

Estimation of Cardiovascular Risk by Using WHO/ISH charts and Correlation & Regression Tree (CART) Analysis in Adults of Rural North Maharashtra India

Amit P Gujarathi^{1*}, Jagdish Powar², Rakesh Patil³, Shekhar Padhyegurjar⁴,
Ashwini Khadatkhar Ghodake⁵, Moshahed Hussain Mushir Ahmed Shaikh⁶

^{1,2,3}Department of Community Medicine, SMBT Institute of Medical Sciences and Research Centre, Nashik, India

⁴Department of Community Medicine, MGM Medical College, Panvel, Navi Mumbai, India

⁵Department of Pathology, SMBT Institute of Medical Sciences and Research Centre, Nashik, India

⁶SMBT Institute of Medical Sciences and Research Centre, Nashik, India

DOI: 10.55489/njcm.161120255801

ABSTRACT

Background: Cardiovascular diseases (CVDs) are a growing cause of morbidity and mortality, especially in rural India. Accurate risk assessment is crucial for effective primary prevention. This study estimated 10-year CVD risk in rural North Maharashtra using Cholesterol & Non cholesterol based WHO/ISH charts, identified data gaps, assessed agreement between charts, and employed Classification and Regression Tree (CART) analysis.

Methods: The Cross-sectional study conducted among (n=110) rural adults (≥40 years). Data included demographics, smoking, diabetes, Systolic blood pressure (SBP), Fasting blood sugar level (FBS), and Total cholesterol (TC). CVD risk was determined using WHO/ISH charts (with/without cholesterol) for SEAR D countries. CART analysis used to predict the binary risk category (Mild Vs Moderate to severe) derived from Cholesterol based WHO/ISH chart, and Kappa statistics assessed chart agreement.

Results: Participants' mean age was 54.9 years. High prevalence included SBP >130 mmHg (62.7%), abnormal TC (33.6%), and diabetes (25.5%). The WHO/ISH chart classified 37.3% as moderate-to-severe risk. CART analysis identified age, FBS, and TC as key predictors. Moderate agreement (kappa = 0.499) was found between the WHO/ISH charts. Age, diabetes, and TC linked significantly to severe risk.

Conclusion: Rural North Maharashtra carries a significant CVD risk factor burden. Age, fasting blood sugar levels, and total cholesterol were key predictor of moderate to severe risk category. Including cholesterol in risk assessment is vital, as its omission may underestimate risk.

Keywords: CVD, Risk Prediction, WHO/ISH chart, CART analysis, Rural India

ARTICLE INFO

Financial Support: None declared

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

Received: 14-07-2025, **Accepted:** 11-10-2025, **Published:** 01-11-2025

***Correspondence:** Dr. Amit P. Gujarathi (Email: gujrathi.amit@gmail.com)

How to cite this article: Gujarathi AP, Powar J, Patil R, Padhyegurjar S, Ghodake AK, Mushir Ahmed Shaikh MH. Estimation of Cardiovascular Risk by Using WHO/ISH charts and Correlation & Regression Tree (CART) Analysis in Adults of Rural North Maharashtra India. Natl J Community Med 2025;16(11):1142-1151. DOI: 10.55489/njcm.161120255801

Copy Right: The Authors retain the copyrights of this article, with first publication rights granted to Medsci Publications.

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-Share Alike (CC BY-SA) 4.0 License, which allows others to remix, adapt, and build upon the work commercially, as long as appropriate credit is given, and the new creations are licensed under the identical terms.

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

INTRODUCTION

Cardiovascular diseases (CVDs), including myocardial infarction (MI) and stroke, are leading global causes of morbidity and mortality.^{1,2} Their burden is significant in low- and middle-income countries (LMICs) like India, where demographic and epidemiological transitions have increased CVD prevalence.^{3,4} While urban populations have historically faced higher risks, rural areas show a growing burden. Based on the ICMR-INDIAB study, while urban areas have a higher diabetes prevalence at 16.4%, the epidemic is clearly shifting to rural India with a prevalence of 8.9%, alongside a high overall burden of hypertension at 35.5% and dyslipidemia at a staggering 81.2%.⁵ The shift of the non-communicable disease (NCD) epidemic to rural areas is evident, requiring urgent public health initiatives and a reorientation of healthcare priorities with prevalence rates of diabetes and other metabolic NCDs much higher than previously reported, necessitating focused risk assessments.^{5,6} Identifying high-risk individuals is critical for primary prevention, encompassing lifestyle modifications and pharmacological interventions.^{4,6}

