ORIGINAL RESEARCH ARTICLE

Assessment of Risk Factors of Cardiovascular Diseases Using Artificial Neural Network (ANN): A Pilot Study

PGN Swamy[1*], Kanika Rai[2]

[1]Chitkara School of Health Sciences, Chitkara University, Rajpura, Punjab, India; Adichunchanagiri Nursing College, Adichunchanagiri University, BG Nagara, Karnataka, India
[2]Chitkara School of Health Sciences; Centre for Evidence Based Practice in Health Care, Chitkara University, Punjab, India

DOI: 10.55489/njcm.161120255963

ABSTRACT

Background: Cardiovascular diseases (CVDs) are the leading cause of death globally, particularly in urban populations. In India, an estimated 62.5 million lives are lost prematurely due to CVD. Traditional risk assessment tools like the Framingham Risk Score, Systematic Coronary Risk Evaluation, and Reynolds Risk Score are widely used. However, Artificial Neural Networks (ANNs) may provide improved prediction of cardiovascular risks. This study aims to assess cardiovascular risk factors among urban populations using ANN models and compare their effectiveness with traditional methods.

Methodology: A sequential exploratory mixed-method approach was used. The qualitative phase included focus group discussions with 35 healthcare professionals to identify perceived cardiovascular risk factors. Analysis via QDA Miner Lite highlighted health education, screening, and training as crucial roles of health workers, especially concerning diabetes, diet, and hypertension. The quantitative phase used a multi-layer perceptron model to analyze data from 60 participants (pilot phase), divided into 77% training and 23% testing datasets.

Results and Conclusion: The ANN model achieved 100% training accuracy and 85.7% testing accuracy, with an AUC of 0.969, showing strong predictive performance. The model effectively identified moderate-risk individuals, suggesting that ANNs outperform traditional methods in cardiovascular risk prediction.

Keywords: Mixed method, Artificial neural network, Cardiovascular disease, Urban India, AI in Public health, CVD epidemiology

ARTICLE INFO

Financial Support: None declared

Conflict of Interest: The authors have declared that no conflict of interests exists.

Received: 30-08-2025, **Accepted**: 08-10-2025, **Published**: 01-11-2025

***Correspondence:** Mr. PGN Swamy (Email: simha.pgn@gmail.com)

How to cite this article: Swamy PGN, Rai K. Assessment of Risk Factors of Cardiovascular Diseases Using Artificial Neural Network (ANN): A Pilot Study. Natl J Community Med 2025;16(11):1131-1141. DOI: 10.55489/njcm.161120255963

www.njcmindia.com │ pISSN: 0976-3325 │ eISSN: 2229-6816 │ Published by Medsci Publications

INTRODUCTION

Cardiovascular diseases accounts for more deaths globally with an estimate of 17.9 million lives every year according to the World Health Organization (WHO); CVD death rate in India is 272 per 1 Lakh population higher than the global average of 235.[1] CVDs consist of range of heart diseases like coronary heart disease, cerebrovascular disease, and hypertension. Urban population is at risk due to factors such as sedentary lifestyle, poor dietary habits, and exposure to pollution. The transition from rural to urban living is associated with an increase in the prevalence of cardiovascular risk factors pointing at targeted intervention strategies in these settings. Risk factors of CVDs include modifiable factors such as unhealthy diet, high blood pressure, physical inactivity, tobacco use, obesity and air pollution. It needs for multisectoral policy interventions focused on prevention, early diagnosis, and management tailored to India's diverse populations.[2] These modifiable risk factors collectively account for over 90% of myocardial infarction risk worldwide. This demonstrated that these risk factors are consistent regardless of ethnicity or geography supporting the global relevance of targeting these factors for myocardial infarction prevention.[3]

Fast-paced urban lifestyle may lead to chronic stress which may be associated with adverse cardiovascular outcomes. Stress can contribute to behaviors like poor diet, smoking and physical inactivity, exacerbating cardiovascular risk.[4]

Traditional models assume linear relationships between variables. In contrast, ANNs can model complex, non-linear interactions between a large number of variables providing more understanding of risk factors and their interplay. ANNs can integrate and analyze a vast array of variables including clinical parameters, patient demographics and even unstructured data like clinical notes providing comprehensive risk assessment. Unlike traditional models, ANNs can continuously learn and improve as they are exposed to more data, potentially increasing their predictive accuracy over time. Austin PC and Steyerberg EW investigated how EPV influences the accuracy and validity of logistic regression models specifically analysing different methods for estimating out-of-sample model performance. Findings of the study indicated that low EPV can lead to substantial overfitting and unreliable prediction while higher EPV improves model stability and validity.[5] Shouval R et al[6] demonstrate that ML methods can improve predictive accuracy over traditional statistical approaches by capturing complex interactions and nonlinearities in clinical data. Caruana R et al[7] showcased that intelligible models can achieve high accuracy while also providing explanations, supporting integration into clinical decision-making. It is important to note that ANNs also have limitations, including their "black box" nature, which can make it challenging to interpret how they arrive at their predictions. They require large datasets for training to perform optimally and avoid overfitting.

Machine learning techniques are most commonly used for predicting coronary heart disease (CHD) out of which neural network is popularly used to improve the performance accuracy of results. Kim JK and Kang S[8] used NN based prediction model using feature correlation analysis (NN-FCA) for predicting the CHD risk among 4146 individuals out of whom 3031 had low CHD risk and 1115 had high risk for CHD. Picture representation of area under receiver operating characteristic (ROC) curve of the model was 0.749 which was larger than the Framingham risk score of 0.393 in ROC curve. To conclude NN-FCA was found to be better in predicting CHD risk than FRS.

A cross-sectional survey was conducted in Mwanza, Tanzania by Malindisa EK et al[9] to evaluate the prevalence of type 2 diabetes mellitus (T2DM), hypertension, dyslipidemia, and abdominal obesity among 245 young adults. The study found alarming prevalence rates: 7.8% for T2DM, 15.5% for impaired glucose tolerance, 11.8% for abdominal obesity, 45.1% for dyslipidemia, and 34.3% for hypertension. The research underscores the significant cardiovascular risk among young adults in urban Tanzania, highlighting the urgent need for targeted interventions to combat these risk factors.

