# scientific reports

# OPEN



# Machine learning analysis of integrated ABP and PPG signals towards early detection of coronary artery disease

Amandeep Minhas<sup>1</sup>, Subhash Chandra Pal<sup>1</sup> & Karan Jain<sup>2</sup>

Every year, Coronary Artery Disease (CAD) claims lives of over a million people. CAD occurs when the coronary arteries, responsible for supplying oxygenated blood to the heart, get occluded due to plaque deposits on their inner walls. The most critical fact about this disease is that it develops gradually over the years and by the time symptomatic changes such as angina or shortness of breath appear, the disease has already become severe. The overall aim of the proposed work is to detect CAD efficiently in its early stage while utilizing (radial) arterial blood pressure (ABP) along with photoplethysmogram (PPG) signals so that necessary clinical measures may be taken timely. To achieve this objective, firstly, ABP and PPG data of 73 CAD and 64 non-CAD (not suffering from any cardiac condition) subjects have been collected from MIMIC-II waveform database with matched subset. Secondly, the collected data is pre-processed using band pass filters having bandwidths of 2.5 to 16 Hz and 1.5 to 16 Hz for ABP and PPG respectively. Thirdly, nineteen features have been extracted from each of the two signals; some of the key features include mean of pulse duration, mean of rising slope and ratio of low frequency to high frequency. Finally, extensive analysis on CAD and non-CAD classification is carried out on the basis of extracted features while employing state-of-the-art classifiers such as support vector machines (SVM), K-nearest neighbors (KNN) and neural networks(NN). The numerical experiments have led to the interpretation that neural network outperforms other classifiers, claiming an accuracy of about 90%. Moreover, accuracy of the proposed approach is found to be better than the state-ofthe-art works reported in literature where one of or combinations of cardiovascular signals, namely, electrocardiogram (ECG), phonocardiogram (PCG) and photoplethysmogram (PPG) have been utilized for the CAD detection.

Keywords Coronary artery disease, ABP, PPG, Classification, Diagnosis, Prognosis

The global mortality rate due to cardiovascular diseases (CVDs) cause a major concern, approximately 17.9 million deaths in 2019 is being attributed to CVDs<sup>1</sup>. Coronary artery disease is one of the major CVDs; in this pathological condition, the coronary arteries, responsible for supplying oxygenated blood to the heart, get occluded due to plaque deposits on their inner walls<sup>2</sup>. This may further lead to symptoms such as chest pain, shortness of breath, heart attack in the advanced stage of disease development. The risk factors include high blood pressure, high cholesterols, diabetes, smoking, obesity and a family history of heart disease<sup>3</sup>. Treatment options for the said condition range from lifestyle changes, medication to surgery. It may be noted that once the disease has become severe, it requires clinical intervention, else, it may prove to be fatal. While coronary catheterization is a gold standard for diagnosing CAD<sup>2</sup>, it is an invasive and expensive procedure conducted by skilled cardiologists. The proposed approach for detecting CAD would be easily accessible, low-cost, non-invasive, and suited for domestic use. Some study's have shown that different machine learning techniques have remarkable impact on heart disease detection<sup>4</sup>.

Medical assessment of CAD typically includes electrocardiography (ECG), treadmill test (TMT), and cardiac catheterization<sup>3</sup>. CAD may not exhibit any of the symptoms or changes in ECG signal until stenosis has become severe<sup>5</sup>. Moreover, cardiac catheterization is a completely invasive process as mentioned above. Also, The sensitivity of TMT is 25–71% for subjects having one artery blockage; the sensitivity of this test improves to 81%

<sup>1</sup>Electrical Engineering Department, National Institute of Technology Durgapur, Durgapur, West Bengal 713209, India. <sup>2</sup>Instrumentation and Control Engineering, Dr B R Ambedkar National Institute of Technology, Jalandhar, Jalandhar, Punjab 144027, India. <sup>Sem</sup>email: amanminhas01@gmail.com

in case of 2–3 arteries' blockage<sup>6</sup>. Therefore, preventive and precise diagnosis of CAD is crucial for reducing mortality rates. Early detection is critical to avoid an increased cardiovascular risk. The overall contribution of this research is to develop a non-invasive and efficient method for screening of CAD subjects while utilizing ABP and PPG measurements. In this context, selected state-of-the art works on detection of CAD are reviewed below.

