JOURNAL OF DATA SCIENCE | Vol.2024:32
eISSN:2805-5160

Early Detection of Cardiovascular Disease Using Photoplethysmography (PPG) Signal Analysis

Padmavathi Y [1], Ushasree R [1]

[1]Dayananda Sagar Academy of Technology and Management, Karnataka, India

Email: [1]padmayallappa10@gmail.com, [2]Ushasree19832000@gmail.com

Abstract

Photoplethysmography (PPG) signals have gained prominence in clinical diagnostics for their non-invasive, cost-effective, and user-friendly applications in detecting cardiovascular diseases (CVDs). This study leverages machine learning techniques to enhance the accuracy of CVD detection from PPG data, addressing critical risk factors such as hypertension and stress, which significantly contribute to elevated blood pressure and, consequently, to cardiovascular disorders. The use of PPG provides a reliable approach for identifying cardiovascular anomalies by monitoring essential parameters like blood pressure and heart rate. In this work, we employ both machine learning and deep learning, specifically neural networks, to assist clinicians in diagnosing CVD, achieving a high accuracy rate of 98% on the PPG-BP dataset. The findings demonstrate the potential of PPG signals combined with advanced algorithms to support early diagnosis and personalized treatment, ultimately reducing mortality rates associated with cardiovascular diseases.

Keywords

Photoplethysmography (PPG); Neural Network (NN); Cardiovascular disease (CVD); Blood Pressure (BP)

Introduction

Photoplethysmography (PPG) is recognized as an effective tool for the rapid diagnosis of cardiac-related disorders. PPG works by measuring blood volume fluctuations within tissues, making it highly effective in the early screening of cardiovascular diseases (CVDs) (Palanisamy & Rajaguru, 2023). The PPG signal captures the circulation of blood from the heart to peripheral areas, providing a continuous assessment of cardiovascular function. CVDs remain a leading cause of death globally; in fact, cardiac abnormalities contributed to 29.6% of worldwide mortality in 2010, highlighting the urgent need for early and accessible diagnostic tools (Shabaan et al., 2020).

PPG devices are relatively easy to use and require minimal training, making them accessible for providing accurate real-time insights (Weng et al., 2023). This study will investigate CVD prediction using data from the PhysioNet database, which includes comprehensive datasets for clinical analysis. For example, PhysioNet's files contain raw signals from electrocardiogram

Submission: 27 June 2024; **Acceptance**: 28 October 2024

(ECG), photoplethysmography (PPG), and arterial blood pressure (ABP) data stored as matrices within MATLAB files, allowing for in-depth analysis of cardiovascular markers.
The primary goal of this research is to improve classifier performance in detecting cardiovascular disease through PPG signals. Using the discrete cosine transform (DCT) to create training and testing sets with K-fold cross-validation, we aim to refine the model's predictive capabilities. Various classifiers, including linear regression and neural networks, are applied to preprocess PPG data. These machine learning and deep learning techniques significantly enhance the ability of classifiers to detect CVD from PPG signals, advancing the reliability of non-invasive cardiac health monitoring (Prabhakar, Rajaguru, & Lee, 2019; Ave et al., 2015).

Photoplethysmography (PPG) is emerging as a promising technique in the early diagnosis of cardiac disorders. It is non-invasive, simple, and measures blood volume fluctuations within tissues to detect various cardiovascular diseases (CVDs) (Prabhakar, Rajaguru, & Lee, 2019). By capturing changes in light absorption that correspond to blood flow dynamics, PPG signals allow clinicians to assess critical aspects of cardiovascular function, including blood flow, heart rate, and circulation time. These measurements, when spectrally analyzed, can reveal variations in heart rate and blood volume levels, enabling the detection of subtle physiological changes associated with cardiovascular issues (Palanisamy & Rajaguru, 2023).

Blood pressure (BP), as a bio-physiological signal, plays a vital role in health monitoring, particularly for conditions such as heart disease, stroke, and kidney failure, all of which are influenced by high BP. Consequently, accurate prediction and continuous monitoring of BP are essential for effective diagnosis, prevention, and treatment of these conditions (Stojanova, Koceski, & Koceska, 2019). A high-quality PPG signal can be recorded by placing a finger on the PPG device, providing a simpler and more accessible alternative to the electrocardiogram (ECG), which typically requires more complex hardware (Ramachandran, Thangapandian, & Rajaguru, 2020).