The Kaunas cohort study by Tamosiunas A et al[10] found that the use of urban green spaces such as parks is associated with lower prevalence of cardiovascular risk factors and diabetes. Increased distance from green spaces was linked to higher risks of both non-fatal and fatal cardiovascular diseases especially among men and women who did not use parks regularly. The study suggests that proximity to and use of green spaces promote physical activity and reduce cardiovascular risks supporting public health policies that integrate urban green areas to enhance cardiovascular health in urban population.

Identification of Research Gaps: A systematic reviews and meta-analysis strategy were used by de Bont J et al[11] through multiple databases between 2010 and 2021 to assess the impact of ambient air pollution on cardiovascular diseases. Out of 56 reviews, most studies were related to stroke (22), all causing CVD mortality and morbidity. Results of the study was evident with short term exposure to particulate matter <2.5 μm, <10μm, and nitrogen dioxide were associated with higher risk of hypertension, MI and stroke. Higher association was found in Asian countries particularly among elderly, cardiac patients and obese people. In conclusion results of the study demonstrated need for reducing air pollution across worldwide especially in Asian countries and vulnerable population.

In another study, Goldman O et al[12] tried to predict coronary heart disease risk using Artificial neural networks and compared the same data sets with FRS. The study used a cohort of 3066 sample from FHS offspring who were subjected to multilayer perceptron ANN model for predicting CHD risk and were compared to FRS scores later. Results shown that lift and gain curves of ANN model outperformed as compared to FRS in terms of percentiles. ANN model had higher sensitivity and specificity than of FRS model. ANN has given good results than FRS with high ACU curves. To conclude, ANN model is a good approach in predicting CHD risk to identify CHD high risk groups.

METHODOLOGY

Informed consent was obtained from the participants involved in qualitative data collection and quantitative phase. Approval from Institutional Human Ethics committee was obtained from the place where research is conducted. Sequential exploratory design was used for qualitative research design wherein it involves collection of qualitative data using focussed group discussion (3 groups i.e. Nursing staff 10, Nursing faculty 15 and medical doctors 10) to explore factors that would possibly influence in determining the cardiovascular health. Qualitative data was collected by asking questions through audio transcripts and organized by using QDA miner lite software. Coding of the data, themes, subthemes and final conclusions of qualitative data were drawn based on the data. The qualitative data collection was followed by quantitative phase in which the effectiveness of ANN algorithms was measured. A sample of 60 participants were selected using purposive sampling technique for the quantitative phase. The clinical and non-clinical data from the patients were collected and subjected to multilayer perceptron for quantitative analysis.

Approval of Institutional Ethical Review: (From the place where research is conducted: - Institutional ethics board of Adichunchanagiri Institute of medical sciences bearing Reg No. EC/NEW/INST/2023/KA/0382. Approval No. AIMS/IEC/90/2024 dated 10.08.2024).

(From the place where PhD is conducted: Institutional Human Ethics Committee, Chitkara University; Approval No. EC/NEW/INST/2024/531/365 dated 12th Dec 2024)

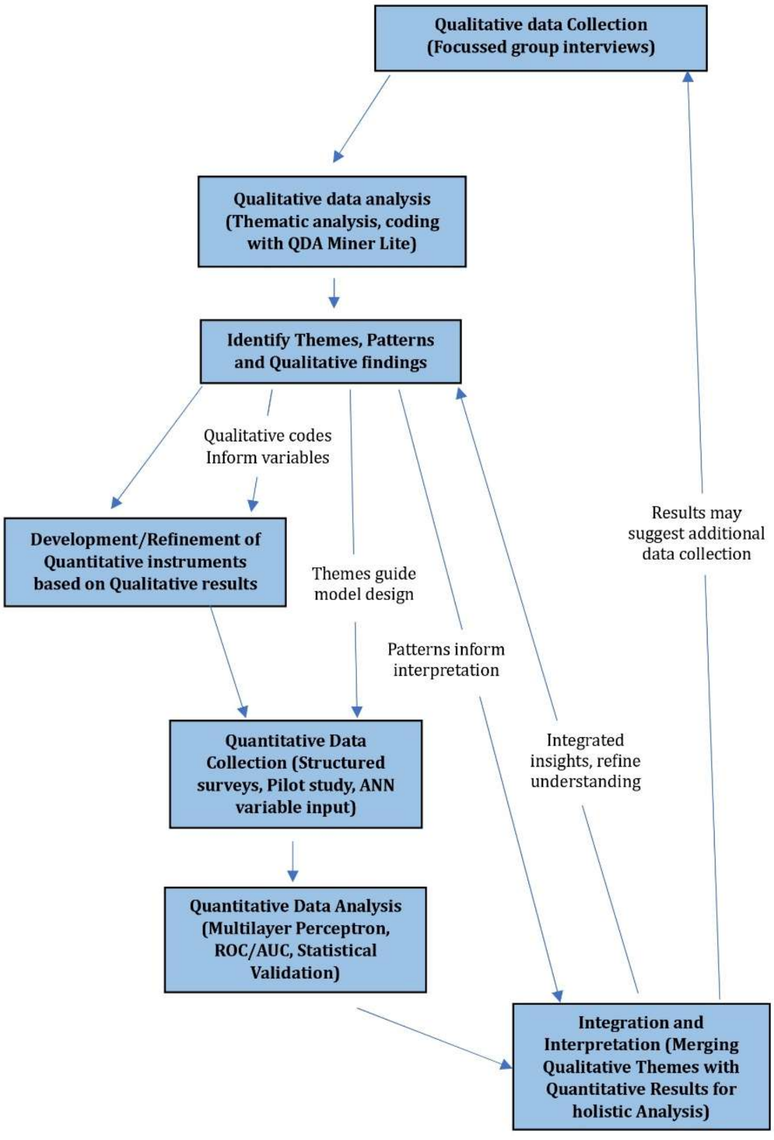

Fig 1: Flow chart of sequential exploratory design

RESULTS

Qualitative Data Analysis Using QDA Miner Lite are as follows: Data preparation phase included collection of qualitative data with focused groups (using 35 samples from three focused groups including staff nurses, teaching faculty and medical doctors) using audio transcriptions (27 questions) followed by formatting of the data wherein responses from all 3 focused groups were collected and organized in a text format which is suitable for supporting QDA Miner Lite software. Each participant's response is clearly collected in one file. All textual data were imported into the software. Data was familiarised carefully to get general essence of the content and was removed in case of repetition of the responses. Out of all responses among 3 focused groups, 4 unique themes were identified under which subthemes were then categorized as listed in the subsequent tables. Codes are assigned that represent specific concepts or categories. Themes (broader concepts) and subthemes (more specific) are identified as listed in the table.