ECG signal has been readily used for detection of cardiac conditions, as reported in the literature. Al. Hosani et al. has used a linear feature extraction approach with KNN, DT(Decision Tree), and SVM classifiers to detect CAD with maximum accuracy of 90%<sup>7</sup>. U. R. Acharya et al. utilized entropy feature extraction technique with the same set of classifiers<sup>8</sup>. V. Jahmunah et al. have utilized convolutional neural networks (CNN) for CAD detection using ECG signals<sup>9</sup>. Advanced machine learning approaches, such as deep neural networks and reinforcement learning, have demonstrated significant potential in analyzing and predicting physiological signals like ECG for cardiac abnormalities<sup>10</sup>. Such techniques can address challenges like missing data segments, providing robust signal predictions and enabling accurate feature extraction for clinical applications. In a similar study using ECG signal, the authors analyzed 303 records, examining 26 features that were selected as target features with input from clinical experts. They tested several key classification algorithms such as Multilayer Perceptron (MLP), Support Vector Machine (SVM), Logistic Regression (LR), J48, Random Forest (RF), K-Nearest Neighbors (KNN), and Naive Bayes (NB)<sup>11</sup>. The accuracy level of ECG-based models has been quite good, but preventive healthcare seems to be difficult in this case, as discussed earlier. Khan et al.<sup>12</sup> utilized PCG signal for CAD-detection and revealed 89% of accuracy, but PCG signals can easily be subjected to noise artefacts which makes the data collection process more complicated and difficult.

Now, PPG based models overcome the limitations of ECG and PCG based models for CAD detection<sup>7,13</sup>. Banerjee R et al. extracted time domain as well as frequency domain features from PPG signal to detect CAD with sensitivity of 73% and specificity of 87% while employing SVM classifier<sup>14</sup>. Paradkar et al. used temporal positioning of systolic and diastolic phases and characteristic points; the sensitivity of 85% and specificity of 78% for detection of CAD have been reported with SVM classifier<sup>15</sup>. Ihsan et al. used three methods for feature extraction respiratory rate (RR) interval, heart rate variability (HRV) features, and time domain features where each technique is used seperately which questions the robustness of the method<sup>16</sup>. Chakraborty et al. utilized statistical features, frequency domain features and time domain features and inter beat variability analysis techniques along with SVM, KNN, DT and LR (Logistic Regression) classifiers for CAD detection, revealing accuracy of 93%<sup>17</sup> but used photo-plethysmographic data, which questions about the robustness of the model.

In some of the recent works, two or more cardiovascular signals have been synchronously used to identify markers of CAD. Advances in signal processing for real-time applications have enabled efficient handling of ECG and PPG data through innovative compression techniques. For instance, Banerjee et al. proposed a real-time lossless compression algorithm for ECG and PPG signals using second-order delta encoding and runlength encoding (RLE), achieving high compression ratios while preserving critical signal features<sup>18</sup>. Banerjee et al. have worked on such a model in which time domain and frequency domain features are extracted from PPG and PCG signals together; the accuracy is reported to be 79%<sup>19</sup>. Zhao et al. demonstrated the effectiveness of VCG and CDG features with machine learning for detecting coronary microvascular dysfunction<sup>20</sup>. CAD progression involves hemodynamic changes, particularly in multi-vessel disease, affecting myocardial perfusion. Harmouche et al. emphasized the importance of collateral circulation in preserving heart function<sup>21</sup>. Our ABP and PPG-based framework builds on these insights to enable non-invasive early CAD detection. To address the aforementioned challenges we propose a novel Integrated Fusion Module which combines the features extracted from both the signals for better classification results and clinical applicability of the proposed framework towards CAD detection. In the proposed work, ABP and PPG signals are together utilized along with state-of-the-art classifiers to identify markers of CAD. The key contributions of this study are as follows:

- A set of Nineteen clinically significant features are extracted from both the ABP and PPG signals.
- An Integrated Fusion Module(IFM) designed for accurate classification. It combines the features extracted from the ABP and PPG signals and fed to the Classification Module.
- Quantitative and qualitative analysis of the predictive probability of the state-of-the-art classification networks for detection of CVDs.
- The proposed model demonstrates reliable and robust performance, validated by a large-scale dataset from the MIMIC-II database.

