DATA ANALYTICS AND APPLIED MATHEMATICS
E-ISSN: 2773-4854
VOLUME 6, ISSUE 1, 2025, 53 – 64
DOI: https://doi.org/10.15282/daam.v6i1.12496

RESEARCH ARTICLE

Development of a deep learning model for prediction of cardiovascular disease

Mohd Syafiq Asyraf Suhaimi[1], Nor Azuana Ramli[1,2*] and Lilik Jamilatul Awalin[3]

[1]Centre for Mathematical Sciences, Universiti Malaysia Pahang Al-Sultan Abdullah, Lebuh Persiaran Tun Khalil Yaakob, 26300 Kuantan, Pahang, Malaysia
[2]Centre for Artificial Intelligence & Data Science, Universiti Malaysia Pahang, Al-Sultan Abdullah, Lebuh Persiaran Tun Khalil Yaakob, 26300 Kuantan, Pahang, Malaysia
[3]Department of Engineering, Faculty of Advanced Technology and Multidicipline, Universitas Airlangga, Jl. Mulyorejo, Surabaya 60115, Indonesia

ABSTRACT - 15.1% of medically certified deaths in 2023 were due to ischemic heart disease (IHD), according to Department of Statistics Malaysia (DOSM) statistics on causes of death in Malaysia published in October 2024. Despite the slight decline, IHD remains a significant health concern in Malaysia, especially among males and individuals aged 41–59 years, where it accounted for 19.8% of deaths in that age group. Regular checks are one approach to preventing heart disease in its early stages; however, they can be expensive and time-consuming. With the advancement of technology, people can now conveniently check their blood pressure, heart rate, and electrocardiogram (ECG) using smartwatches. However, since some people lead busy lives and occasionally forget to track or monitor their health through the applications, monitoring alone is insufficient. The primary goal of this research was to develop a deep learning model for predicting cardiovascular disease (CVD) using data from smartwatches, which offer non-invasive and real-time health monitoring capabilities. The research employs two deep learning techniques: Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks. ECG and heart rate data were collected from 20 volunteers using local smartwatches supplemented with a publicly available dataset from Kaggle. Data pre-processing involved denoising ECG signals and normalising heart rate readings to ensure accuracy and reliability. The models were evaluated using precision, recall, F1-score, and accuracy metrics, achieving over 99 per cent across all measures. While both models demonstrated high predictive power, the LSTM model outperformed the CNN in computational efficiency, completing model training in 31 minutes compared to 87 minutes for the CNN. The study highlights the potential for wearable devices for real-time CVD monitoring and early diagnosis. Future work will explore the inclusion of additional data sources and advanced modelling involving ensemble techniques.

ARTICLE HISTORY

Received	:	5th Nov 2024
Revised	:	21st Feb 2025
Accepted	:	15th Mac 2025
Published	:	31st Mac 2025

KEYWORDS
Deep learning
Cardiovascular disease
Long short-term memory
Convolutional neural networks
Predictive modelling

1. INTRODUCTION

The health issues brought on by chronic illnesses get worse as the current civilization ages. Ischemic heart disease (IHD) was the second-highest cause of death in Malaysia in 2023, accounting for 15.1% of medically certified fatalities. Pneumonia was the principal cause of death in 2023, accounting for 18,181 deaths or 15.2 per cent of the medically certified deaths [1]. IHD remains a significant health concern in Malaysia, particularly among males and individuals aged 41–59 years, accounting for 19.8% of deaths in this age group despite a declining trend [1]. The absolute risk of cardiovascular disease increases with age; however, the risk factors for cardiovascular disease vary by age group. Since physiological risk factors, such as lipid metabolism and vascular health, have a limited impact on cardiovascular disease in young people, its occurrence is lower in this age group compared to other age groups.

Physiological factors gradually alter the risk factors for cardiovascular disease in middle age compared to young individuals. However, it has been observed that these risk factors take some time to deteriorate sufficiently. Consequently, middle-aged individuals are categorized solely as a prospective risk category for cardiovascular disease and do not garner as much focus as the real danger group, which is the elderly [2]. Therefore, if middle-aged individuals, who are presently considered the at-risk population, persist in maintaining the unhealthy habits they have adopted since their younger years, they are highly likely to transition into a high-risk group for cardiovascular issues [3]. Hence, it is crucial to anticipate and alert individuals who are at risk of developing cardiovascular disease. A tool to predict the development of cardiovascular disease would be highly advantageous.

Anticipating cardiovascular illness is not a novel concept. By employing various data mining approaches, it is significantly more feasible to detect coronary heart disease at an early stage. Data mining aims to extract data from a specific dataset and transform it into a format that is understandable for further utilisation. To incite a reaction to the paintings, it is most effective to begin by collecting statistics on each component closely related to the system's objective.

***CORRESPONDING AUTHOR | N.A. Ramli** | ✉ azuana@umpsa.edu.my

The study focused on the fundamental causes and factors that influence cardiovascular health. Certain aspects, such as age, sex, and familial background, are unmodifiable. However, specific characteristics, such as blood pressure and heart rate, can be controlled using particular methods. The utilization of machine learning is experiencing significant growth in the healthcare industry, as it reduces human error through computational analysis and improves accuracy. Machine learning approaches provide highly accurate diagnoses of illnesses [4]. Machine learning classification algorithms, such as logistic regression, decision tree, Naïve Bayes, k-nearest neighbour (k-NN), support vector machine (SVM), artificial neural network (ANN), ensemble learning, and deep learning, can be used to predict various illnesses that are like heart disease, including liver disease, diabetes, and neoplasms. Deep learning algorithms are predominantly utilised in the realm of medical diagnosis for most tumour predictions [5].

Smartwatches can enhance daily health by facilitating self-monitoring of individual activity, providing feedback based on activity metrics, conducting on-site surveys to identify behaviour patterns, and enabling two-way communication with healthcare professionals and family members [6]. Favoriot, a local startup company, has already launched the Raqib Smartwatch, which can measure and record heart rate, blood pressure, and electrocardiogram (ECG). This smartwatch is primarily designed for hajj pilgrims. It can track the wearer's whereabouts, monitor their health, and facilitate continuous communication. Nevertheless, the Raqib smartwatches were employed in this research to record the participants' electrocardiogram (ECG), heart rate, blood pressure, and everyday activities. The data was obtained from the server and serves as the validation dataset for the predictive model constructed in this work, utilising deep learning techniques and the Kaggle dataset.

Deep learning is recognised for achieving higher levels of accuracy in recognition compared to other machine learning algorithms [7]. This study employed Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) techniques. LSTM is a recurrent neural network (RNN) capable of acquiring knowledge about long-term relationships. [8]. Someone initially presented them; many developed and popularised them in their future work. LSTM models have exceptional efficacy in addressing diverse problems and are now widely applied. LSTM is specifically engineered to mitigate the issue of prolonged dependence. Long-term memory is an automatic routine for the brain, requiring minimal effort to acquire.