Health education is an important factor where participants strongly recommend this in educating communities. Screening and in service education are also needed for early detection and ongoing health workers training. Follow up care and managing risks shows attention towards continuity of care. Less mentioned probably are low focussed but relevant are coordination, diet, advice, updating knowledge, physical activity, records and evidence-based practice (Table 1).

Blood pressure, Diabetes and junk food were found to be the leading risk factors indicating that participants need to emphasize these influences strongly. Smoking, tobacco usage and alcohol are moderate risk factors. Obese people and who are living with kidney disfunctions are leading risk categories. Less identified but relevant factors are stress, lifestyle, family history, age gender, noncompliance to treatment and arrhythmias.

Table 1: Interpretation of Role of health workers and related subthemes

Subtheme	Count (%)	Cases (%)	Nb Words (%)	Interpretation
Health Education	10 (23.3)	10 (27.8)	-	Most important in community awareness and prevention
Screening Role	7 (16.3)	6 (16.7)	-	Early detection through screening is an important function
In service Education	7 (16.3)	7 (19.4)	-	Need for Continuous professional development
Follow Up Care	5 (11.6)	5 (13.9)	-	Need for Monitoring and ongoing support
Coordination	3 (7)	2 (5.6)	-	Collaborative approach for integrated care
Management of Risk	4 (9.3)	4 (11.1)	-	Active role in reducing and managing health risks
Diet Advice	2 (4.7)	2 (5.6)	-	Nutritional guidance is part of health education
Updating Knowledge	2 (4.7)	2 (5.6)	-	Keeping up-to-date knowledge is necessary
Physical Activity	1 (2.3)	1 (2.8)	-	Encouraging exercise but given less importance
Records and Reports	1 (2.3)	1 (2.8)	-	Documentation is important even though less mentioned
Evidence Based Practice	1 (2.3)	1 (2.8)	-	Use of scientific evidence is less frequently discussed

Table 2: Risk Factors (Main theme) and related subthemes

Subtheme	Count (%)	Cases (%)	Nb Words (%)	Interpretation
Diabetes	8 (12.3)	7 (19.4)	7 (1.3)	Most likely recognized risk factor
Junk Food	7 (10.8)	7 (19.4)	-	Dietary risk linked to disease
Blood Pressure	5 (7.7)	5 (13.9)	30 (5.7)	Hypertension identified as a major modifiable risk
Smoking	5 (7.7)	5 (13.9)	4 (0.8)	Biochemical markers are well elaborated
Alcohol	4 (6.2)	4 (11.1)	-	Recognized lifestyle risk factor
Tobacco	4 (6.2)	4 (11.1)	-	Substance use contributing to risk
Kidney Functions	4 (6.2)	4 (11.1)	14 (2.7)	Tobacco use identified
Obesity	4 (6.2)	4 (11.1)	5 (0.9)	Kidney health linked to cardiovascular risk
Stress	3 (4.6)	3 (8.3)	11 (2.1)	Excess weight is a recognized risk factor
Lifestyle	5 (7.7)	6 (16.7)	8 (1.5)	Psychosocial factors are important but less discussed
Family History	2 (3.1)	2 (5.6)	22 (4.2)	General lifestyle patterns impact risk
Age	2 (3.1)	2 (5.6)	-	Genetic predisposition is also identified as risk factor
Gender	2 (3.1)	2 (5.6)	-	Non-modifiable risk factor
Occupation	1 (1.5)	1 (2.8)	-	Biological differences identified as an evident risk
Noncompliance	1	1	17	Work environment influencing the risk but less recognized
Arrhythmias	1	1	-	Treatment adherence is an important concern
Homocysteine	1	1	2	Specific cardiac condition identified
Lipid Profile	6	6	47	Biochemical markers are less stressed

Table 3: Lifestyle Choices (Theme)

Subtheme	Count (%)	Cases (%)	Nb Words (%)	Interpretation
Sedentary Lifestyle	20 (39.2)	19 (52.8)	44 (9.3)	Considered as most recognized risk factor
Imbalanced Diet	16 (31.4)	16 (44.4)	72 (15.3)	Dietary risk is mostly and widely discussed risk
Tobacco Dependence	14 (27.5)	11 (30.6)	21 (4.4)	Addiction identified as significant lifestyle risk

Table 4: Weightage of Risk Factor

Subtheme	Count (%)	Cases (%)	Nb Words (%)	Interpretation
Combined BP and Diabetes	6(17.6)	5(13.9)	40 (7.6)	High risk perception for combined hypertension and diabetes
High Blood Pressure	2 (5.9)	2 (5.6)	30 (5.7)	Standalone hypertension recognized as key factor
Combined Smoking and High BP	2 (5.9)	2 (5.6)	16 (3)	Smoking & hypertension seen as particularly harmful
Kidney Disorder with Diabetes	2 (5.9)	2 (5.6)	14 (2.7)	Combined kidney and diabetes risk acknowledged
Combined BP, Diabetes & Lifestyle Habits	1 (2.9)	1 (2.8)	21 (4)	Multi-factor risk combination less frequent but important
Kidney Disorder with High BP	1 (2.9)	1 (2.8)	14 (2.7)	Additional comorbid conditions are noted
Lifestyle and Biochemical Changes	1 (2.9)	1 (2.8)	18 (3.4)	Interaction of lifestyle with lab markers

Chida Y et al[13] explained that urban areas, especially highly industrialized cities, tend to have higher levels of air pollution. Long-term exposure to pollutants is a known risk factor for cardiovascular diseases. Vlahov D et al[14] identified that high population density can contribute to stress, pollution, and the spread of infectious diseases, all of which can indirectly influence cardiovascular health. Galea S et al[15] identified that urbanization is associated with transition towards diets higher in fats, sugars and processed foods contributing to obesity, hypertension and dyslipidaemia which all are risk factors for cardiovascular disease .