The World Health Organization/International Society of Hypertension (WHO/ISH) risk charts, estimating 10-year risk based on age, sex, smoking status, systolic blood pressure (SBP), diabetes, and sometimes total cholesterol (TC), are prediction tools used in India to assess CVD risk.^{2,7,8} Studies in rural India report varying prevalence: 17% of those over 40 in rural South India had moderate-to-high risk⁹; 15.2% of rural Karnataka residents had a $\geq 30\%$ chance of MI or stroke within ten years³; and 44.4% in rural North India had a CVD risk of $\geq 10\%$.⁶ Regional variation is further highlighted by research in Mysuru⁹ and Puducherry.^{2,9} CVD risk also differed among occupational and urban cohorts; for example, urban CVD risk in Tamil Nadu was 25.6%¹⁰, and Bengaluru police officers showed 9.3% high CVD risk¹. A Tamil Nadu study found 1.12% of adult males had $>30\%$ 10-year CVD risk via Framingham score.¹¹ These differences emphasize region-specific evaluations.¹¹

WHO/ISH risk charts are practical for risk stratification, but alternative methodologies like Classification and Regression Trees (CART) offer advantages.^{7,12} CART, a non-parametric statistical learning technique, partitions data into risk categories based on predictor variables, capturing complex interactions traditional models miss. While CVD risk models have been compared across populations^{13,14}, their applicability in diverse Indian cohorts requires further evaluation. Thulani UB et al¹⁵ (2021) found the WHO/ISH (SEAR-B) chart overestimated risk in Sri Lanka, highlighting calibration's importance. Research from 2018 noted wide geographic and socio-demographic variations in India, reinforcing the need for tailored tools.¹⁶

This study aims to fill a critical gap in region-specific

CVD risk estimation. India belongs to the WHO/ISH SEAR D country group (with Sri Lanka, Nepal, Bhutan, Pakistan, Bangladesh). While WHO/ISH charts are widely applied in India, rural North Maharashtra remains understudied. Comparing WHO/ISH (Type D)¹⁷ with CART analysis offers a unique opportunity to explore a data-driven approach's predictive capabilities in this population.

This primary objective was to quantitatively assess the 10-year cardiovascular disease (CVD) risk in a cohort of adults from rural North Maharashtra by using two distinct methodologies. First, the cholesterol-based WHO/ISH risk chart (Type D)¹⁷ was applied to establish baseline risk estimations. Concurrently, the predictive efficacy of the Classification and Regression Tree (CART) model was evaluated. The investigation also sought to determine the level of concordance between the cholesterol-based and non-cholesterol-based WHO/ISH charts. A core hypothesis of the study was that the cholesterol-based model would identify a higher prevalence of at-risk individuals, while the CART model would reveal non-linear interactions among key predictors, offering a more nuanced understanding of CVD risk factors.

The Classification and Regression Tree (CART) analysis excels at uncovering novel, non-linear relationships among risk factors that traditional, more rigid models might miss. Its key advantage over many other machine-learning methods is providing a transparent, visual decision tree that clearly shows how specific risk factor combinations lead to an outcome, making the model's logic both powerful and interpretable in a clinical context.

Integrating insights from both approaches, this study aimed to provide a comprehensive CVD risk assessment, informing targeted prevention strategies in rural North Maharashtra.

METHODOLOGY

Study Design and Setting: This cross-sectional, community-based study was conducted at the Rural Health Training Centre (RHTC) of a tertiary healthcare institute located in Nashik, Maharashtra, India. The RHTC serves the surrounding rural population, providing a representative setting to assess cardiovascular disease (CVD) risk profiles within a distinct regional context.

Study Duration: The study duration was September 2022 to May 2023.

Sample Size and Participant Selection: The target sample size for this study was calculated using OpenEpi software, Version 3. Based on an assumed coronary heart disease prevalence of 13%¹⁸⁻²⁰, a 95% confidence interval, a 5% margin of error, and a design effect of 1, the estimated sample size was 104 participants, which was rounded up to 110 to account for potential non-response or incomplete data.

Participants were selected via simple random sampling from the adult population aged 40 years and above residing in the RHTC's service area. Participants of aged 40 years or older with no prior history of documented cardiovascular events such as stroke, myocardial infarction (MI), ischemic heart disease (IHD), or cardiomyopathy were included. Participants who were not willing to either participate or provide written informed consent were excluded. Ethical approval for the study was obtained from the Institutional Ethics Committee of SMBT, IMSRC, Nashik (IEC approval of protocol no. SMBT/IEC/2019/51, with reference no. 544/SMBT/02/SS/UG/IEC/20/11 dated 18/01/2020). and comprehensive written informed consent was secured from all participants prior to data collection.

