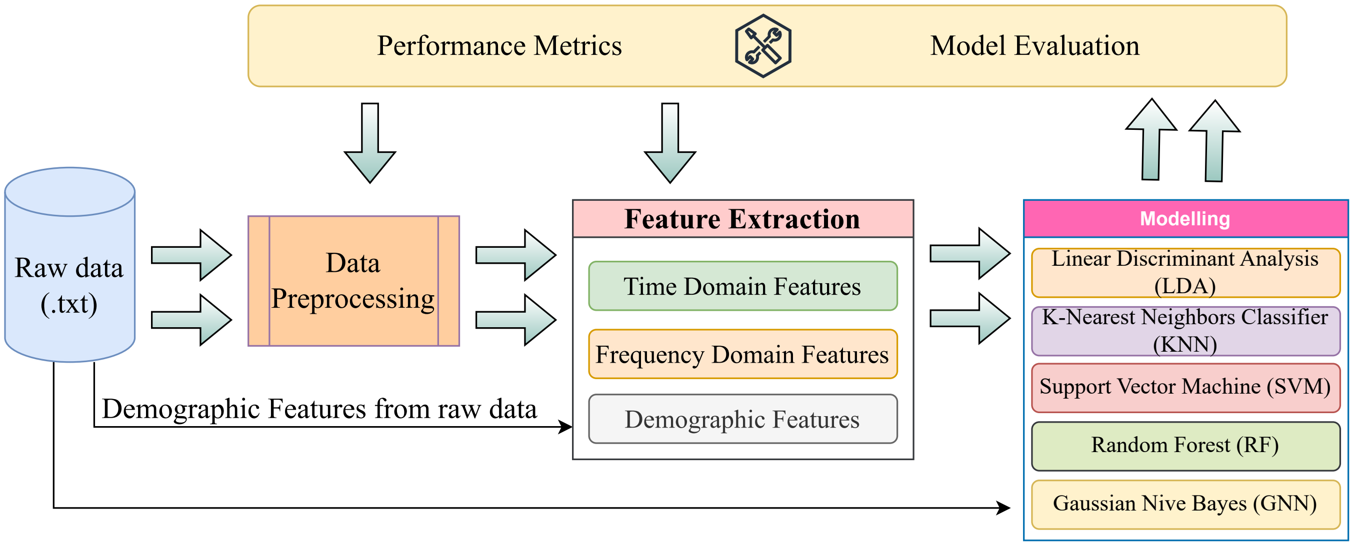

Fig. 1. Process flow of the presented work

The PPG signals were acquired from the finger and wrist using non-invasive sensors. Blood pressure measurements were taken simultaneously using an oscillometric cuff at the interval of 1 minute for a total of 10 minutes. The sampling frequency of PPG signal acquisition is 125 Hz. The dataset consists of both healthy individuals and patients with hypertension. Participants were between the ages of 23 and 71, with a mean age of 45.4. The dataset also includes demographic information such as body mass index (BMI) and sex for each participant.

B. Pre-Processing

The PPG signals are downloaded in text format and loaded into Python 3.0 using Pandas. The noise or artefacts are filtered out using a bandpass filter. It resulted in the desired frequency range between 0.5 to 90 Hz from the signals. The filtered signals are normalized using a standard scaler from sklearn. Normalization is done to rescale the values between 0 and 1. It standardizes a feature by subtracting the average and dividing it by the standard deviation. After feature scaling, resampling is performed, and the resulting data are tested for outliers. Finally, the data frame is converted into the array using NumPy for further feature extraction process.

C. Feature Extraction

The PPG-BP data have 8400 data points, i.e., raw features for each recording. Along with the raw PPG signal, demographic features of the subject were also provided, which included: age, sex, height, SBP, weight, heart rate, and diastolic blood pressure. Additionally, for a few subjects, the presence of another disease, i.e., diabetes, cerebral infarction, cerebrovascular disease, is provided, which is not used in this work. Here, frequency-domain attributes and time-domain attributes are obtained from the raw PPG signal with the aid of the Scipy and Neurokit2 libraries. The frequency and time domain features are later combined with the demographic features to prepare the final set of features for the presented work. All the features are listed in Table II.