#### Methodology

The proposed Machine Learning framework as depicted in Fig. 1 illustrates the framework for the integrated fusion-based analysis of ABP and PPG signals, aimed at early detection of CAD. The framework consists of three primary stages: pre-processing, the Integrated Fusion Module (IFM), and the classification module. Each stage contributes to optimizing the diagnostic potential of the fused signal data.

In the pre-processing stage, the raw ABP and PPG signals are subjected to noise removal and filtering to eliminate unwanted artifacts and enhance signal clarity. This ensures that only the most reliable and relevant signal components are used for further processing. Once cleaned, these signals are forwarded to the IFM. In the Integrated Fusion Module, feature extraction is performed separately on the ABP and PPG signals, yielding 19 features from each signal. These features capture key physiological characteristics, such as pulse waveforms, amplitudes, and variability, which are crucial for identifying CAD-related abnormalities. The extracted features from both signals are then fused into an integrated feature set, leveraging the complementary nature of ABP and PPG to provide a comprehensive representation of the cardiovascular state.

Finally, the classification module utilizes the integrated feature set to train and test machine learning classifiers, including KNN, SVM, and neural networks. The classifiers are trained using a portion of the dataset and then tested to predict whether a subject falls into the CAD or non-CAD category. This stage evaluates the



**Fig. 1**. An overview of the Machine Learning Analysis of Integrated ABP and PPG Signals. This framework includes the Integrated Fusion Module (IFM) which combines the features from both PPG and ABP signals for better predictive analysis of CAD.

diagnostic performance of the fusion-based approach, assessing metrics like accuracy, sensitivity, specificity, and positive predictive value. By combining ABP and PPG signals, the framework enhances diagnostic accuracy and reliability, offering a non-invasive and efficient tool for early CAD detection. This integration emphasizes the potential of multi-signal fusion to improve the identification of critical cardiovascular conditions.

## Dataset

ABP and PPG signals have been analysed from 137 subjects out of which 73 suffer from CAD, and 64 do not suffer from any cardiac condition (labeled as non-CAD). Each of the analysed signals have a sampling rate of 125 Hz. The source of this data is MIMIC-II<sup>22</sup> waveform database with matched subset, was obtained from PhysioBank ATM<sup>23</sup>. Figure 2 shows the sample of data while showcasing some of the features.

## Pre-processing

The data from MIMIC-II waveform database requires certain pre-processing in order to remove unwanted noise and motion artifacts. A body at rest has a low respiratory rate, but a person sprinting may have a very high respiratory rate<sup>24</sup>. This rise and fall in respiratory rate might result in low frequency disruptions in the recorded signal. Even the smallest action of a human muscle can result in low frequency sounds<sup>25,26</sup>. Therefore, it's essential to keep the subject motionless while recording. The raw signal must be cleaned up since it might take up low-frequency noise during long recordings. Further, motion artifacts and distorted signals were eliminated subsequently. For ABP signals, a fourth order Butterworth band pass filter with a frequency range of 1.5–16 Hz and for PPG signals, a sixth order Butterworth band pass filter with a frequency range of 2.5–16 Hz are utilised. Figure 3 shows raw signal sample before and after filtering; here boxes indicate the affected area.

#### **Feature extraction**

In total, 19 features are extracted from each of ABP as well as PPG signals (Each feature equation presented in Supplementary Table S1 and in Figure S1). Table 1 shows average values of the features extracted from ABP and PPG for CAD as well as non-CAD subjects. Firstly, Heart Rate Variability (HRV) of the signal is used to extract relevant time domain features. HRV is a neuro-cardiac function of the brain and heart signals that contain changes or variations in the heartbeat<sup>27</sup>. Further, HRV is split into its components VLF, LF, and HF rhythms that operate within distinct frequency ranges; these are listed as frequency domain features<sup>27</sup>. The power spectrum of the HRV signal is divided into three bands: Very Low Frequency (VLF, 0–0.04 Hz), Low Frequency (LF, 0.04–0.15 Hz), and High Frequency (HF, 0.04–0.30 Hz) (HF, 0.15–0.4 Hz), which are also listed in Table 1.