This project aims to develop a deep learning model for predicting cardiovascular disease (CVD) utilising LSTM and CNN techniques. The proposed model was trained and tested using data from Kaggle. The validation was conducted using data collected from the Raqib smartwatch to assess its capability to predict the occurrence of cardiovascular disease. This work is organised as follows: the next section highlights the research gap between this study and earlier studies. The following section then details every method utilised in this study, followed by a discussion of all the findings and a brief conclusion with suggestions for future research.

2. LITERATURE REVIEW

The healthcare industry is trending toward the use of artificial intelligence and data analytics applications due to the increased volume of data and improved processing capabilities. Based on a study conducted by [9], there was a notable increase in the utilisation of data mining and machine learning in healthcare studies between 2003 and 2015, with 51.5% of the investigations employing data mining techniques and 39.3% utilising machine learning approaches. Numerous studies have focused on data mining and machine learning techniques in predicting diseases. Some research has also been conducted to predict cardiovascular disease. However, these studies predict cardiovascular disease using data features that can only be measured if individuals with medical conditions seek medical attention.

Traditionally, global cardiovascular disease risk assessment relies on clinical risk scores that calculate the risk over a ten-year period. Nevertheless, most of these scores fail to accurately represent the fluctuating variations in individualised risk that closely correspond to lifestyle patterns. Integrating subjective lifestyle behaviours into risk assessment has posed difficulties. Hence, utilising objective data obtained from wearables presents a fresh prospect for enhancing the precision, comprehensiveness, and dynamism of assessing the risk of cardiovascular disease throughout one's lifetime. Multiple studies have shown that physical activity, as evaluated by wearables, is inversely related to the risk of death from any cause. Triaxial accelerometers were used to quantify moderate-to-vigorous physical activity (MVPA), which was linked to a reduced mortality risk compared to light physical activity or sedentary behaviour. This association has been observed in numerous US cohorts and a population-based cohort in Sweden [10]. A separate investigation involving women with an average age of 72 years demonstrated that a mere 4,400 steps per day were notably linked to a 41% decrease in mortality compared to only 2,700 steps per day. However, the advantages seemed to plateau at 7,500 steps per day [11]. It is worth noting that the intensity of stepping did not correlate with mortality, even after accounting for the daily number of steps taken.

Wearable data can also enable the implementation of real-time behavioural change techniques (BCTs), such as just-in-time adaptive interventions. These interventions are designed to continuously evaluate user needs and deliver the right amount and type of intervention at the appropriate moment. Multiple experiments were conducted to assess the advantages of wearable-guided behaviour change technologies. One of the studies was the mActive trial, where 48 outpatients were recruited from an academic cardiovascular facility. The study participants had a prevalence rate of 50% for hypertension, 23% for diabetic mellitus, and 29% for coronary heart disease. The trial participants were randomly divided into two groups: one group was blinded to their Fitbug ORB activity, while the other group was unblinded. The participants were then examined in two phases. The first stage consisted of utilising solely the tracking device, whereas the second stage included employing intelligent texts with behaviour change techniques (BCTs) for the unblinded group. The intelligent communications were automated and tailored (coaching SMS messages created by the physician investigators and influenced by real-time data from the device). The messages were categorised into positive reinforcement messages, sent to participants who were either on track or had already achieved their daily target of 10,000 steps, and boosting messages aimed at motivating participants who were not on track to reach their goal. The text messaging group exhibited a daily increase of 2,534 steps compared to the unblinded group, which did not engage in texting, and a 3,376-step increase compared to the blindfolded control group [12].

A separate investigation by [13] included 110 individuals scheduled for vascular or cardiac treatments. Participants were given an iPhone and an Apple Watch, both equipped with the VascTrac study app. The participants were then monitored for a period of six months. The application passively gathered activity data, specifically focusing on daily step counts. Logistic regression was employed to evaluate the predictive ability of home 6MWT data and passive data in determining ‘frailty’, as measured by the supervised 6MWT gold standard. Frailty was defined as the inability to walk 300 meters or less on a hill during the 6-Minute Walk Test (6MWT). This study utilised longitudinal observation to examine the relationship between passive activity data acquired from an iPhone and an Apple Watch and the performance of incline walking in the 6-Minute Walk Test (6MWT). The findings revealed that the passive activity data accurately predicted the performance of incline walking in the 6MWT. This discovery implies that it may be possible to remotely monitor and evaluate the frailty and functioning of individuals with cardiovascular disease, offering hope for the future of patient care.

Based on an extensive literature review, this study distinguishes itself from previous research in terms of variables, approaches, methodologies, and experimental design. Specifically, this study concentrates on middle-aged individuals in Malaysia. The data obtained from smartwatches manufactured by a domestic Malaysian manufacturer were compared with the lifestyle of the volunteers. Additional distinctions between this investigation and prior research are illustrated in Table 1.

Table 1. Research gap between the previous studies and the present study

	Previous Studies	Present Study
Case Study	Most previous studies focus on CVD prediction from available open or closed-source data.	The present study uses data from a local smartwatch and the Kaggle dataset to predict heart disease.
Factors of Heart Disease in the Studies	Most previous studies were based on existing data, followed by the development of predictive modelling for CVD using machine learning. Not many researchers included factors that can contribute to CVD, such as hypertension, high blood cholesterol, smoking, obesity, diabetes, and physical inactivity, as well as other valuable information on the impact of related diseases in their studies. Some researchers also studied factors such as blood triglycerides and high-density lipoprotein (HDL) cholesterol levels, age, gender, and psychosocial problems.	Wearable technology, such as smartwatches, has revolutionised personal health monitoring, making it especially beneficial for tracking cardiovascular disease (CVD) factors, including electrocardiogram (ECG) data. As these devices can now measure ECG in real-time, they offer a convenient way for people, mainly working adults with busy schedules, to monitor their heart health without needing frequent visits to a medical facility. Monitoring heart rate variability, irregular rhythms, and other ECG data can alert users to potential issues early, helping manage CVD risk factors such as high blood pressure, arrhythmias, and abnormal heart rates.
Methodology	Several studies have applied deep learning, such as CNN and LSTM, to predict CVD. Some previous studies also employed a data collection method similar to the current study (dataset from an open-source source). The only difference lies in the smartwatch data, where other studies utilised existing smartwatches on the market, such as Fitbit and Samsung smartwatches, or their developed smartwatch.	This research aims to develop a deep learning model that utilises Convolutional Neural Networks (CNN) and Long-Short-Term Memory (LSTM) networks to predict cardiovascular disease using smartwatch data. To align with the main objective, this study was conducted in three major phases: data collection (Phase 1), predictive modelling (Phase 2), and performance evaluation (Phase 3).

3. METHODOLOGY

To align with the stated objectives, the project was executed in phases. The research framework for this study is illustrated in Figure 1.