Operational Definitions: For consistency and clarity, certain operational definitions were adopted. Diabetes was defined as a fasting plasma glucose level of ≥ 126 mg/dL, a postprandial glucose level of ≥ 140 mg/dL (confirmed on two separate occasions), or current use of insulin or oral hypoglycemic agents. For reliable blood pressure measurement, two readings of systolic blood pressure (SBP) were taken at least five minutes apart. The participant was seated, with their right upper arm at heart level during both readings, and the final SBP value was represented by mean of these two measurements. Participants were classified as smokers if they were current smokers or had ceased smoking within the past year. For Total Cholesterol (TC), levels exceeding 200 mg/dL were considered high. For conversion, mg/dL values were divided by 38 to obtain mmol/L.

Data Collection: Data were systematically collected through a multi-faceted approach. A semi-structured questionnaire, which had been pilot-tested for clarity and validity, was administered to gather demographic information, lifestyle factors, and relevant medical history. Standardized anthropometric measurements were performed, including height (using a stadiometer), weight (using a calibrated digital scale), and waist circumference (using a non-stretchable measuring tape). Blood pressure readings were obtained using a calibrated sphygmomanometer, with measurements taken after a 5-minute rest period in a seated position.

For biochemical analyses, all 110 participants were asked to observe an 8-hour overnight fast before venous blood samples drawn. These samples were used to measure fasting blood sugar and total cholesterol levels. Cholesterol testing was specifically performed using the enzymatic CHOD-POD method. To ensure accuracy and consistency, all blood analyses were conducted at a NABL certified laboratory affiliated with the tertiary healthcare institute.

Cardiovascular Risk Assessment: The 10-year risk of myocardial infarction (MI) and stroke for all 110 participants was estimated using the both WHO/ISH risk prediction charts specifically designed for the South-East Asia Region, Type D (SEAR-D region)¹⁷.

Two versions of the WHO/ISH charts were utilized:

Basic WHO/ISH Chart: This chart assesses CVD risk based on a set of core parameters: age, sex, systolic blood pressure (SBP), smoking status, and diabetes status. (Figure 1)

Cholesterol-Based WHO/ISH Chart: This advanced chart incorporates all the parameters of the basic chart but additionally includes serum total cholesterol levels, providing a more comprehensive risk assessment. (Figure 2)

The risk estimation process was conducted through a standardized, stepwise methodology. The appropriate risk chart was first selected according to the participant's diabetes status, with distinct charts applied for diabetic and non-diabetic individuals. Within each chart, gender-specific tables were employed to ensure precision in risk calculation. Participants' smoking status was identified and categorized as current smoker, recent quitter (within the past year), or non-smoker. Age was classified into predefined groups (40-49, 50-59, and 60-69 years) based on chart-specific criteria. Systolic blood pressure (SBP) was measured in the seated position, and the corresponding SBP category on the chart was used for risk determination. For cholesterol-based charts, serum total cholesterol concentrations were integrated into the calculation to refine cardiovascular risk estimation.

CVD risk classification was determined by color-coded categories as per the WHO/ISH charts: Green (<10% risk), Yellow (10% to <20% risk), Orange (20% to <30% risk), Red (30% to <40% risk), and Deep Red (>40% risk).

Statistics: All data collected were meticulously entered and analyzed using SPSS statistical software, Version 22. Descriptive statistics, including frequencies and percentages, were employed to characterize the distribution of risk factors and the 10-year cardiovascular disease (CVD) risk categories within the study population. Pearson's chi-square test was applied to assess associations between categorical risk factors and CVD risk classifications. The agreement between the cholesterol and non-cholesterol based WHO/ISH charts was evaluated using Kappa statistics. A p-value of less than 0.05 ($p < 0.05$) was considered to indicate statistical significance. Importantly, there were no missing data, and all 110 participants were included in the final analysis.

Classification and Regression Tree (CART) Analysis Model: The data for CART analysis was imported meticulously and analyzed using Minitab® Version 22 with inbuilt Classification and Regression Tree (CART) algorithm. Furthermore, CART analysis was conducted using specialized statistical software which randomly partitioned the dataset into 80% for training and 20% for testing. This analysis was based on the variables from the cholesterol-based WHO/ISH chart, with risk categories dichotomized

into mild versus moderate to severe CVD risk, enabling a detailed evaluation of predictive associations between individual risk factors and the primary study outcomes. The outcome variable was defined as a binary classification, categorizing participants into two distinct groups: mild CVD risk versus moderate-to-severe CVD risk. The CART algorithm iteratively partitioned the dataset into increasingly homogeneous subsets by identifying the optimal split points for predictor variables. This process, driven by the Gini index, ensured that each node in the tree