**Fig. 2**. Basic ABP Signal with some features (Tc = Total Time for one Cycle, Ts = Systolic Time, Td = Diastolic Time).





## Integrated fusion module (IFM)

ABP measurements, which include systolic and diastolic pressures are useful for detecting hypertension a major risk factor for CAD. Additionally, the analysis of ABP waveform characteristics provides insights into arterial compliance and resistance, which are critical for diagnosing CAD. Meanwhile, PPG signals are employed to calculate pulse wave velocity (PWV), an indicator of arterial stiffness and a predictor of CAD. In the IFM we have combines the features extracted from both the signals. Nineteen features from both PPG and ABP signals are extracted which are of important biomarkers for detecting CAD. To enhance the classification accuracy of the signals features of the fused ABP and PPG signlas are used for training and testing.

#### Classification

Three state-of-the-art classifiers, namely, K-nearest neighbours (KNN), Support Vector Machines (SVM) and Neural Networks (NN) have been utilized for CAD/non-CAD classification. Ten-fold cross-validation technique has been used to validate the models. This kind of validation protects against over-fitting by partitioning the data set into folds and estimating accuracy on each fold. In total 70% of data were used for training, for validation 15% of data were used and for testing also 15% data were employed. In case of KNN, weighted KNN with 10 Neighbours has been used and for distance matric, the Euclidean method has been utilized. In case of SVM, Fine Gaussian SVM has been utilized. In case of Neural Network, 100 iterations has been utilized.

# Results and discussion

# **Experimental results**

The classification results are tabulated in Table 2. Performance parameters such as accuracy, sensitivity, specificity and PPV for each of the classifiers and for each of the models have been summarized in Table 2. Also, Confusion Matrix and ROC (Receiver Operating Characteristic) curves for each of the models are presented in Fig. 4. The

|        |  | PPG range |         | ABP range |         |
|--------|--|-----------|---------|-----------|---------|
| S. no. | Feature  | CAD       | Non-CAD | CAD       | Non-CAD |
| 1      | Mean of pulse duration                                   | 0.810     | 0.777   | 0.754     | 0.741   |
| 2      | Standard deviation (SD) of pulse duration                | 0.187     | 0.139   | 0.061     | 0.047   |
| 3      | Mean of relative crest time                              |           | 0.025   | 0.024     | 0.023   |
| 4      | SD of relative crest time                                | 0.005     | 0.003   | 0.002     | 0.001   |
| 5      | Mean of relative diastolic time                          |           | 0.005   | 0.055     | 0.056   |
| 6      | SD of relative diastolic time                            |           | 0.003   | 0.002     | 0.016   |
| 7      | Mean of time ratio                                       | 0.019     | 0.019   | 0.020     | 0.020   |
| 8      | SD of time ratio   | 0.069     | 0.065   | 0.027     | 0.025   |
| 9      | Mean of pulse height                                     | 1.098     | 4.232   | 2.241     | 2.786   |
| 10     | Mean of rising slope                                     | 0.041     | 0.148   | 0.083     | 0.107   |
| 11     | Mean of falling slope                                    | 0.007     | 0.025   | 0.012     | 0.014   |
| 12     | SD of normal peak-to-peak intervals                      | 0.063     | 1.113   | 0.491     | 0.379   |
| 13     | SD of normal diastolic peak intervals                    | 0.110     | 1.118   | 0.494     | 0.383   |
| 14     | Root mean square of Successive Deviations Between Normal | 0.093     | 1.580   | 0.663     | 0.521   |
| 15     | SD of successive deviation                               | 0.096     | 1.673   | 0.692     | 0.548   |
| 16     | Very low frequency                                       | 0.192     | 0.115   | 0.776     | 0.112   |
| 17     | Low frequency  | 0.136     | 0.121   | 0.761     | 0.117   |
| 18     | High frequency   | 0.014     | 0.020   | 0.013     | 0.017   |
| 19     | Ratio of LF and HF                                       | 9.470     | 6.840   | 7.434     | 7.860   |

Table 1. Extracted features and their range (for ABP and PPG signals for both CAD and Non-CAD).