Sedentary lifestyle is frequently coded risk suggesting lack of physical activity as a primary concern. Imbalanced diet is also extensively highlighted by the participants. Tobacco dependence is an important but less discussed than physical inactivity and diet (table 3).

Participants expressed the combined risks of having both high blood pressure and diabetes as significant factors. Combination of smoking, kidney disorders and lifestyle factors appear less frequently which reflect that multiple risk factors frequently coexist and affect outcomes (table 4).

Diabetes (19.4) and junk food (19.4) are the major risk factors followed by sedentary life style (16.7) and lipid profile (16.7) highlighting need to address these risk factors (Fig 2).

To conclude, role of health workers includes health education, screening and in-service education paying attention on prevention and capacity building of health workers. Emerging factors associated with metabolic and combined risks focus on integrated care.

Quantitative data analysis using Multi-layer Perceptron (MLP) model is as follows: Multilayer Perceptron is a type of ANN made up of several layers. Input Layers are the layers through which data is entered. e.g. patients age, cholesterol, smoking habits etc. Hidden Layers process the data through many connected nodes to find complex relationships. Output Layer gives the final output. e.g. predicting whether a patient is at moderate or high cardiovascular risk. It’s called as ‘multilayer’ because it has more hidden layers between input and output which helps it to learn more complex patterns than simple models.

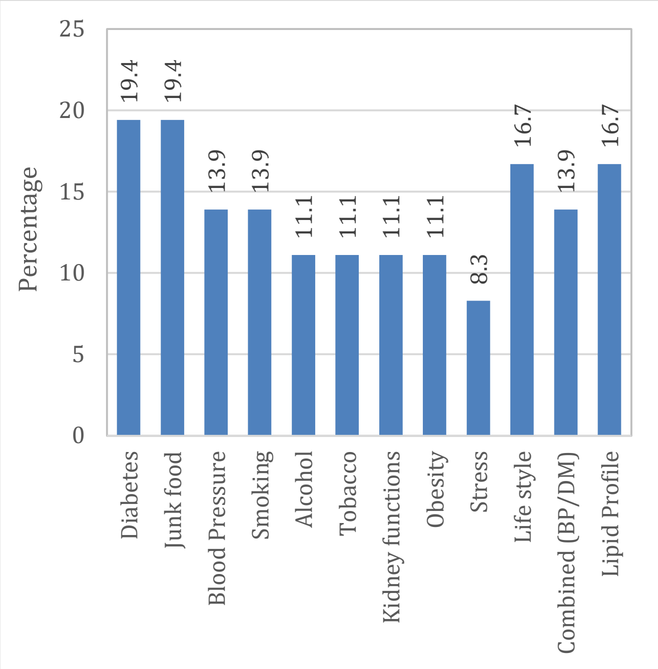

Fig 2: Overall risk factors influencing cardiovascular disease risk factors

A total of 60 participants were used with data on their health and risk factors with no missing values. About 76.7% (46 people) of the participants were randomly selected as training sample to teach the neural network model to learn patterns, e.g. How age, smoking, and cholesterol relate to cardiovascular risk. Remaining 23.3% (14 people) were not used in the training i.e. they were used as testing sample to test the model's ability to predict risk on new unseen cases. This is done to check if the model can generalize its learning beyond just memorizing the training data.

Network Information: Input variables are the details that the model uses to make predictions. Categorical variables include Gender, History of Diabetes Mellitus, Lifestyle, Occupation, Tobacco chewing history, Alcohol consumption and Smoking habit. Covariates include Age, BMI, Total cholesterol, LDL, HDL, HbA1c, CRP, Homocysteine and Creatinine levels. Since few factors were categorical and required encoding, the total input units used were 26. Model contains one hidden layer with four units. These are intermediate processing points where the model learns complex relationships between inputs and the outcome. All numeric inputs are adjusted to a common scale (with mean 0 and standard deviation 1) so no one variable dominates just because of its scale (e.g. Age vs Cholesterol values) (table 5).

Activation functions: Hidden layer uses Hyperbolic tangent (tanh) allows the network to model complex non-linear relationships by transforming the inputs into values between -1 and 1. Output layer uses SoftMax to convert the output into probabilities for each risk category (Moderate or High risk) allowing the model to decide in which class the participants are belonging to. Cross-entropy measures the difference between predicted probabilities and actual outcomes. Lower cross-entropy indicates better prediction.

Model Summary:

The training cross-entropy error was 0.765, indicating that the model's predictions were close to the actual outcomes and demonstrating a good fit to the training data. The training error rate was 0.0%, meaning that the model correctly classified all 46 training participants without any misclassification. During the testing phase, the cross-entropy error increased to 5.435, which is expected when the model is applied to unseen data, as prediction errors typically rise outside the training set. The testing error rate was 14.3%, indicating that the model misclassified approximately 2 out of 14 testing participants. Overall, while the model performed perfectly on the training data, its higher testing error suggests some decline in generalization accuracy when predicting new cases.

Predicted probabilities clustered near to 1 for people who have moderate risk, some have low predicted probabilities (outliers). People with high actual risk, predicted probabilities spread between 0.2 to 0.8. (Fig 4)

Table 5: Network Information

Input Layer	
Factors	1. Gender 2. H/o Diabetes Mellitus 3. Lifestyle 4. Occupation 5. H/o of Tobacco chewing 6. H/o alcohol consumption 7. Smoking habit
Covariates	1. age 2. BMI 3. Total cholesterol 4. LDL 5. HDL 6. HbA1C 7. CRP 8. Homocysteine 9. Creatinine
No. of Units[a]	26
Rescaling Method for Covariates	Standardized
Hidden Layer(s)	
No. of Hidden Layers	1
No. of Units in Hidden Layer 1[a]	4
Activation Function	Hyperbolic tangent
Output Layer	
Dependent Variables	1. CVD risk
No. of Units	2
Activation Function	SoftMax
Error Function	Cross-entropy