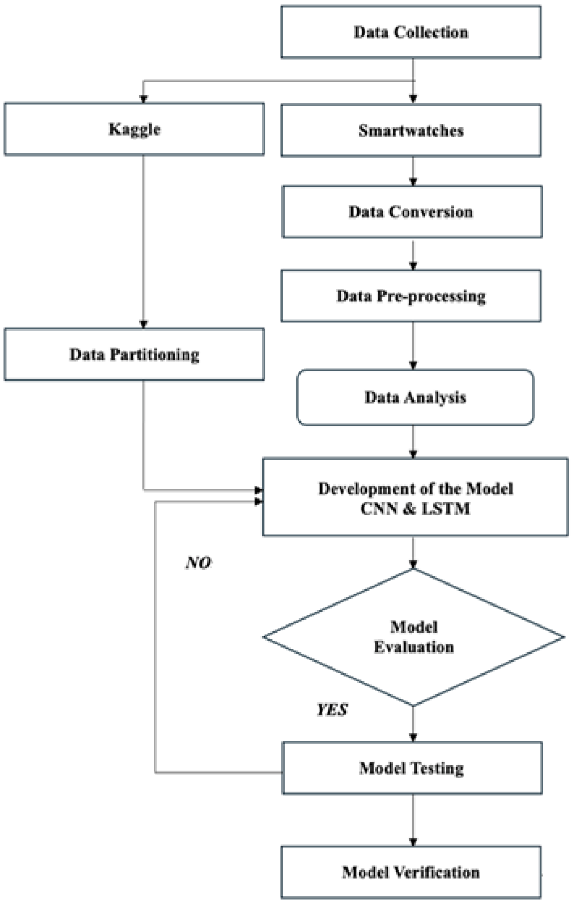


Figure 1. Research framework

3.1 Data collection

Data for this study was collected through Raqib smartwatches from Favoriot. Four smartwatches are used for our data collection purposes. In our experiment, four individuals were randomly selected as our weekly test subjects, ensuring a consistent and reliable data collection process. These individuals, who were not subject to any specific requirements other than being from a middle-aged group, wore smartwatches continuously for a week. After a week, a different person was tested, and this process was repeated for 6 months. All the datasets from the smartwatches are readily available on Raqib's web app, providing easy access to the data for further analysis. However, the datasets were visualised by day in line graph form, and another process needs to be performed to obtain the datasets in comma-separated values (CSV) format.

The data conversion process was straightforward and efficient, facilitated by the user-friendly Postman API platform. Initially presented in a line graph, the dataset was easily converted into a table, with all the parameters neatly organised. The datasets are in JSON format and converting them to CSV is a quick and straightforward process that can be done through websites such as CSV Online Converter. (https://www.convertcsv.com/json-to-csv.htm).

Table 2. Retrieve ECG raw data – Query Response (JSON)

Parameter	Description
timestamp	Time in EPOCH format
data	An array of ECG value data (used for plotting the ECG pattern)
heart_rate	Heart rate data
sdnn	SDNN value used for determining physiological age
lf	Low-frequency value used for determining fatigue
hf	High-frequency value used for determining fatigue
samplingFreq	Sampling frequency

The second dataset was acquired from Kaggle using the following URL: https://www.kaggle.com/datasets/shayanfazeli/heartbeat/data. The dataset was utilised to examine heartbeat classification through deep neural network structures and to evaluate specific transfer learning capabilities. The signals correlate with the electrocardiogram (ECG) waveforms of heartbeats in everyday situations and cases affected by different arrhythmias and myocardial infarction. The signals underwent preprocessing and segmentation, with each segment corresponding to a single heartbeat. The following information provides an overview of the dataset:

- The total number of samples is 109,446.

- The number of categories is five, and the sampling frequency is 125 Hz.
- The data source used is the MIT-BIH Arrhythmia Dataset from Physionet.

The data was sourced from Kaggle [14], as the data collected from the smartwatch alone is insufficient to generate a highly accurate predictive model. This dataset was used for model development.

3.2 Data pre-processing

Prior to utilising this data, it is essential to conduct data preprocessing. Data preprocessing, a data preparation component, encompasses any processing carried out on raw data to ready it for subsequent processing. Traditionally, it has been a crucial initial step in machine learning and deep learning. Data preprocessing involves transforming data into a more manageable format for machine learning, deep learning, and other data science tasks. These methods are frequently employed throughout the initial stages of machine learning and artificial intelligence development to obtain reliable results.

This study's preprocessing involved data cleaning, transformation, denoising, segmentation, and normalisation. The data preprocessing phase focuses more on the smartwatch dataset since the data obtained from Kaggle is already clean. From our raw data, it was found that the timestamp needed to be in the proper date format. In this case, the timestamp was in an epoch format, which can be converted to a readable date using the Epoch Converter website (https://www.epochconverter.com/). After the date conversion, the dataset was filtered by date, as different people generated different results. The analysis was conducted on an individual, and the primary analysis involved an ECG signal analysis, followed by assessments of heart rate, steps, blood pressure, and the person's lifestyle and behaviour.

In the case of ECG signals, three primary noise sources must be filtered out to achieve signal denoising. The reason for this is that ECG signals depict the electrical actions of the heart, and they are sometimes tainted by disruptions of both high and low frequencies. ECG signal denoising can be achieved by utilising the Discrete Wavelet Transform (DWT), which has proven to be an effective technique in numerous research investigations. Python was used to denoise ECG signals utilising the PyWavelets library package. Once the ECG signals have been filtered to remove noise, the dataset was divided into time windows of a specified length, each lasting 5 seconds. By dividing the ECG records into shorter intervals, a more significant amount of cardiac data can be utilised as training and testing data to input into the model. Finally, all the datasets were normalised using the MinMax scaling method, as it can improve model training stability, speed, and performance by ensuring consistent feature scaling. The formula for normalisation is shown in equation (1).

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{1}$$

3.3 Model development

Deep learning is a specific branch of machine learning that involves neural networks with more than two hidden layers. This study employed two renowned deep learning techniques: a Convolutional Neural Network (CNN) and a Long-Short Term Memory (LSTM) network, which is a type of recurrent neural network (RNN) architecture. The architecture of CNN closely resembles the neural network structure of the human brain. Like the brain, CNN consists of billions of neurons organised specifically. The neurons in a CNN are structured in a manner that closely resembles the organisation of the brain's frontal lobe, which is responsible for the processing of visual stimuli. This architecture guarantees that the entire visual field is encompassed, preventing the fragmented image processing challenge that conventional neural networks encounter when presented with low-resolution image segments. Compared to previous networks, CNN performs better when presented with visual inputs [15].

CNN comprises three distinct layers: convolutional, pooling, and fully connected (FC). The complexity of the CNN increases as it progresses from the convolutional layer to the FC layer. The increasing complexity of CNN allows it to detect more intricate portions of an image, ultimately identifying the entire object. A CNN can incorporate many layers trained to identify specific features within an input image. Each image undergoes a filter or kernel application, resulting in an output that progressively enhances and increases in detail with each layer. Filters in the lower layers may initially manifest as basic characteristics. Existing networks can be utilised to retrain and expand the capabilities of CNNs to perform novel recognition tasks. These advantages create opportunities for utilising CNN in practical applications without increasing computational complexity or cost. Due to parameter sharing, CNNs exhibit superior computing efficiency compared to conventional neural networks. Implementing the concept is straightforward and may be used on any device.