Given that cardiovascular diseases remain one of the leading causes of death globally, early detection is critical for reducing CVD-related morbidity and mortality (Ave et al., 2015). In response, various predictive models, such as those using linear regression and neural networks, have been developed to analyze PPG data and predict cardiovascular events effectively (Li, 2018). The design and functionality of PPG technology are straightforward, yet it allows for in-depth analysis when paired with large datasets like the PhysioNet database, significantly aiding CVD detection (Rajaguru, Shankar, Nanthakumar, & Murugan, 2023).

The PPG process operates through infrared light sensors that measure cardiovascular signals at peripheral sites on the body. This technology is widely applied in pulse oximetry and other healthcare areas, owing to its cost-effectiveness and ease of use. By analyzing PPG signals, it is possible to achieve up to 98% accuracy in cardiovascular disease predictions, a crucial benchmark in improving patient outcomes (Prabhakar, Rajaguru, & Kim, 2020). Utilizing neural networks, which have shown significant promise in healthcare applications, PPG-based algorithms can capture essential features such as heart rate variability and pulse transit time for CVD detection. Deep learning models, in particular, are highly effective in analyzing these data for clinical applications (Weng et al., 2023).

Critical advancement in PPG technology involves signal denoising, which enhances the accuracy of CVD predictions. Effective denoising of PPG signals before analysis can improve performance metrics, making this preprocessing step valuable in clinical settings. Independent of other clinical analytics, denoising demonstrates the potential to refine physiological signal data and elevate diagnostic efficacy (Ukil et al., 2016).

Methodology

The methodology of this study involves several key steps as depicted in Figure 1. First, data collection is carried out, gathering essential information on patients, including gender, age, height, weight, blood pressure, and heart rate. These variables are instrumental in distinguishing between normal and abnormal cardiovascular cases. Data preprocessing follows, involving reshaping the data, noise removal, and signal filtering. These steps are essential for preparing the data for accurate analysis, ensuring that the PPG signals are free from artifacts and ready for input into the neural network model. Through this multi-stage approach, the study aims to develop a robust predictive model for early cardiovascular disease detection based on PPG signals.

The objective of this study is to utilize various subject features to differentiate between normal and diseased cardiovascular cases using photoplethysmography (PPG) signal analysis. By gathering data on demographic and physiological variables, this research aims to develop a model capable of identifying cardiovascular disease patterns within PPG signals, enhancing diagnostic accuracy and early detection in clinical settings.

Age plays a significant role in this analysis, as it directly affects vascular stiffness, which in turn influences the shape and characteristics of PPG signals. As an individual age, changes in vascular flexibility can alter the PPG waveform, making age a critical factor in assessing cardiovascular health. Additionally, gender influences various physiological indicators; research has shown that men and women experience different cardiovascular disease (CVD) events, such as heart failure, due to gender-specific factors (Regitz-Zagrosek et al., 2006). Including gender in the data collection provides insights into the diverse CVD manifestations across populations.

Body metrics, such as height and weight, are used to calculate the Body Mass Index (BMI), which serves as a general indicator of health and potential cardiovascular risk. Blood pressure (BP), however, is the primary measure of cardiovascular health, as hypertension or abnormal BP levels often signal underlying cardiovascular issues. Heart rate also acts as a valuable indicator; variations in heart rate can reflect abnormalities in heart function, stress levels, and levels of physical activity, providing further information for assessing cardiovascular risk.

For classification, the study applies machine learning and deep learning techniques, particularly using a neural network (NN) model. The neural network model is trained on PPG signal features, employing K-fold cross-validation to ensure robust performance. The model is designed to predict the likelihood of cardiovascular disease based on observed PPG characteristics, utilizing TensorFlow as the primary framework for implementing and optimizing the neural network.