[a] Excluding the bias unit; H/o = History of

Table 6: Classification of Results

Observed	Predicted		
	Moderate risk	**High Risk**	**Percent Correct**
Training			
Moderate risk	44	0	100.0%
High Risk	0	2	100.0%
Overall Percent	95.7%	4.3%	100.0%
Testing			
Moderate risk	12	0	100.0%
High Risk	2	0	0.0%
Overall Percent	100.0%	0.0%	85.7%

Dependent Variable: CVD risk

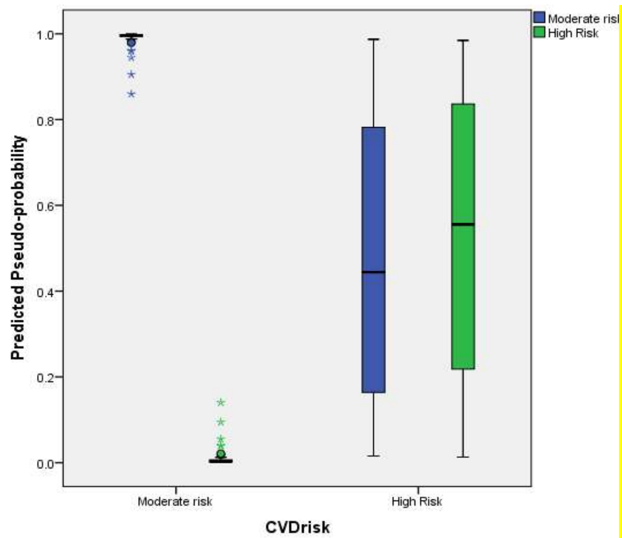

Fig 4: Predicted probability of having cardiovascular disease (CVD) risk

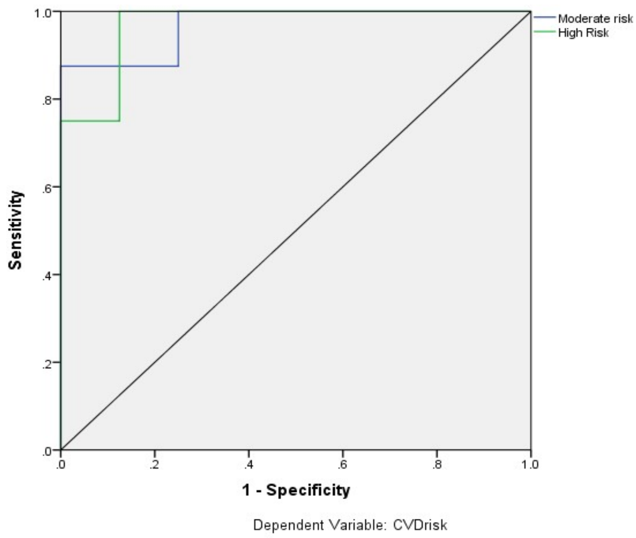

Fig 5: Prediction model (ROC) separating moderate risk with high risk

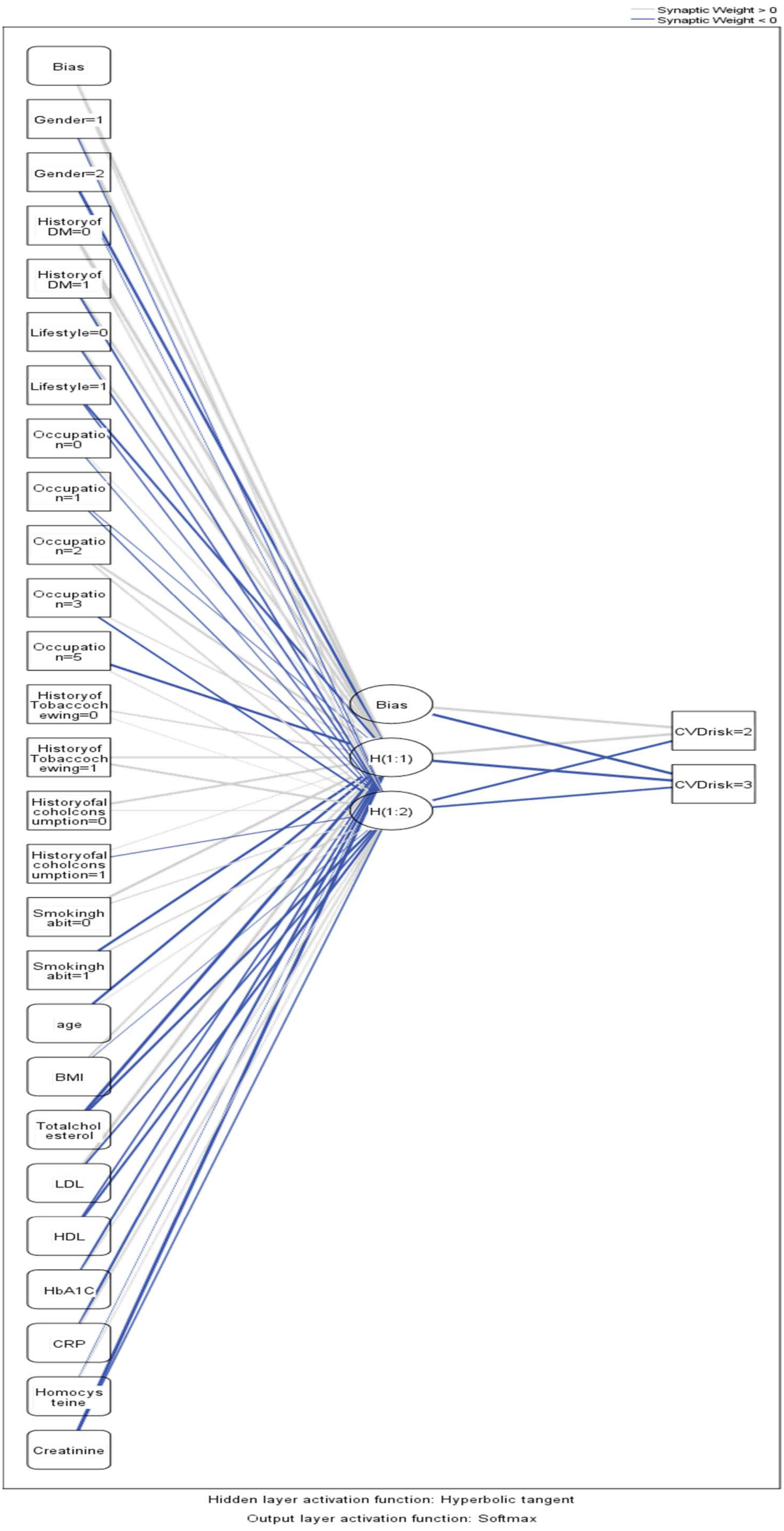

Fig 3: Multilayer perceptron model on network information

Specificity represents the rate of false positive cases. Sensitivity shows true positive cases. Blue and green lines show the model's performance for moderate and high risk. Diagonal black line from bottom-left to top-right is what would be got if the model made random guesses no predictive ability. Closer the coloured line is to the top-left corner, better the model is at correctly identifying risk groups. Blue and green lines are near to the top-left indicates the model has good accuracy.