| S. no. | Model       | Classifier | Accuracy (%) | Sensitivity (%) | Specificity (%) | PPV (%) |
|--------|-------------|------------|--------------|-----------------|-----------------|---------|
| 1      | PPG         | KNN        | 84.6         | 70.5            | 84.0            | 80.92   |
|        |             | SVM        | 83.5         | 65.8            | 80.9            | 75.5    |
|        |             | NN         | 82.0         | 69.7            | 82.8            | 80.0    |
| 2      | ABP         | KNN        | 91.5         | 78.0            | 91.1            | 89.5    |
|        |             | SVM        | 88.8         | 72.1            | 86.2            | 82.8    |
|        |             | NN         | 89.1         | 76.2            | 89.1            | 87.37   |
| 3      | Fused Model | KNN        | 89.5         | 74.7            | 88.2            | 85.8    |
|        |             | SVM        | 88.2         | 70.5            | 85.0            | 80.9    |
|        |             | NN         | 90.2         | 78.0            | 90.9            | 89.5    |

Table 2. Performance comparison of different classifiers using PPG, ABP, and the fused model.

empirical formulae for Accuracy (Acc Eq. 1), Sensitivity (Se Eq. 2), Specificity (Sp Eq. 3) and Positive Predictive Value (PPV Eq. 4) are detailed below<sup>28–30</sup>.

• Accuracy (Acc) The proportion of correctly classified cases (both CAD and non-CAD) to the total number of cases. Which Indicates the overall effectiveness of the classification model.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \tag{1}$$

where:

- TP (True Positives): CAD cases correctly identified as CAD.
- TN (True Negatives): Non-CAD cases correctly identified as non-CAD.
- FP (False Positives): Non-CAD cases incorrectly identified as CAD.
- FN (False Negatives): CAD cases incorrectly identified as non-CAD.
- Sensitivity (Se) The ability of the model to correctly identify CAD cases. Measures how well the model identifies actual CAD cases, minimizing missed diagnoses.

$$Se = \frac{TP}{TN + FN} \times 100 \tag{2}$$



**Fig. 4**. Confusion Matrix and ROC Curve of best model from ABP(Model 1), PPG(Model 2) and Fused Model(Modle 3). Where "0" represents NON-CAD and "1" represents CAD patients.

• **Specificity (Sp)** The ability of the model to correctly identify non-CAD cases. Measures the model's effectiveness in avoiding false alarms by correctly classifying non-CAD cases.

$$Sp = \frac{TP}{TN + FP} \times 100 \tag{3}$$

• **Positive predictive value (PPV)** The proportion of positive predictions (CAD cases) that are correctly classified. Reflects the reliability of positive predictions and reduces false positives in the CAD diagnosis.

$$PPV = \frac{TP}{TP + FP} \times 100 \tag{4}$$

- Receiver operating characteristic (ROC) curve A graphical plot that illustrates the trade-off between the True Positive Rate (Sensitivity) and the False Positive Rate (1 Specificity) across different thresholds is shown in Fig. 4. ROC Evaluates the classifier's performance irrespective of the decision threshold. In ROC higher area under the ROC curve (Area under the curve) indicates better model performance.
- Confusion matrix In Fig. 4 confusion matrix is used to summarize the model's predictions, showing the distribution of TP, TN, FP, and FN. Provides a comprehensive view of the classification performance, helping identify where the model excels or struggles. Where "0" represents NON-CAD patients and "1" represents CAD patients.

# Discussion

In this study, several machine learning models were trained and evaluated to classify Coronary Artery Disease (CAD) and non-CAD cases using arterial blood pressure (ABP) and photoplethysmogram (PPG) signals. The results demonstrate that the Neural Network (NN) classifier, utilizing features from both ABP and PPG signals through a fused dataset, outperformed other models in terms of accuracy, sensitivity, and specificity. The NN model achieved the highest classification accuracy of 90.2%, surpassing K-Nearest Neighbors (KNN) and Support Vector Machines (SVM), which also showed competitive but lower performance metrics. Among the single-signal models, ABP-based models performed better than PPG-based models, yet both were significantly improved by the integrated fusion approach.