TABLE II
LIST OF FEATURES

Demographic features	Time-domain features	Frequency-domain fetaures
age (years) sex (F/M) weight (kg) height (cm) SBP (mmHg) DBP (mmHg) Heart rate (b/m)	Mean Variance Standard deviation RMS Min Max Range Kurtosis Skewness	Energy Dominant frequency Spectral entropy

D. Modelling

Five famous machine learning algorithms, i.e., LDA, RF, KNN, SVM, and GNB, are selected to model the classifier to predict the instances of hypertension. The hyper-parameters for each of the classifier is listed in Table III-D.

The classification was performed twice with each algorithm, once with the processed PPG signal to have 8400 dimensions and again with 20 features (i.e., time-domain, frequency-domain, and demographic features). Every classifier showed significant improvement with features compared to the PPG signal. A further result

ML Classifier	Hyperparameters
RF Classifier	n_estimator: 100.0 criterion: ”gini” min_samples_split: 2.0 min_samples_leaf:: 1.0 max_features: ’sqrt’ ccp_alpha: 0.0 bootstrap: True
KNN Classifier	n_neighbors: 5.0 weights: ’uniform’ algorithm: ’auto’ leaf_size: 30 p: 02 metric: ’minkowski’ n_jobs: 1
SVM Classifier	C: 1.0 kernel: ’rbf’ degree: 3.0 gamma: ’scale’ shrinking: True tol: 1e-3 max_iter: 1 decision_function_shape: ’ovr’
LDA Classifier	solver: ’svd’
GNB Classifier	Priors: None Var_smoothing: 1e-9

of each of the classifiers in both cases is discussed in the next section.

IV. Results and Discussion

Following the extraction of features, data is prepared for modelling. Eighty percent of the data are used for training, while twenty percent are used for assessment. Five classifiers are utilised to categorise the various cases of hypertension.

Based on the previous literature the five classifiers are selected, which are Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest (RF), Gaussian Naive Bayes (GNB), and Linear Discriminant Analysis (LDA) [17]. Five- Fold Cross Validation is used to compare the performance of the presented five classifiers. Along with accuracy, recall, precision, and F1 score are used as performance measures [18]. All the reported values are average of five-fold Cross Validation. Table III gives each classifier’s performance measure with processed PPG signal.

TABLE III
Performance metrics of classifiers with processed PPG signal

Classifier(s)	Precision	Recall	F1 score	Accuracy
LDA	84	83	84	83
GNB	71	70	70	70
KNN	79	79	79	79
SVM	67	65	64	65
RF	73	72	72	72

TABLE IV
Performance metrics of classifiers with extracted features

Classifier(s)	Precision	Recall	F1 score	Accuracy
LDA	91	90	90	92
GNB	86	86	85	88
KNN	84	84	84	86
SVM	80	76	78	80
RF	91	89	90	93

Further, the performance measures of each classifier with extracted features (i.e. time-domain and frequency-domain) along with the demographic feature are listed in Table IV. The LDA performs better with raw PPG signal, recording the blood volume variation with a linear relation with Hypertension.

Although, Random Forest performs best with the extracted features due to the following reasons: (a) The features include binary, categorical, and numerical or continuous variables (b) RF is robust to non-linear features (c) It handles the unbalanced data quite well. The class-wise performance matrices for each of the classifier with processed PPG signal is given in Table V, and with extracted features, it is given in Table VI.

The presented work is compared with previous work and the same is listed in Table IV. It is evident from the table that the presented machine learning model classifier’s performance measures are much better than the others. Hence, the set of extracted and demographic features gives the best prediction for different hypertension classes.