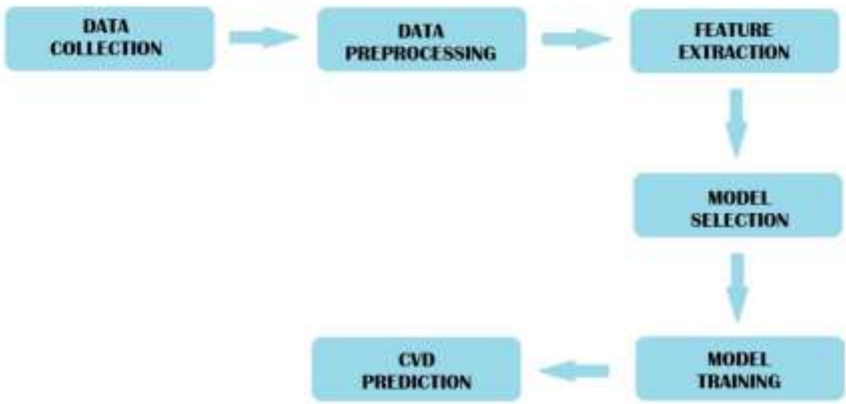

Figure 1. Schematic diagram for the implementation

Feature extraction is a vital step in preparing PPG signals for analysis, as it enhances the relevant information needed for accurate cardiovascular disease predictions. In this study, cross-correlation techniques are applied to align and synchronize PPG signals, which optimizes the extraction of essential signal features for the predictive model. Additionally, the Discrete Cosine Transform (DCT) is employed for signal compression, noise removal, and dimensionality reduction. By improving the quality of signal representation, DCT facilitates the isolation of key characteristics within the PPG data, thus supporting the accurate detection of cardiovascular anomalies.

Model selection is conducted using K-fold cross-validation, a method that divides the data into K subsets, training and validating the model iteratively on each subset. This technique ensures that the chosen model performs well on average across all K iterations, thereby enhancing reliability and reducing the likelihood of overfitting. Through this process, the neural network model selected is highly optimized for detecting cardiovascular disease based on PPG signal features. This neural network model, specifically trained on relevant data, predicts the likelihood of cardiovascular disease by recognizing the characteristic patterns observed in PPG data.

The model training phase involves using a subset of the data rows to accelerate the training process while maintaining accuracy. This subset approach allows for efficient training, particularly for large datasets, without sacrificing performance. The neural network model, implemented within TensorFlow, is iteratively trained to identify patterns that correlate with cardiovascular disease, refining its accuracy with each iteration.

In the final step, prediction, the trained neural network model is applied to new data, enabling it to make predictions about cardiovascular health. By visualizing the model's predicted results, an accuracy rate of 98% is achieved, demonstrating the effectiveness of using PPG data combined with deep learning techniques. This high accuracy highlights the potential of PPG signals and machine learning to aid in the early detection of cardiovascular diseases, offering a powerful, non-invasive solution for clinical practice.

Results

The graph in Figure 2 illustrates the comparison between true blood pressure (BP) values and the predicted BP values over a series of 100 samples. The blue line represents the actual BP values, while the orange line shows the predicted BP values generated by a machine-learning model. From the visual representation, it is evident that the predicted BP values closely follow the general trend of the true BP values, although there are occasional deviations, particularly during peaks and troughs in the true BP data.

This alignment indicates that the model has achieved a reasonable level of accuracy in estimating BP values, albeit with some limitations in tracking rapid fluctuations or extreme values accurately. Such results suggest that the model may be effective in capturing average BP trends, but further optimization may be required to improve its performance during abrupt changes. Overall, this comparison highlights the potential of predictive models in estimating BP values non-invasively using alternative methods such as PPG signals, which can support continuous monitoring of cardiovascular health. This can be seen in the figure 2.