The Receiver Operating Characteristic (ROC) curves provided further validation of the model's performance, clearly demonstrating the superior ability of the fused model to distinguish between CAD and NON-CAD cases.





The confusion matrix for the NN classifier revealed a balanced trade-off between true positive and true negative rates, further highlighting the robustness of the fused dataset in addressing classification challenges.

In contrast, models trained on single-signal datasets lacked the discriminative power provided by the fused approach. The results underscore the importance of combining complementary physiological signals for robust and reliable classification. This study provides strong evidence that the proposed fusion-based methodology, validated on a large-scale dataset, holds significant potential for early, non-invasive CAD detection and demonstrates practical applicability for real-world clinical use.

#### **Clinical relevance**

The clinical relevance of this work is significant, as it aims to detect CAD at an early stage using radial ABP and PPG signals. Early identification of CAD is critical for initiating timely clinical interventions that can prevent the disease from progressing to more severe stages. By utilizing these non-invasive and readily accessible monitoring techniques, the approach can facilitate regular screening and monitoring, making early detection more feasible and widespread. This can lead to prompt medical responses, such as lifestyle modifications, medications, and other therapeutic measures, which are crucial for improving patient outcomes and reducing the incidence of heart attacks. Additionally, early detection through ABP and PPG signals can improve patient compliance and follow-up, as these methods are less intimidating and more comfortable for patients compared to traditional techniques. Consequently, integrating ABP and PPG signals for early CAD detection enhances the classification results significantly which in-turn holds promising impact on preventive healthcare and reducing the overall risk of cardiovascular diseases.

## Conclusion and future work

Coronary artery disease is one of the most prevalent diseases and kills millions of people annually. In order to treat this condition and thereby reduce mortality rates, it is crucial to detect it early. To achieve this objective, arterial blood pressure and photoplethysmogram data have been collected for 137 subjects from MIMIC-II waveform database. It is concluded that Neural Network reveals best results with 90% prediction accuracy, 93.3% for patients with CAD, and 89.3% for patients without CAD. Despite the above innovation our study has several limitations. A comparison shown in Fig. 5 and Table 3 can demonstrate the effectiveness of our proposed method over other techniques used. Firstly, validation using external datasets from different institutions or equipment was not conducted. Secondly, the proposed method could not be compared with previous methods within the same analytical environment. Additionally, in this study we have extracted the features manually further the use of deep neural networks for classification on CAD can be implemented. For future work, enhancements can focus on improving the generalization of the dataset and incorporating deep learning methodologies. Additionally, a more comprehensive analysis of other potential features could be explored to further refine the model's performance.

| Model type                 | S. no. | Feature extraction techniques                                    | Classifier used  | Results     |
|----------------------------|--------|--|------------------|-------------|
|                            | 1      | Time + Frequency <sup>14</sup>                                   | SVM, KNN, DT     | Se = 73%%   |
|                            |        | Time + Prequency   |                  | Sp = 87     |
|                            | 2      | Temporal Position of Systelic and Diastolic Phases <sup>15</sup> | SVM              | Se = 85%,   |
| PPG                        |        | remporar rostion or systeme and Diastone r hases                 |                  | Sp = 78%    |
|                            | 3      | Peak + Notch + Valley Features <sup>24</sup>                     | SVM, KNN, DT     | Acc = 72%   |
|                            |        |  |                  | Se = 62%    |
|                            |        |  |                  | Sp = 75%    |
|                            | 1      | Time - Frequency <sup>19</sup>                                   | SVM              | Se = 60%    |
|                            |        | Time + Prequency   |                  | Sp = 93%    |
| PPG and PCG                | 2      | Time + Frequency <sup>31</sup>                                   | SVM              | Acc = 79%   |
|                            |        |  |                  | Se = 80%    |
|                            |        |  |                  | Sp = 70%    |
|                            | 1      | Linear Feature Extraction7                                       | KNN DT SVM       | Acc = 90%   |
| FCC                        |        |  | KININ, DI, SVIVI | Sp = 75%    |
| ECG                        | 2      | Linear Feature Extraction <sup>9</sup>                           | Gabor CNN        | Acc = 99%   |
|                            | 3      | Entropy Features <sup>8</sup>                                    | SCM, DT, KNN     | Acc = 99%   |
|                            | ) 1    | Time + Frequency + Shape   | SVM, KNN, NN     | Acc = 90.2% |
| Proposed Model (PPG + ABP) |        |  |                  | Se = 78%    |
|                            |        |  |                  | Sp = 90.9%  |

 Table 3. Comparison of different state-of-the-art models with the proposed model.