The vanishing gradient problem renders the deployment of RNNs in actual applications unfeasible, necessitating the introduction of LSTM to mitigate the issue of multiplying gradients less than zero. The LSTM model consists of an input gate, a forget gate, and an output gate, which are used to process the current network input, x_t, and the previous cell output, y_{t-1}. An input gate utilising the tanh activation function supplied by:

$$k = \tanh\left(b^{(k)} + x_t W_1^{(k)} + y_{t-1} W_2^{(k)}\right) \tag{2}$$

The input gate, represented as a sigmoid-activated node in the hidden layer, is indicated as

$$I = \sigma\left(b^{(1)} + x_t W_1^{(1)} + y_{t-1} W_2^{(1)}\right) \tag{3}$$

The input gate's output is

$$k \circ I \tag{4}$$

The inner state of the LSTM, referred to as o_t, is shifted by one time step and then combined with the equation mentioned before. This establishes an internal feedback loop that acquires knowledge about the correlation between the inputs provided at different instances. This stage comprises a forget gate, a node activated by the sigmoid function that determines which previous state to retain in memory. The term defines the forget gate:

$$F = \sigma\left(b^{(F)} + x_t W_1^{(F)} + y_{t-1} W_2^{(F)}\right) \tag{5}$$

This stage's output is

$$o_t = o_{t-1} {^\circ} F + k {^\circ} I \tag{6}$$

where o_{t-1} is the preceding cell's inner state.

The final phase of the LSTM model is the output gate, which is composed of a tanh squashing function and an output sigmoid function, as defined in equation (7):

$$L = \sigma\left(b^{(L)} + x_t W_1^{(L)} + y_{t-1} W_2^{(L)}\right) \tag{7}$$

As a result, the cell's output is provided by

$$y_t = \tanh(o_t) {^\circ} L \tag{8}$$

At each step, $b^{(i)}$, $W_1^{(i)}$, and $W_2^{(i)}$ represent the input bias, input weight, and weight of the preceding cell's output. Weights and biases are computed during the training period.

The model architecture in this study was constructed utilising the Keras library with the TensorFlow framework as the underlying technology. Keras is a versatile open-source deep learning library that may be deployed as a Python package. The CNN and LSTM models were constructed by defining the Sequential() model method.

3.4 Performance evaluation

Following the development of the models, their performance was assessed using precision, recall, and F1-score, which can be obtained through the confusion matrix and formula. Table 3 presents a confusion matrix that facilitates a detailed examination of the model's predictive accuracy and error distribution across different classes, providing critical insights for assessing the efficacy of the developed model.

Table 3. Confusion matrix

	Actual Class 1	Actual Class 2
Predicted Class 1	True Positive (TP)	False Positive (FP)
Predicted Class 2	False Negative (FN)	True Negative (TN)

True Positive (TP) refers to instances where the model correctly predicts the positive class. False Positive (FP) indicates cases where the model incorrectly predicts the positive class when the actual class is negative. False Negative (FN) signifies instances where the model incorrectly predicts the negative class when the actual class is positive. True Negative (TN) denotes cases where the model correctly predicts the negative class.

The formula for precision:

$$\text{Precision} = \text{True Positives} / (\text{True Positives} + \text{False Positives}) \tag{9}$$

The formula for the recall:

$$\text{Recall} = \text{True Positives} / (\text{True Positives} + \text{False Negatives}) \tag{10}$$

The formula used to calculate the F1-score:

$$\text{F1-score} = 2\ (\text{precision recall}) / (\text{precision} + \text{recall}) \tag{11}$$

Finally, the formula used to calculate accuracy:

$$\text{Accuracy}=(\text{TP}+\text{TN})/(\text{FP}+\text{FNTP}+\text{TN})\times 100\% \tag{12}$$

4. RESULTS AND DISCUSSION

This paper presents the classification outcomes of CNN and LSTM models on the Kaggle datasets. This section is divided into three parts. The first part presents the results gained from data preprocessing. Subsequently, Scikit-learn was employed to generate data training and validation sets, utilising a dataset sourced from Kaggle, before commencing model construction. The CNN and LSTM models were assessed using precision, recall, and F1-score at the end stage to determine the optimal model for further testing with smartwatch data.

This Kaggle dataset comprises two sets of heartbeat signals obtained from two renowned datasets in heartbeat classification: the MIT-BIH Arrhythmia Dataset and the PTB Diagnostic ECG Database. Both collections include ample samples to adequately train a deep neural network. The Arrhythmia Dataset has 109446 samples, divided into five distinct categories. The sampling frequency used for this dataset is 125 Hz. The PTB Diagnostic ECG Database contains 14,552 samples, classified as either normal or pathological. The samples were collected at a sampling frequency of 125hz.

4.1 Data Pre-processing

The data preprocessing stage mainly concentrates on the smartwatch dataset. Figure 2 displays the original plot of one of the smartwatch's data in Python.

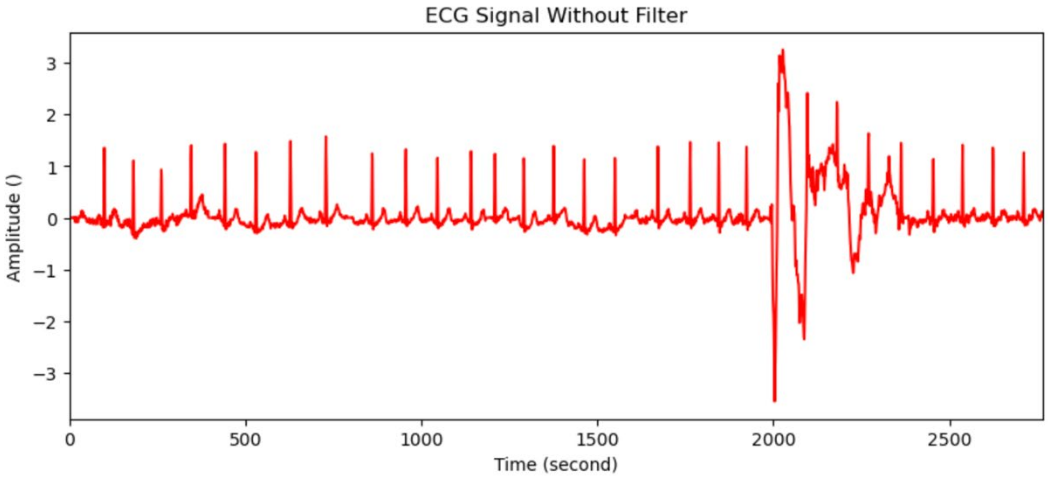


Figure 2. Smartwatch ECG signal without filter

Then, the median filter was applied to the ECG dataset to denoise it. Figure 3 shows the original ECG signal and the denoised signal (orange line), where the denoised signal effectively preserves the key morphological features of the original ECG waveform, such as the R-peaks and overall periodic structure, while significantly reducing high-frequency noise and small fluctuations. The original signal (blue line) exhibits sharp spikes and irregularities, likely caused by motion artefacts or electrical interference, whereas the filtered signal is smoother and cleaner. This demonstrates that the median filter successfully suppresses outliers without distorting the important cardiac features essential for further analysis, such as heartbeat detection or arrhythmia classification.