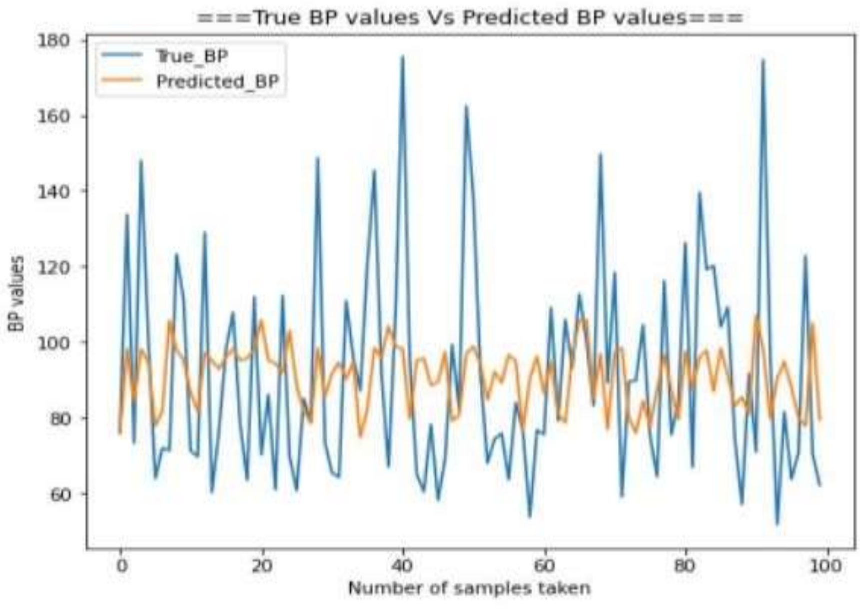

Figure 2. Visualization of predicted values using linear regression

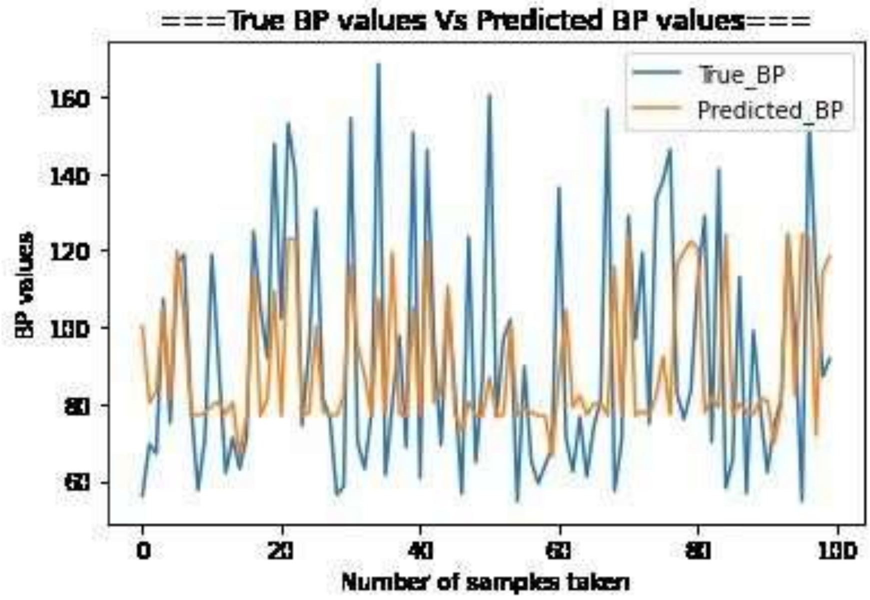

Figure 3. Visualization of Predicted BP and the true BP of Neural Network

The graph in Figure 3 depicts the comparison between true blood pressure (BP) values and predicted BP values across 100 samples, with the blue line representing the true BP values and the orange line indicating the predicted BP values generated by a model. The alignment between the true and predicted values is relatively consistent, suggesting that the model can capture general BP trends. However, it also shows some noticeable discrepancies, especially during sudden spikes and drops, where the predicted values do not fully capture the intensity of these fluctuations.

This pattern suggests that while the model performs well in following the average trend of BP changes, it may struggle to accurately track rapid variations in BP. This could be due to limitations in the model's sensitivity to extreme or abrupt changes, which are challenging to predict precisely. Despite these challenges, the overall alignment between the two lines indicates that this model could be beneficial for monitoring BP patterns, potentially assisting in continuous cardiovascular monitoring through non-invasive methods like PPG signals. Further refinement may enhance its accuracy, especially in capturing rapid BP fluctuations.

Conclusion

PPG signals hold significant promise for the early detection and prevention of cardiovascular disease (CVD) by continuously monitoring vital signs and identifying abnormal patterns. By incorporating machine learning algorithms, PPG-based systems have the potential to transform cardiovascular healthcare, enabling timely interventions and enhancing patient outcomes. This project focused on detecting CVD using PPG signals, leveraging datasets processed through classifiers to achieve high classification accuracy with minimal false positives. While linear regression classifiers were less effective across all five dimensionality reduction techniques, a more targeted approach—using one-minute segments of raw PPG signals—proved successful. The optimized neural network model achieved an impressive classification accuracy of 98%, underscoring the effectiveness of advanced machine learning techniques in supporting non-invasive CVD diagnostics.