#### Data availability

The datasets generated and analysed during the current study are available in the The MIMIC II Waveform Database repository, https://archive.physionet.org/physiobank/database/mimic2wdb/.

Received: 5 September 2024; Accepted: 6 March 2025 Published online: 12 March 2025

#### References

- 1. Organization, W. H. et al. Cardiovascular diseases (cvds). *Media centre* (World Health Organization, 2021) https://www.who.int/n ews-room/fact-sheets/detail/noncommunicable-diseases.
- 2. Taki, A., Kermani, A., Ranjbarnavazi, S. M. & Pourmodheji, A. Overview of Different Medical Imaging Techniques for the Identification of Coronary Atherosclerotic Plaques 1st edn. (Elsevier, Amsterdam, 2017).
- 3. Willerson, J. T. & Holmes Jr., D. R., Coronary Artery Disease / edited by James T. Willerson, David R. Holmes, Jr. Cardiovascular Medicine, 1st ed. (Springer London, London, 2015).
- Chang, V., Bhavani, V. R., Xu, A. Q. & Hossain, M. An artificial intelligence model for heart disease detection using machine learning algorithms. *Healthc. Anal.* 2, 100016. https://doi.org/10.1016/j.health.2022.100016 (2022).
- Jain, K., Maka, S. & Patra, A. Modeling of cardiovascular circulation for the early detection of coronary arterial blockage. *Math. Biosci.* 304, 79–88 (2018).
- 6. Bonow, R. O., Mann, D. L., Zipes, D. P. & Libby, P. Braunwald's Heart Disease e-book: A Textbook of Cardiovascular Medicine (Elsevier Health Sciences, Amsterdam, 2011).
- Alhosani, A. et al. Automatic detection of coronary artery disease (CAD) in an ECG signal. In *ICECS 2017 24th IEEE International* Conference on Electronics, Circuits and Systems 2018-January, 414–418, https://doi.org/10.1109/ICECS.2017.8292036 (2018).
- Acharya, U. R. et al. Entropies for automated detection of coronary artery disease using ECG signals: A review. Biocybern. Biomed. Eng. 38, 373–384. https://doi.org/10.1016/j.bbe.2018.03.001 (2018).
- Jahmunah, V., Ng, E. Y., San, T. R. & Acharya, U. R. Automated detection of coronary artery disease, myocardial infarction and congestive heart failure using GaborCNN model with ECG signals. *Comput. Biol. Med.* 134, 4457. https://doi.org/10.1016/j.comp biomed.2021.104457 (2021).
- 10. Banerjee, S. & Singh, G. K. Deep neural network based missing data prediction of electrocardiogram signal using multiagent reinforcement learning. *Biomed. Signal Process. Control* 67, 102508 (2021).
- 11. Garavand, A. et al. Efficient model for coronary artery disease diagnosis: A comparative study of several machine learning algorithms. J. Healthc. Eng. 2022, 5359540 (2022).
- Khan, M. U., Aziz, S., Malik, A. & Imtiaz, M. A. Detection of myocardial infarction using Pulse Plethysmograph signals. In Proceedings - 2019 International Conference on Frontiers of Information Technology, FIT 2019 95–100, https://doi.org/10.1109/FIT 47737.2019.00027 (2019).
- Kumar, S. S. & Vijayalakshmi, K. Coronary Artery Disease Detection from PCG signals using Time Domain based Automutual Information and Spectral Features. In *Proceedings - 2020 International Conference on Computing, Electronics and Communications* Engineering, iCCECE 2020 69–74, https://doi.org/10.1109/iCCECE49321.2020.9231107 (2020).
- Banerjee, R., Choudhury, A. D., Vempada, R., Pal, A. & Mandana, K. M. Identifying coronary artery disease from photoplethysmogram. In UbiComp 2016 Adjunct - Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing 1084–1088, https://doi.org/10.1145/2968219.2972712 (2016).
- Paradkar, N. & Roy Chowdhury, S. Coronary artery disease detection using photoplethysmography. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 100–103, https://doi.org/10.1109/EMBC. 2017.8036772 (2017).
- Ihsan, M. F., Mandala, S. & Pramudyo, M. Study of feature extraction algorithms on photoplethysmography (ppg) signals to detect coronary heart disease. In 2022 International Conference on Data Science and Its Applications (ICoDSA), 300–304 (IEEE, 2022).