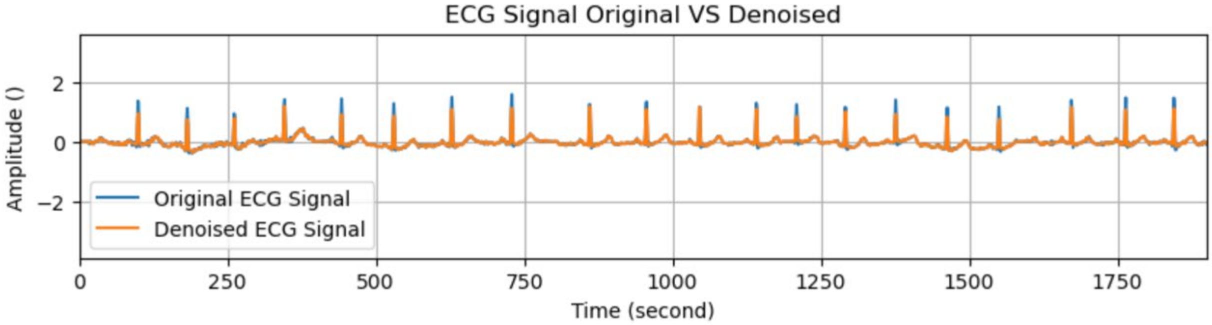


Figure 3. ECG signal original VS denoised

Lastly, the final denoised ECG signal is shown in Figure 4, where the ECG waveform appears clean and well-preserved after the median filter is applied. The R-peaks remain prominent and consistently spaced, indicating a stable heart rhythm, while high-frequency noise and minor fluctuations have been effectively suppressed. The baseline is relatively stable, and no significant signal distortion is observed, confirming that the denoising process successfully enhances signal quality while retaining the essential morphological features needed for accurate physiological analysis or further processing tasks such as heartbeat classification and arrhythmia detection.

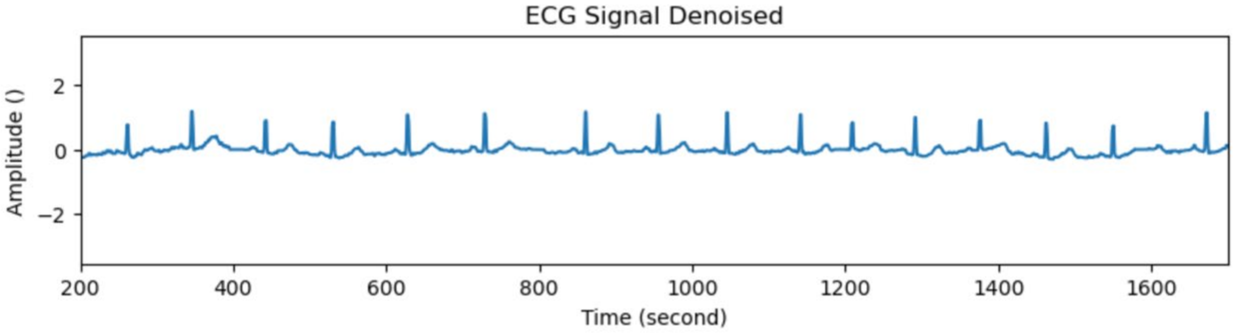


Figure 4. Denoised ECG signal

4.2 Model development

As stated earlier, models in this study were developed using the Kaggle dataset. This dataset was loaded into Pandas' DataFrame, and all the features needed to be converted to a float type. Once the conversion is done, a dictionary maps the target indices to class names. The class names are:

0: Normal
1: Atrial Premature
2: Premature ventricular contraction
3: Fusion of ventricular and normal
4: Fusion of paced and normal

Data normalisation can be performed once the input and target are in the correct shape. Following normalisation, the data was divided into training and validation sets using scikit-learn. The class imbalance of each class in both the training and validation sets was examined to ensure a balanced distribution of data. Due to the non-uniform distribution of classes, assigning weights to each class was necessary, which were then normalised.

The next step involved model development, which began with training the model using a CNN. Figure 5 illustrates the validation and training loss obtained from the CNN model training over 50 epochs. The model is not overfitting, as the line shows a decreasing trend and consistent improvement in generalisation performance. However, it is worth noting that model training for CNN is time-consuming, taking almost 1.5 hours, even when using a Mac Laptop with an Intel i5 2-core processor.

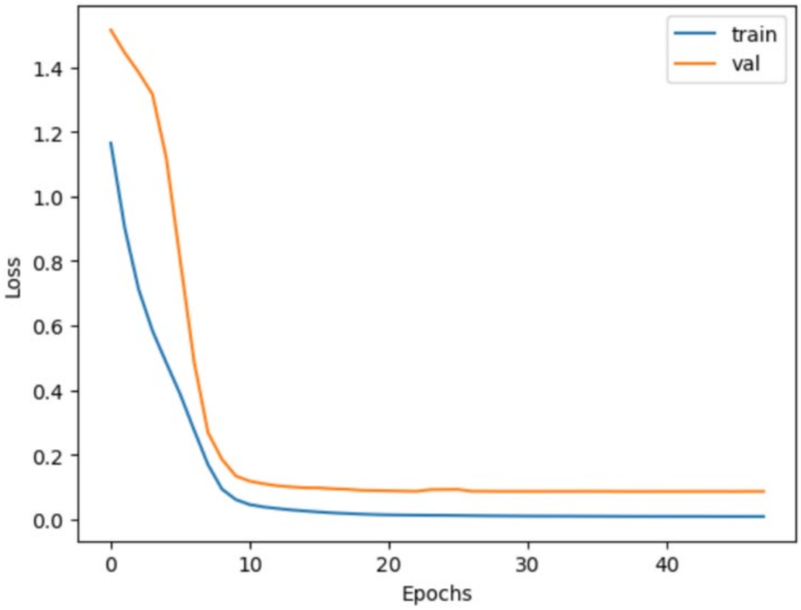
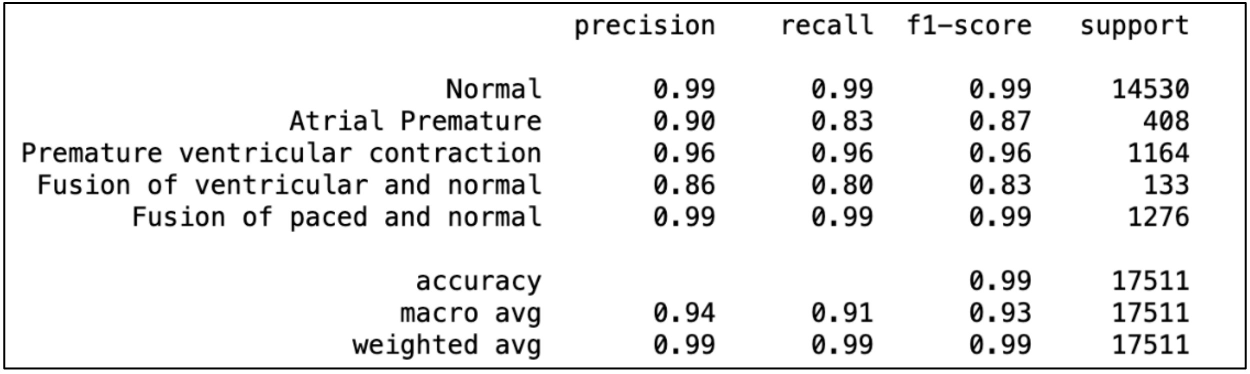