Overall result shows that in training data, all 44 participants were correctly identified as having moderate risks for CVDs. All participants were correctly identified as high risk (2 Participants) and in testing data, all 12 participants were correctly classified as moderate risk. two participants were misclassified as moderate risk instead of high risk. Overall testing accuracy was 85.7% (table 6).

Area Under the Curve (AUC): This measures the model's ability to correctly distinguish between risk groups. For Moderate risk, it is 0.969 and for high risk, it is 0.969 (Fig 5).

The Gain chart helps us to understand how well a predictive model identifies people with the outcome of interest. Gain is more gradual for moderate risk group (blue line) that shows that the model is less sharp in identifying moderate-risk people still better than random (fig 6).

The lift is very high for high-risk group at low percentages that means the model is 7.5 times better than random at identifying high-risk people in the top 10% selected. The lift decreases as more people are selected approaching random chance as selected everyone. Lift stays close to 1 for moderate risk indicating the model is not much better than random at identifying moderate risk people (Fig 7).

Artificial Neural Network (ANN) Performance Metrics: The model demonstrated excellent predictive performance on the training dataset, achieving an overall accuracy of 100% with 0.0% incorrect predictions and a cross-entropy error of 0.765. This indicates that the ANN model perfectly classified all training cases, reflecting a strong fit to the training data.

When evaluated on the testing dataset, the model achieved an overall accuracy of 85.7%, with 14.3% incorrect predictions and a cross-entropy error of 5.435. This decline in accuracy and increase in error are expected, as the testing dataset represents new, unseen cases.

The classification results further show that among the testing samples, the model correctly classified all 12 participants with moderate CVD risk (100% accuracy). However, it failed to correctly classify any of the 2 participants with high CVD risk (0% accuracy). This suggests that while the model performs very well for moderate-risk prediction, it has limitations in accurately identifying high-risk cases.

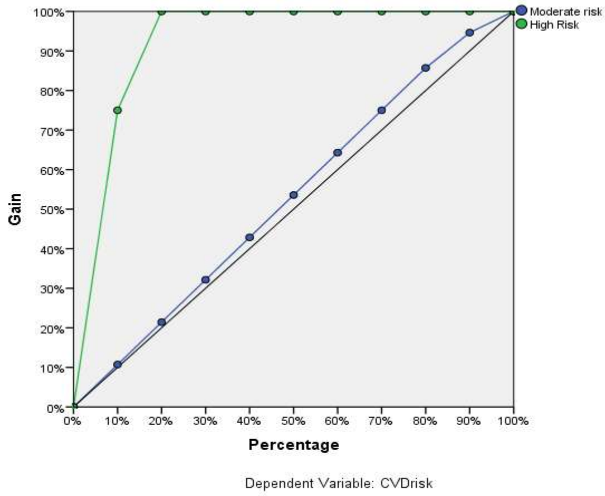

Fig 6: Gain chart: predicting the outcome

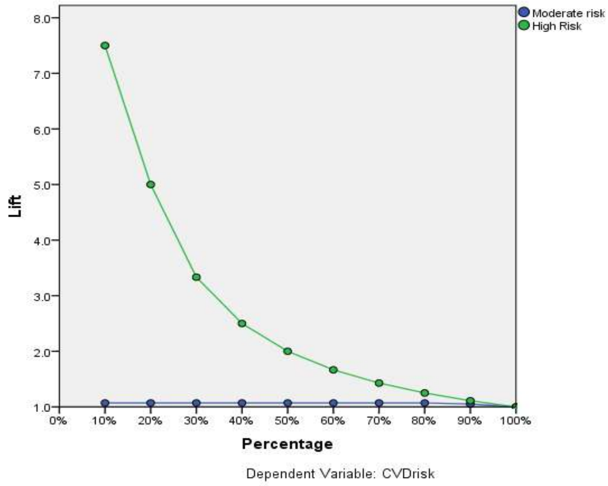

Fig 7: Lift chart: predicting the risk compared to random chance

AUC with confidence intervals: Moderate risk: AUC – 0.969; and High risk: AUC – 0.969. AUC curve reflects good model discrimination whereas confidence intervals from similar literatures suggest a range of ±0.01 for large sample. With limited sample size of high-risk cases, the study needs more data for robust CI reporting in urban context.

Sub group thematic analysis of urban population using focussed groups: The study samples from Urban Nagamangala Taluk identifies heterogeneous risk profiles influenced by lifestyle, socio economic factors and environmental influences. Obesity is notably high in urban areas compared with semi urban and rural areas. Lee IM et al[16] suggests that decreasing physical inactivity by 10% to 25% could prevent more than half a million to over 1.3 million deaths annually. Additionally, eliminating inactivity could boost global life expectancy by about 0.68 years, emphasizing its major impact on health and longevity. Air pollution, sedentary life style and poor dietary pattern make differences in CVD risk affecting elderly people and obese people as well as urban setting in which people living. Kalra K et al[17] advocated for more comprehensive patient-centred approach that included outcomes related to functional status, cognitive health, mental well-being and overall quality of life. They also emphasized the importance of involving older adults in research as they often have complex health profiles and value different outcomes than younger patients requiring tailored treatment strategies.