V. Conclusion

The performance of five classifiers is studied with a processed PPG signal and with the feature(s) extracted from the PPG signal. It has been concluded that the LDA works best with processed PPG signals and has a five-fold mean cross-validation score of 83%. At the same time, ensemble machine learning-based classifier RF gives the best accuracy, i.e., 93%, with extracted features. Overall, the second approach is better as it uses only 20 features (i.e., nine time-domain, three freq-domain, and eight demographic features) compared to the first approach, which uses 8400 features/dimensions of the PPG signal. In terms of recall, which is considered one of the critical performance measures in disease classification, the second approach is better. Although, LDA gives the best recall in both approaches. As all the classifiers here are used with their default parameters in both cases, there is scope for improvement by tuning the models’ hyperparameter.

TABLE V

CLASS-WISE PERFORMANCE METRICS OF EACH CLASSIFIER WITH PROCESSED PPG SIGNALS

	LDA			GNB			KNN			SVM			RF		
Class	Precision	Recall	F1 Score	Precision	Recall	F1 Score	Precision	Recall	F1 Score	Precision	Recall	F1 Score	Precision	Recall	F1 Score
Normal	77	94	85	76	76	76	79	86	82	70	63	66	78	72	75
Pre-hypertension	88	72	79	65	70	68	78	75	76	57	75	65	65	81	72
Stage-I Hypertension	90	80	85	65	63	64	78	71	74	78	47	59	80	57	66
Stage-II Hypertension	83	90	86	81	62	70	81	82	82	81	55	66	83	57	68

TABLE VI

CLASS-WISE PERFORMANCE METRICS OF EACH CLASSIFIER WITH EXTRACTED FEATURES SIGNALS

	LDA			GNB			KNN			SVM			RF		
Class	Precision	Recall	F1 Score	Precision	Recall	F1 Score	Precision	Recall	F1 Score	Precision	Recall	F1 Score	Precision	Recall	F1 Score
Normal	91	98	95	90	98	94	89	95	92	85	85	85	95	99	97
Pre-hypertension	95	91	93	94	83	88	90	82	86	77	81	79	95	94	95
Stage-I Hypertension	90	88	89	75	86	80	74	83	78	74	73	73	87	86	86
Stage-II Hypertension	88	82	85	85	76	80	84	75	80	84	66	74	87	75	81

TABLE VII

COMPARISON OF THE PRESENTED WORK WITH SOME PREVIOUS LITERATURE

Previous Work	Class Included	Dataset	ML Classifier	Performance Metrics			
				Accuracy	Precision	Recall	F1 score
Evdochim et-al, 2022 [10]	Normal and Hypertension	MIMIC III	Quadratic SVM	72.90	84.50	71.00	77.00
Martinez-rios et-al, 2022 [12]	Normal and Prehypertension	PPG-BP	SVM	71.42	65.51	90.47	76.00
Yen et-al, 2021 [14]	Normal, Pre-hypertension, Stage 1 hypertension, and Stage 2 hypertension	PPG BP	Xception + BILSTM	76.00	48.00	45.00	–
Presented work	Normal, Pre-hypertension, Stage-1 hypertension, and Stage-2 hypertension	PPG BP	Random Forest	**93.00**	**91.00**	**90.00**	**90.00**

REFERENCES

[1] A. Chandrasekhar, M. Yavarimanesh, K. Natarajan, J.-O. Hahn, and R. Mukkamala, “Ppg sensor contact pressure should be taken into account for cuff-less blood pressure measurement,” *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 11, pp. 3134–3140, 2020.

[2] S. Gupta, A. Singh, A. Sharma, and R. K. Tripathy, “Higher order derivative-based integrated model for cuff-less blood pressure estimation and stratification using ppg signals,” *IEEE Sensors Journal*, vol. 22, no. 22, pp. 22 030–22 039, 2022.