Figure 5. Training and validation loss for CNN

The model needs to be evaluated using a classification report with the validation set to ensure it does not miss any minor classes. The results are presented in Figure 6, where the model's accuracy is relatively high, at 99%.



```
                                   precision    recall  f1-score   support

                            Normal      0.99      0.99      0.99     14530
                  Atrial Premature      0.90      0.83      0.87       408
 Premature ventricular contraction      0.96      0.96      0.96      1164
  Fusion of ventricular and normal      0.86      0.80      0.83       133
        Fusion of paced and normal      0.99      0.99      0.99      1276

                          accuracy                          0.99     17511
                         macro avg      0.94      0.91      0.93     17511
                      weighted avg      0.99      0.99      0.99     17511
```

Figure 6. Classification report for CNN obtained from Python

Detailed results are presented in the confusion matrix in Figure 7, which illustrates the model's performance in classifying the five cardiac beat types. The model achieves high true positive rates for Normal (0.99), Premature Ventricular Contraction (0.96), and Fusion of Paced and Normal (0.99), indicating strong classification accuracy for these classes. However, Atrial Premature beats are misclassified as Normal in 16% of cases, and Fusion of Ventricular and Normal beats are misclassified as Normal (14%) and Premature Ventricular Contraction (6%), suggesting challenges in distinguishing between morphologically similar arrhythmias.

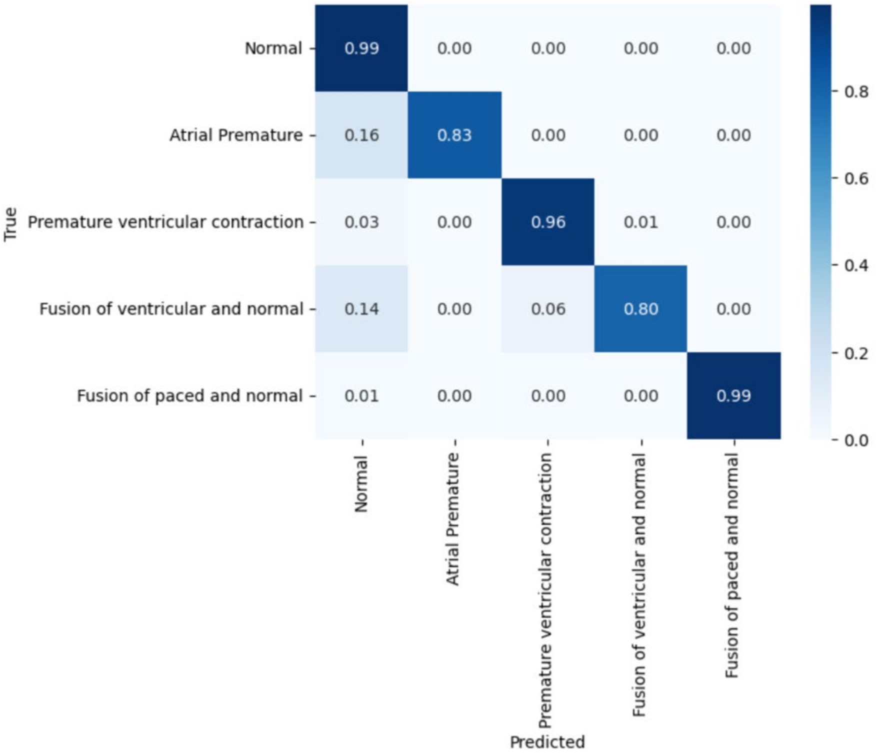
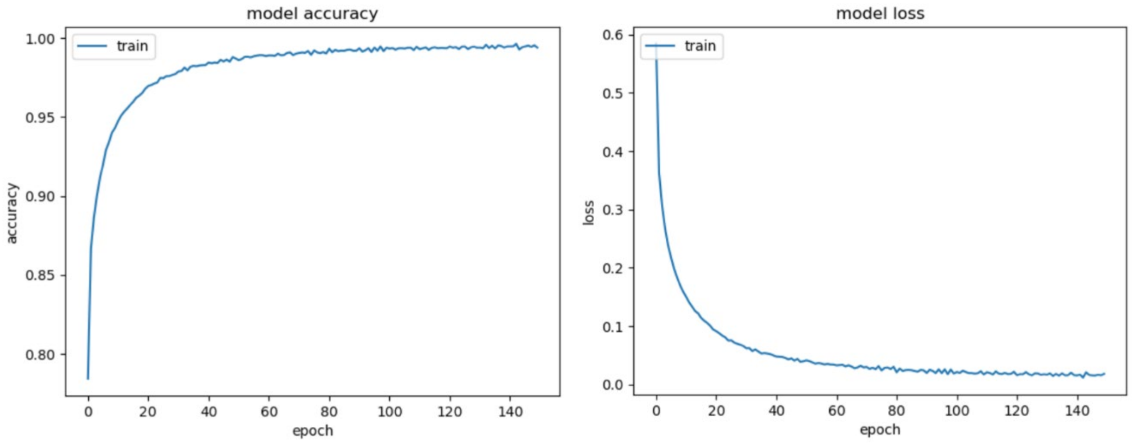


Figure 7. Confusion matrix for the CNN model.

When it comes to LSTM model development, it is essential to highlight its ability to avoid overfitting, a key factor contributing to its impressive performance. The training time for LSTM is significantly faster, taking only 20 minutes, compared to CNN. The process for model development using LSTM is similar to that of CNN, with the only notable difference being the epoch size, which was set at 150 for this model. Figure 8 presents the outcomes of the model's training, demonstrating a consistent increase in model accuracy and a simultaneous decrease in model loss. The model's ability to generalise is solid, with no signs of overfitting, further underlining its impressive performance.