Clinical Interpretation of Results:

Indication of results to health care providers: The model performs good in identifying moderate-risk patients so that it can help patients who may not need immediate intensive interventions but should be monitored and managed proactively. The model misses high-risk patients. High AUC values show that the model can distinguish patients properly by overall risk but the exact cut-off for classifying someone as high risk might need adjustment. False negatives are critical to avoid because these patients might not get timely care. Additional screening or monitoring needed for patients among whom model classifies as moderate risk and has other warning signs.

Discussion

The ANN model showed high accuracy and strong area under curve (AUC) data (0.969 for both moderate and high-risk groups), classifying participants by CVD risk based on various clinical and life style variables of urban population. Key variables include age, gender, diabetes mellitus, lifestyle, BMI, HDL, LDL and other biomarkers. This model correctly identified moderate risk individuals in tested data and showed perfect discrimination in training data indicating good recognition and fit for moderately risk groups.

Concerns on Overfitting and small high-risk subgroups: In spite of high AUC statistics, the model classified all high-risk cases among testing data set as moderate risk suggesting two main limitations; **Potential Overfitting**: Model performed correctly on training data set but was less effective for a smaller number of high-risk testing data set. This indicates overfitting where ANN memorizes training data set patterns and does not recognize well to unseen subgroups. **Class imbalance**: Testing data set had small number of high-risk individuals making it hard for the model to distinguish cases which is a well-known phenomenon in machine learning in health care when class distribution is classified.

Implications for urban personalized risk assessment: Results of the study confirm that ANNs can integrate complex, non-linear data and provide more individual specific risk stratifications in urban population especially for moderate-risk groups who are sometimes missed by traditional methods. Singh M et al[18] (2024) demonstrated that ANNs can integrate complex clinical and behavioural data to provide more precise individual risk stratification in cardiovascular disease (CVD). Models like Neural CVD -DSM showed improved reclassification for moderate-risk individuals compared to traditional scores, aligning with current CVD prevention guidelines. The use of multiple clinical (e.g. cholesterol, creatinine) and behavioural (e.g. smoking, alcohol use) factors are the significant improvement factors over traditional scoring like Framingham risk or SCORE. This alignment with contemporary CVD prevention guidelines is of high value for urban health management, where risk factor diversity and social determinants are prominent. Weng WH et.al. used deep learning PPG-based cardiovascular disease risk score (DLS) in this study, predicts the 10-year probability of major adverse cardiovascular events (MACE) using only age, sex, smoking status and photoplethysmography (PPG) data. It demonstrated non-inferior predictive accuracy compared to the office-based refit-WHO score achieving a C-statistic of 71.1% versus 70.9%. The calibration of the DLS was satisfactory with a mean absolute calibration error of 1.8%. Adding DLS features to the traditional risk score improved the discrimination slightly by 1.0%. The interpretability analysis indicated that the DLS features extracted from the PPG waveform morphology were independent of heart rate. This study supports the potential of PPG-based deep learning models as accessible low-cost tools for community-level primary prevention of CVD especially in resource-limited settings where traditional clinical measurements are less accessible[19]. Asadi F et al[20] (2024) demonstrated that advanced tree-based machine learning algorithms outperform traditional linear models for detecting cardiovascular disease, by leveraging a large dataset with diverse risk factors and accounting for data clustering. The GMERF model showed the highest predictive accuracy and sensitivity, underscoring the improvements offered by machine learning approaches in personalized and population-level CVD risk prediction for urban settings. Meder B et al[21] (2025) discussed how real-world AI systems integrating cholesterol, creatinine, smoking and alcohol use enable population-level risk management for urban communities. AI can process diverse environmental and social determinants and identify urban-specific exposures proving more adaptable and precise than classical models.

Model Robustness and Need for Large-scale Validation: The findings showed the necessity for follow-up studies using larger, more diverse urban cohorts to test the model's robustness and performance on rare, high-risk cases. Expanding the sample size and ensuring adequate representation of high-risk categories will allow the ANN to find out more examples which supports both generalization and avoidance of overfitting. Agarbattiwala MH et al[22] also signifies that AI models consistently performs better than the traditional methods in predicting cardiovascular events. Additional strategies could include transfer learning from other urban cohorts or registries; External validation in different cities or urban populations; Interpretability enhancements (to clarify model decisions to clinicians).

Limitations

Limited sample size attributed to inaccurate risk profiling however this will be substituted using large sample in main study.

Conclusion

The study concluded that Artificial Neural Networks (ANNs) demonstrated higher accuracy and predictive performance in assessing cardiovascular risk as compared to traditional methods. With high model accuracy and strong discrimination (AUC = 0.969) ANNs effectively identified moderate-risk individuals, highlighting their potential as a reliable tool for early cardiovascular disease risk assessment in urban population.

Future Directions:

The model is expected to improve as it is exposed to larger and more heterogeneous urban populations, enhancing its generalizability and predictive accuracy. Future work should emphasize continuous training as more data become available, particularly focusing on high-risk subgroups to refine the ANN's sensitivity. Additionally, integrating the model with clinician judgment and traditional risk assessment tools can strengthen its clinical utility in identifying cardiovascular risks. Furthermore, deploying the model within digital health platforms could facilitate personalized CVD risk prediction, enable ongoing monitoring and ensure transparency in model operation.

Acknowledgement: The author expresses sincere gratitude to all participants for their valuable support and contribution to the successful conduct of this study.

Individual Authors' Contributions: PS initiated the study concepts, structured the research design, carried out data collection and analysis and drafted the initial version of the manuscript. **KR** provided overall mentorship and expert input on the study design and analytical framework, guided data interpretation and critically revised the manuscript to enhance its academic quality, approved the final manuscript for publication

Availability of Data: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declaration of Non-use of Generative AI Tools: This article was prepared without the use of generative AI tools for content creation, analysis, or data generation. All findings and interpretations are based solely on the authors' independent work and expertise.