- Chakraborty, A., Sadhukhan, D., Pal, S. & Mitra, M. Automated myocardial infarction identification based on interbeat variability analysis of the photoplethysmographic data. *Biomed. Signal Process. Control* 57, 101747. https://doi.org/10.1016/j.bspc.2019.101747 (2020).
- Banerjee, S. & Singh, G. K. A new real-time lossless data compression algorithm for ecg and ppg signals. *Biomed. Signal Process.* Control 79, 104127 (2023).
- Banerjee, R., Choudhury, A. D., Datta, S., Pal, A. & Mandana, K. M. Non invasive detection of coronary artery disease using PCG and PPG. In *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST* 181 LNICST, 241–252, https://doi.org/10.1007/978-3-319-49655-9\_32 (2017).
- 20. Zhao, X. et al. Early detection of coronary microvascular dysfunction using machine learning algorithm based on vectorcardiography and cardiodynamicsgram features. *IRBM* 44, 100805 (2023).
- 21. Harmouche, M., Maasrani, M., Verhoye, J.-P., Corbineau, H. & Drochon, A. Coronary three-vessel disease with occlusion of the right coronary artery: What are the most important factors that determine the right territory perfusion?. *IRBM* **35**, 149–157 (2014).
- 22. Saeed, M. et al. Multiparameter intelligent monitoring in intensive care ii: A public-access intensive care unit database. *Crit. Care Med.* **39**, 952–960 (2011).
- 23. Goldberger, A. L. et al. Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. *Circulation* **101**, e215–e220 (2000).
- 24. Banerjee, S., Sood, P., Ghose, S. & Das, P. P. Coronary artery disease classification from photoplethysmographic signals. In *Proceedings of the 4th International Conference on Medical and Health Informatics*, 246–251 (2020).
- Ihsan, M. F., Mandala, S. & Pramudyo, M. Study of feature extraction algorithms on Photoplethysmography (PPG) signals to detect coronary heart disease. In 2022 International Conference on Data Science and Its Applications, ICoDSA 2022 300–304, https://doi. org/10.1109/ICoDSA55874.2022.9862855 (2022).
- Paradkar, N. & Chowdhury, S. R. Cardiac arrhythmia detection using photoplethysmography. In 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 113–116 (IEEE, 2017).
- Shaffer, F. & Ginsberg, J. P. An overview of heart rate variability metrics and norms. Front. Public Health 5, 1–17. https://doi.org/1 0.3389/fpubh.2017.00258 (2017).
- Hossin, M. & Sulaiman, M. N. A review on evaluation metrics for data classification evaluations. Int. J. Data Min. Knowl. Manag. Process. 5, 1 (2015).
- 29. Fawcett, T. An introduction to roc analysis. Pattern Recognit. Lett. 27, 861-874 (2006).
- Luque, A., Carrasco, A., Martín, A. & de Las Heras, A. The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognit.* 91, 216–231 (2019).
- Choudhury, A. D., Banerjee, R., Pal, A. & Mandana, K. M. A fusion approach for non-invasive detection of coronary artery disease. PervasiveHealth: Pervasive Computing Technologies for Healthcare 217–220, https://doi.org/10.1145/3154862.3154871 (2017).

#### Author contributions

The authors confirm their contribution to the paper as follows: study conception, design, and implementation: A.M., K.J.; analysis and interpretation of results: A.M., S.C.P., K.J.; draft manuscript preparation: A.M., S.C.P., K.J.; All authors reviewed the results and approved the final version of the manuscript.

# Declarations

## **Competing interests**

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at https://doi.org/1 0.1038/s41598-025-93390-x.

Correspondence and requests for materials should be addressed to A.M.

Reprints and permissions information is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025