(a) (b)

Figure 8. Model accuracy and model loss for LSTM

4.3 Performance evaluation

The overall performance of both CNN and LSTM is considered good, with accuracy scores of 98.62% and 98.63%, respectively. However, certain considerations, such as model complexity and the time required to train the model, must be taken into account. Regarding complexity, the development of the CNN model was more complex than that of the LSTM model. As for the time taken to complete a model training, more time is needed to run the CNN model, which, in this case, CNN uses 50 epochs only, but it took almost 1 hour and a half to complete a model training while LSTM needs 20 minutes to complete 150 epochs, and this can save a lot of time. Generally, not everyone has a high-performance computer to run a very complex system, so it is recommended to consider the LSTM model instead of the CNN, since both models provide good results with high accuracy.

Table 4. Comparison of the classification performance of the proposed method with previous studies

Methods	Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Sowmya et al., 2022 [16]	CNN	96.20	93.70	95.10	N/A
Sowmya et al., 2022 [16]	CNN- LSTM	99.50	91.60	99.20	N/A
Mehmood et al., 2020 [17]	CNN	97.00	97.06	96.35	96.70
Proposed 1	CNN	98.62	99.00	99.00	99.00
Proposed 2	LSTM	98.63	99.00	99.00	99.00

In Table 4, the minimal classification gap indicates that all these classifiers are reliable for prediction with over 95% accuracy. However, most research does not specify the time to complete model training with these classifiers.

4.4 Model testing

Both models were utilised during the testing phase to predict cardiovascular disease (CVD) using data derived from smartwatches. The testing process was implemented in Python, leveraging the Torch library for model evaluation. Prior to prediction, the models were properly set to evaluation mode and instantiated with the appropriate input and output dimensions, ensuring reliable performance. The predicted class labels correspond to five categories based on the Kaggle dataset classification: Class 0 – Normal, Class 1 – Atrial Premature, Class 2 – Premature Ventricular Contraction, Class 3 – Fusion of Ventricular and Normal, and Class 4 – Fusion of Paced and Normal. It is essential to note that only Classes 0 through 4 are represented in the system output; therefore, class identifiers should be consistently referred to as Class 0 through Class 4. For enhanced interpretability, the prediction results can be visualised in various formats. In this instance, a histogram is employed to illustrate the distribution of predicted classes, as shown in Figure 9.

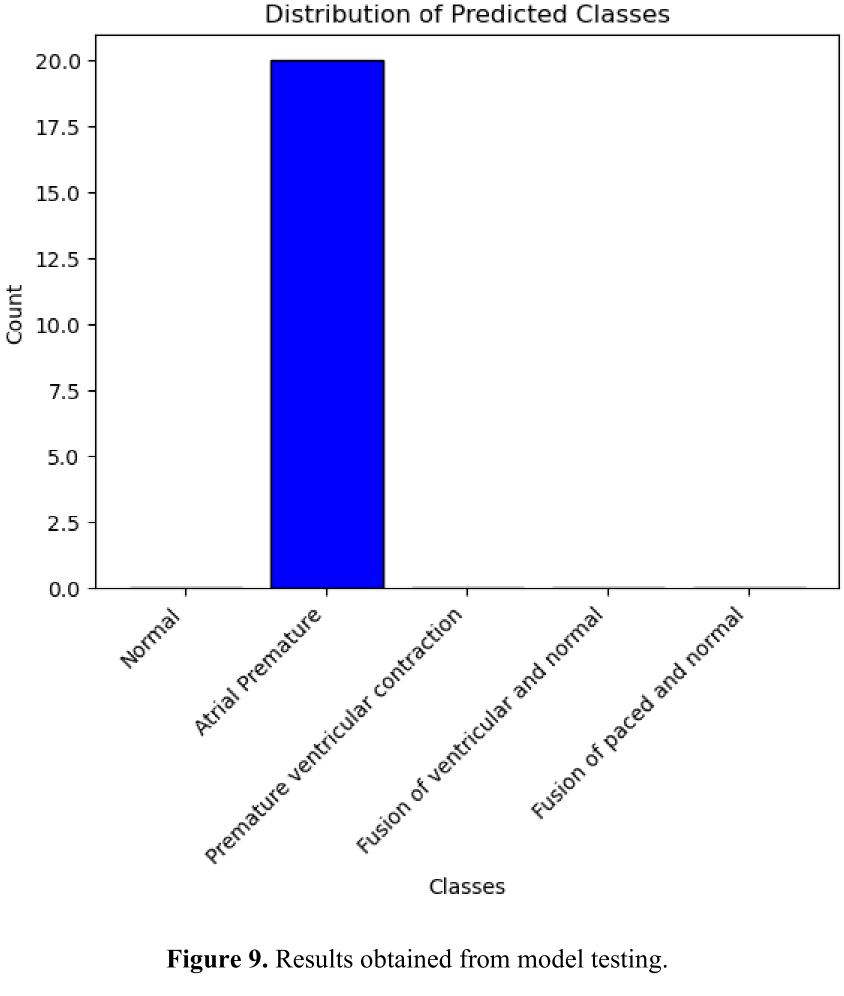


Figure 9. Results obtained from model testing.

The bar chart in Figure 9 illustrates the distribution of predicted classes resulting from the cardiovascular disease (CVD) classification model, which utilises smartwatch data. Among the five possible classes, which are Normal, Atrial Premature, Premature Ventricular Contraction, Fusion of Ventricular and Normal, and Fusion of Paced and Normal, only the Atrial Premature class shows any predictions, with a total count of 20. The remaining classes have a count of zero, indicating that the model exclusively classified all test instances into a single category.

- This highly imbalanced prediction outcome may suggest several possibilities:
- Model bias or overfitting toward the Atrial Premature class.
- Insufficient diversity in the test data, potentially containing only samples of Atrial Premature.
- Issues with class representation in training data, causing the model to generalize poorly across classes.
- Potential label mapping or data preprocessing errors, leading to incorrect class assignments.

Such results highlight the need for further investigation into the model's training and evaluation pipeline, including class distribution, sampling strategy, and performance metrics to ensure the model can robustly distinguish between all target classes.

4. CONCLUSIONS

This study presents the development of a deep learning model designed to predict cardiovascular disease (CVD). The methodology was structured into several distinct phases. Initially, data were gathered from Raqib smartwatches worn by 20 volunteers. However, several challenges were encountered, including difficulties in volunteer recruitment, technical issues with the Favoriot server, and inconsistent data capture by some participants. To strengthen our model's development, we supplemented our data sources with a dataset obtained from Kaggle. Throughout the study, we adhered to established health metrics and benchmarks derived from relevant literature to guide our data interpretation and model evaluation, thereby ensuring a rigorously objective analytical framework. A predictive model for CVD was iteratively refined using Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) architectures, executed in Python with the TensorFlow library. The CNN and LSTM models demonstrated impressive performance, achieving an average precision, recall, and F1 score of 99 per cent and an overall accuracy of approximately 98 per cent on the Kaggle dataset. Notably, the LSTM model demonstrated significantly faster execution than the CNN model, thereby enhancing its practical applicability. In summary, this study accomplished its objectives, overcoming data collection challenges and developing a predictive model characterized by high accuracy, precision, recall, and F1-score. The LSTM model's efficiency during training further highlights its potential for real-world applications in predicting cardiovascular disease.