[3] G. R. Nidigattu, G. Mattela, and S. Jana, “Non-invasive modeling of heart rate and blood pressure from a photoplethysmography by using machine learning techniques,” in *2020 International Conference on COMmunication Systems & NETworkS (COMSNETS)*. IEEE, 2020, pp. 7–12.

[4] N. Sinha, T. Jangid, A. M. Joshi, and S. P. Mohanty, “icardo: A machine learning based smart healthcare framework for cardiovascular disease prediction,” *arXiv preprint arXiv:2212.08022*, 2022.

[5] D. Li, X. Zeng, Y. Huang, H. Lei, G. Li, N. Zhang, and W. Huang, “Increased risk of hypertension in young adults in southwest china: impact of the 2017 acc/aha high blood pressure guideline,” *Current hypertension reports*, vol. 21, no. 3, p. 21, 2019.

[6] N. Sinha and A. M. JOSHI, “Predicting the presence of left ventricular hypertrophy using este’s criteria with support vector machine,” 2022.

[7] N. Sinha, A. M. Joshi, and S. P. Mohanty, “icardo 2.0: A smart healthcare framework for cardiovascular disease accurate prediction by using t-wave morphology of ecg,” in *2022 IEEE International Symposium on Smart Electronic Systems (iSES)*. IEEE, 2022, pp. 343–348.

[8] F. D. Fuchs and P. K. Whelton, “High blood pressure and cardiovascular disease,” *Hypertension*, vol. 75, no. 2, pp. 285–292, 2020.

[9] P. Jain, A. M. Joshi, and S. Mohanty, “Everything you wanted to know about noninvasive glucose measurement and control,” *arXiv preprint arXiv:2101.08996*, 2021.

[10] L. Evdochim, D. Dobrescu, S. Halichidis, L. Dobrescu, and

S. Stanciu, “Hypertension detection based on photoplethysmography signal morphology and machine learning techniques,” *Applied Sciences*, vol. 12, no. 16, p. 8380, 2022.

[11] E. Brophy, M. De Vos, G. Boylan, and T. Ward, “Estimation of continuous blood pressure from ppg via a federated learning approach,” *Sensors*, vol. 21, no. 18, p. 6311, 2021.

[12] E. Martinez-Ríos, L. Montesinos, and M. Alfaro-Ponce, “A machine learning approach for hypertension detection based on photoplethysmography and clinical data,” *Computers in Biology and Medicine*, vol. 145, p. 105479, 2022.

[13] A. Dagamseh, Q. Qananwah, H. Al Quran, and K. S. Ibrahim, “Towards a portable-noninvasive blood pressure monitoring system utilizing the photoplethysmogram signal,” *Biomedical Optics Express*, vol. 12, no. 12, pp. 7732–7751, 2021.

[14] C.-T. Yen, S.-N. Chang, and C.-H. Liao, “Deep learning algorithm evaluation of hypertension classification in less photoplethysmography signals conditions,” *Measurement and Control*, vol. 54, no. 3-4, pp. 439–445, 2021.

[15] Y. Liang, Z. Chen, G. Liu, and M. Elgendi, “A new, short-recorded photoplethysmogram dataset for blood pressure monitoring in china,” *Scientific Data*, vol. 5, no. 1, p. 180020, Feb 2018. [Online]. Available: https://doi.org/10.1038/sdata.2018.20

[16] Y. Liang, G. Liu, Z. Chen, and M. Elgendi, “Ppg-bp database,” 9 2022.

[17] S. Pancholi and A. M. Joshi, “Intelligent upper-limb prosthetic control (iulp) with novel feature extraction method for pattern recognition using emg,” *Journal of Mechanics in Medicine and Biology*, vol. 21, no. 06, p. 2150043, 2021.

[18] G. Sharma, A. M. Joshi, R. Gupta, and L. R. Cenkeramaddi, “Depcap: A smart healthcare framework for eeg based depression detection using time-frequency response and deep neural network,” *IEEE Access*, 2023.