References

1. World Health Organization. Cardiovascular diseases. Geneva: World Health Organization; [cited 2025 Aug 13]. Available from: https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1
2. Prabhakaran D, Jeemon P, Roy A. Cardiovascular Diseases in India: Current Epidemiology and Future Directions. Circulation. 2016 Apr;133(16):1605-1620. DOI: https://doi.org/10.1161/CIRCULATIONAHA.114.008729 PMid:27142605
3. Yusuf S, Hawken S, Ounpuu S, Dans T, Avezum A, Lanas F, McQueen M, Budaj A, Pais P, Varigos J, Lisheng L; INTERHEART Study Investigators. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. Lancet. 2004 Sep 11-17;364(9438):937-952. DOI: https://doi.org/10.1016/S0140-6736(04)17018-9 PMid:15364185
4. Hajat C, Stein E. The global burden of multiple chronic conditions: A narrative review. Prev Med Rep. 2018 Oct 19;12:284-293. DOI: https://doi.org/10.1016/j.pmedr.2018.10.008 PMid:30406006 PMCid:PMC6214883
5. Austin PC, Steyerberg EW. Events per variable (EPV) and the relative performance of different strategies for estimating the out-of-sample validity of logistic regression models. Stat Methods Med Res. 2017 Apr;26(2):796-808. DOI: https://doi.org/10.1177/0962280214558972 PMID: 25411322.
6. Shouval R, Bondi O, Mishan H, Shimoni A, Unger R, Nagler A. Application of machine learning algorithms for clinical predictive modeling: a data-mining approach in SCT. Bone Marrow Transplant. 2014; 49(3): 332-337. DOI: https://doi.org/10.1038/bmt.2013.146 PMid:24096823
7. Caruana R, Lou Y, Gehrke J, et al. Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2015:1721-1730. DOI: https://doi.org/10.1145/2783258.2788613
8. Kim JK, Kang S. Neural Network-Based Coronary Heart Disease Risk Prediction Using Feature Correlation Analysis. J Healthc Eng. 2017;2017:2780501. DOI: https://doi.org/10.1155/2017/2780501 PMID: 29065583; PMCID: PMC5606055
9. Malindisa EK, Balandya E, Mashili F, Iddi S, Njelekela M. The magnitude of type 2 diabetes mellitus and cardiovascular disease risk factors among young adults in urban settings: a cross-sectional survey in Mwanza, Tanzania. Pan Afr Med J. 2022 May 10;42:19. DOI: https://doi.org/10.11604/pamj.2022.42.19.22184 PMID: 35812257 PMCID: PMC9228915
10. Tamosiunas A, Grazuleviciene R, Luksiene D, Dedele A, et al. Accessibility and use of urban green spaces, and cardiovascular health: findings from a Kaunas cohort study. Environ Health. 2014 Mar 19;13(1):20. DOI: https://doi.org/10.1186/1476-069X-13-20 PMid:24645935 PMCid:PMC4000006
11. de Bont J, Jaganathan S, Dahlquist M, Persson Å, Stafoggia M, Ljungman P. Ambient air pollution and cardiovascular diseases: An umbrella review of systematic reviews and meta-analyses. J Intern Med. 2022 Jun;291(6):779-800. DOI: https://doi.org/10.1111/joim.13467 PMid:35138681 PMCid:PMC9310863
12. Goldman O, Raphaeli O, Goldman E, Leshno M. Improvement in the Prediction of Coronary Heart Disease Risk by Using Artificial Neural Networks. Qual Manag Health Care. 2021 Oct-Dec 01;30(4):244-250. DOI: https://doi.org/10.1097/QMH.0000000000000309

PMid:34326290

13. Chida Y, Steptoe A. The association of anger and hostility with future coronary heart disease: a meta-analytic review of prospective evidence. J Am Coll Cardiol. 2009 Mar 17;53(11):936-946. DOI: https://doi.org/10.1016/j.jacc.2008.11.044 PMid:19281923
14. Vlahov D, Freudenberg N, Proietti F, Ompad D, Quinn A, Nandi V, Galea S. Urban as a determinant of health. J Urban Health. 2007 May;84(3 Suppl):i16-26. DOI: https://doi.org/10.1007/s11524-007-9169-3
15. Galea S, Vlahov D. Urban health: evidence, challenges, and directions. Annu Rev Public Health. 2005;26:341-365. DOI: https://doi.org/10.1146/annurev.publhealth.26.021304.144708 PMid:15760293
16. Lee IM, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT; Lancet Physical Activity Series Working Group. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet. 2012 Jul 21;380(9838):219-229. DOI: https://doi.org/10.1016/S0140-6736(12)61031-9 PMID: 22818936 PMCID:
17. Kalra K, Moumneh MB, Nanna MG, Damluji AA. Beyond MACE: a multidimensional approach to outcomes in clinical trials for older adults with stable ischemic heart disease. Front Cardiovasc Med. 2023;10:1276370. DOI: https://doi.org/10.3389/fcvm.2023.1276370 PMid:38045910 PMCid:PMC10690830
18. Singh M, Kumar A, Khanna NN, Laird JR, Nicolaides A, et al. Artificial intelligence for cardiovascular disease risk assessment in personalised framework: a scoping review. E Clinical Medicine. 2024 May 27;73:102660. DOI: https://doi.org/10.1016/j.eclinm.2024.102660 PMid:38846068 PMCid:PMC11154124
19. Weng WH, Baur S, Daswani M, Chen C, Harrell L, Kakarmath S, et al. Predicting cardiovascular disease risk using photoplethysmography and deep learning. PLOS Glob Public Health. 2024 Jun 4;4(6):e0003204. DOI: https://doi.org/10.1371/journal.pgph.0003204 PMID: 38833495 PMCID: PMC11149850
20. Asadi F, Homayounfar R, Mehrali Y, Masci C, Talebi S, Zayeri F. Detection of cardiovascular disease cases using advanced tree-based machine learning algorithms. Sci Rep. 2024 Sep 27;14(1):22230. DOI: https://doi.org/10.1038/s41598-024-72819-9 PMid:39333550
21. Meder B, Asselbergs FW, Ashley E. Artificial intelligence to improve cardiovascular population health. Eur Heart J. 2025 May 21;46(20):1907-1916. DOI: https://doi.org/10.1093/eurheartj/ehaf125 PMID: 40106837
22. Agarbattiwala MH, Shaikh AI, Patel P. Comparative Study of Artificial Intelligence-Driven Diagnostic Models and Traditional Risk Scores in Predicting Cardiovascular Events. Eur J Cardiovasc Med. 2025;15(4):323-326. DOI: https://doi.org/10.61336/ejcm/25-04-60