However, like all studies, this research acknowledges inherent limitations, particularly in data collection and sampling. A primary challenge was volunteer commitment to consistently wearing the smartwatch, which introduced variability in data collection and impacted the dataset's completeness. Furthermore, the short-term nature of data collection limited our ability to capture fluctuations in health metrics, potentially missing transient cardiovascular indicators relevant to risk assessment [18]. The study's sample size and diversity also present limitations, as a small and relatively homogenous sample may not reflect broader populations and can introduce biases. Larger and more diverse cohorts could enhance model generalizability. The variability in smartwatch data quality due to device malfunctions or user errors may also affect heart rate accuracy and ECG readings [19].

Using supplementary data from Kaggle introduced additional challenges, including differences in data collection methods and potential inconsistencies with the smartwatch data. This reliance on external data necessitated adjustments in the training process, highlighting the need for more robust, locally sourced datasets for future research. Finally, the study's dependency on technology implies that outcomes are reliant on the accuracy of the devices and software. Technical issues could lead to data inaccuracies, underscoring the importance of reliable data collection protocols. Addressing these complexities is crucial for refining CVD prediction models and enhancing the validity and applicability of research findings in public health.

The limitations identified in this study point to opportunities for improvement. First, exploring other smartwatch models could improve data quality and reduce costs associated with data formatting and SIM card services. Alternative devices offer better cost-efficiency and usability, encouraging consistent data capture from volunteers. Future collaborations with healthcare institutions could enhance model validation by integrating clinical datasets, which are generally more accurate than self-reported questionnaire data. Incorporating clinical health records would improve the model's reliability and predictive power. Building on this study's model, there is potential for smartwatch developers to integrate continuous ECG readings with personal health data, enabling real-time alerts for users with elevated CVD risks. Continued research in CVD prediction could significantly impact global health by supporting preventive strategies, ultimately aiming to reduce the incidence of CVD worldwide. As the adage goes, "prevention is better than cure," and further exploration in this area can contribute to proactive healthcare practices. Improvements in volunteer engagement and extended data collection timelines will also be crucial, ensuring consistent data quality and better capturing the range of health metrics needed to refine CVD risk assessment tools.

ACKNOWLEDGEMENTS

Institution(s)

All the authors would like to express their gratitude to Universiti Malaysia Pahang Al-Sultan Abdullah and Universiti Airlangga for the support and facilities.

Fund
This paper was created within the project “Development of a predictive algorithm for heart disease using an ensemble of logit model and nearest neighbour tree method based on smartwatch data”. Project registration number RDU210328 under UMPSA Internal Grant.

Individual Assistant
NA

AUTHOR CONTRIBUTIONS

Nor Azuana Ramli (Conceptualization; Supervision; Editing), Mohd Syafiq Asyraf Suhaimi (Methodology; Formal analysis; Investigation; Visualization; Writing - original draft; Resources), Lilik Jamilatul Awalin (Review & Editing).

DECLARATION OF ORIGINALITY

The authors declare no conflict of interest to report regarding this study.

REFERENCES

[1] Department of Statistics Malaysia (DOSM). Statistics on Causes of Death, Malaysia, 2024. Retrieved from https://www.dosm.gov.my/portal-main/release-content/statistics-on-causes-of-death-malaysia-2024.
[2] Grundy SM, Pasternak R, Greenland P, Smith S, Fuster V. Assessment of cardiovascular risk by use of multiple-risk-factor assessment equations: a statement for healthcare professionals from the American Heart Association and the American College of Cardiology. Journal of the American College of Cardiology. 1999;34(4):1348-59.
[3] Park K, Hwang SY. 10-year risk for cardiovascular disease among male workers in small-sized industries. Journal of Cardiovascular Nursing. 2015;30(3):267-73.
[4] Ahsan MM, Luna SA, Siddique Z. Machine-learning-based disease diagnosis: A comprehensive review. In Healthcare 2022 Mar 15 (Vol. 10, No. 3, p. 541). MDPI.
[5] Ahmad I, Alqurashi F. Early cancer detection using deep learning and medical imaging: A survey. Critical Reviews in Oncology/Hematology. 2024;204:104528.
[6] Parthiban G, Srivatsa SK. Applying machine learning methods in diagnosing heart disease for diabetic patients. International Journal of Applied Information Systems. 2012;3(7):25-30.
[7] Taye MM. Understanding of machine learning with deep learning: architectures, workflow, applications and future directions. Computers. 2023;12(5):91.
[8] Hochreiter S, Schmidhuber J. Long short-term memory. Neural computation. 1997;9(8):1735-80.
[9] Kim MJ. Building a cardiovascular disease prediction model for smartwatch users using machine learning: Based on the Korea national health and nutrition examination survey. Biosensors. 2021;11(7):228.
[10] Fawagreh K, Gaber MM, Elyan E. Random forests: from early developments to recent advancements. Systems Science & Control Engineering: An Open Access Journal. 2014;2(1):602-9.
[11] Lee HY, Shin J, Kim GH, Park S, Ihm SH, Kim HC, Kim KI, Kim JH, Lee JH, Park JM, Pyun WB. 2018 Korean Society of Hypertension Guidelines for the management of hypertension: part II-diagnosis and treatment of hypertension. Clinical hypertension. 2019;25(1):20.
[12] Martin A, Fitzsimons C, Jepson R, Saunders DH, van der Ploeg HP, Teixeira PJ, Gray CM, Mutrie N. Interventions with potential to reduce sedentary time in adults: systematic review and meta-analysis. British journal of sports medicine. 2015;49(16):1056-63.
[13] Rens N, Gandhi N, Mak J, Paul J, Bent D, Liu S, Savage D, Nielsen-Bowles H, Triggs D, Ata G, Talgo J. Activity data from wearables as an indicator of functional capacity in patients with cardiovascular disease. Plos one. 2021;16(3):e0247834.
[14] Kachuee M, Fazeli S, Sarrafzadeh M. Ecg heartbeat classification: A deep transferable representation. In 2018 IEEE international conference on healthcare informatics (ICHI) 2018 (pp. 443-444). IEEE.
[15] Albawi S, Mohammed TA, Al-Zawi S. Understanding of a convolutional neural network. 2017 International Conference on Engineering and Technology (ICET), Antalya, Turkey, 2017:1-6.
[16] Sowmya S, Jose D. Contemplate on ECG signals and classification of arrhythmia signals using CNN-LSTM deep learning model. Measurement: Sensors. 2022;24:100558.
[17] Mehmood A, Iqbal M, Mehmood Z, Irtaza A, Nawaz M, Nazir T, Masood M. Prediction of heart disease using deep convolutional neural networks. Arabian Journal for Science and Engineering. 2021;46(4):3409-22.
[18] Moulton S. Long-term health monitoring using wearable technology: Promises and pitfalls. Journal of Health Informatics. 2019:25(1), 28-32.
[19] Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health information science and systems. 2014;2(1):3.