Acknowledgment

The authors extend sincere gratitude to Dayananda Sagar Academy of Technology and Management (DSATM) for providing the resources and facilities essential for conducting this research on “CVD Detection Using Photoplethysmography (PPG) Signal Data.” We also thank our families, especially our mothers, for their unwavering support, encouragement, and financial assistance, which enabled us to complete this project with dedication. Our heartfelt appreciation goes to all who contributed to the success of this research.

References

Ave, A., Fauzan, H., Adhitya, S. R., & Zakaria, H. (2015). Early detection of cardiovascular disease with photoplethysmogram (PPG) sensor. In *2015 International Conference on Electrical Engineering and Informatics (ICEEI)* (pp. 676-681). IEEE. https://doi.org/10.1109/ICEEI.2015.7352584

Li, G. (2018). Application of finite mixture of logistic regression for heterogeneous merging behavior analysis. *Journal of Advanced Transportation, 2018*, 1-9. https://doi.org/10.1155/2018/1436521

Palanisamy, S., & Rajaguru, H. (2023). Machine learning techniques for the performance enhancement of multiple classifiers in the detection of cardiovascular disease from PPG signals. *Bioengineering, 10*(6), 678. https://doi.org/10.3390/bioengineering10060678

Prabhakar, S. K., Rajaguru, H., & Kim, S.-H. (2020). Fuzzy-inspired photoplethysmography signal classification with bio-inspired optimization for analyzing cardiovascular disorders. *Diagnostics, 10*(10), 763. https://doi.org/10.3390/diagnostics10100763

Prabhakar, S. K., Rajaguru, H., & Lee, S. W. (2019). Metaheuristic-based dimensionality reduction and classification analysis of PPG signals for interpreting cardiovascular disease. *IEEE Access, 7*, 165181-165206. https://doi.org/10.1109/ACCESS.2019.2950220

Rajaguru, H., Shankar, M. G., Nanthakumar, S. P., & Murugan, I. A. (2023). Performance analysis of classifiers in detection of CVD using PPG signals. In *AIP Conference Proceedings, 2725*(1). AIP Publishing. https://doi.org/10.1063/5.0125222

Ramachandran, D., Thangapandian, V. P., & Rajaguru, H. (2020). Computerized approach for cardiovascular risk level detection using photoplethysmography signals. *Measurement, 150*, 107048. https://doi.org/10.1016/j.measurement.2019.107048

Regitz-Zagrosek, V., Lehmkuhl, E., & Weickert, M. O. (2006). Gender differences in the metabolic syndrome and their role for cardiovascular disease. *Clinical Research in Cardiology, 95*, 147-147. https://doi.org/10.1007/s00392-006-0351-5

Shabaan, M., Arshid, K., Yaqub, M., Jinchao, F., Zia, M. S., Bojja, G. R., Iftikhar, M., Ghani, U., Ambati, L. S., & Munir, R. (2020). Survey: Smartphone-based assessment of cardiovascular diseases using ECG and PPG analysis. *BMC Medical Informatics and Decision Making, 20*, 1-16. https://doi.org/10.1186/s12911-020-01199-7

Stojanova, A., Koceski, S., & Koceska, N. (2019). Continuous blood pressure monitoring as a basis for ambient assisted living (AAL)–Review of methodologies and devices. *Journal of Medical Systems, 43*, 1-12. https://doi.org/10.1007/s10916-018-1138-8

Ukil, A., et al. (2016). Cardiac condition monitoring through photoplethysmogram signal denoising using wearables: Can we detect coronary artery disease with higher performance efficacy? In *2016 Computing in Cardiology Conference (CinC)*. IEEE. https://doi.org/10.22489/CinC.2016.082-334

Weng, W. H., Baur, S., Daswani, M., Chen, C., Harrell, L., Kakarmath, S., ... & Ardila, D. (2023). Predicting cardiovascular disease risk using photoplethysmography. *arXiv preprint arXiv:2305.05648*. https://doi.org/10.1371/journal.pgph.0003204