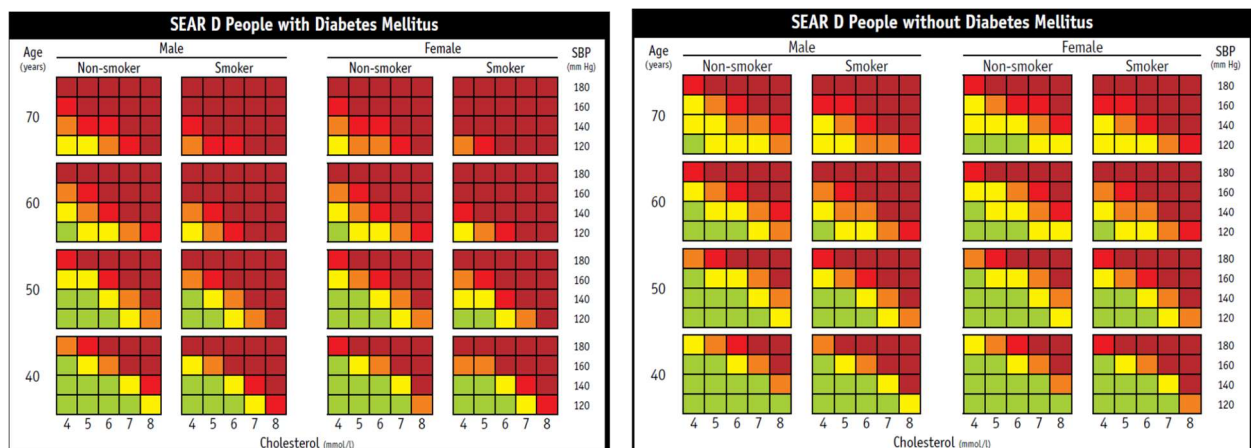
optimally divided the outcome classes. This inbuilt CART model uses 10-fold cross validation which systematically divides data into 10 subsets, where each fold served as a test set against the remaining nine folds used for training.

The dataset was Model performance was comprehensively evaluated using a range of metrics, including the confusion matrix (Sensitivity, Specificity) and the Area Under the Receiver Operating Characteristic (ROC) curve.



Green (<10% risk), Yellow (10% to <20% risk), Orange (20% to <30% risk), Red (30% to <40% risk), and Deep Red (>40% risk).

Figure 1: WHO-ISH Risk Prediction chart (non-cholesterol based)



Green (<10% risk), Yellow (10% to <20% risk), Orange (20% to <30% risk), Red (30% to <40% risk), and Deep Red (>40% risk).

Figure 2: WHO-ISH Risk Prediction chart based on Cholesterol level

RESULTS

A total of 110 participants were enrolled, with a mean age of 54.97 (± 10.75) years. The majority (73.6%) were aged 41-60 years, and males constituted 70.9% of the study population. Elevated systolic blood pressure (≥ 130 mmHg) was observed in 62.7% of participants, while 17.3% were current smokers. Abnormal total cholesterol (> 5.16 mmol/L) and fasting blood sugar (> 126 mg/dL) were reported in 33.6% and 21.8% of participants, respectively.

Postprandial hyperglycemia was present in 66.4%, and 25.5% reported a family history of diabetes mellitus. These findings indicate that the study population carried a substantial cardiometabolic burden (Table 1).

Risk stratification by WHO/ISH charts is presented in Table 2. Using the cholesterol-based chart, 37.3% of participants were classified into the moderate-severe risk category, whereas the non-cholesterol-based chart identified only 24.6% in this group.

Table 1: Baseline characteristics of study participants (N = 110)

Variable	Participants (%)
Age group (years)	
40-50	45 (40.9)
51-60	36 (32.7)
61-70	20 (18.2)
>70	9 (8.2)
Median (Mean $\pm$ SD)	53 (54.97 $\pm$ 10.75)
Gender	
Female	32 (29.1)
Male	78 (70.9)
Systolic BP (mm Hg)	
≤ 120	23 (20.9)
121-129	18 (16.4)
130-139	35 (31.8)
140-179	34 (30.9)
Smoking status	
No	91 (82.7)
Yes	19 (17.3)
Total cholesterol	
Normal (≤ 5.16 mmol/L)	73 (66.4)
Abnormal (> 5.16 mmol/L)	37 (33.6)
Median (Mean $\pm$ SD)	4.99 $\pm$ 1.22
Fasting blood sugar level	
Normal (< 126 mg/dl)	86 (78.2)
Abnormal (> 126 mg/dl)	24 (21.8)
Median (Mean $\pm$ SD)	116 (115.57 $\pm$ 12.61)
H/O Diabetes mellitus	
No	82 (74.5)
Yes	28 (25.5)

Table 2: Cardiovascular disease (CVD) risk categories by WHO/ISH charts with and without cholesterol (N =110)

Risk Categories	With cholesterol	Without cholesterol
Risk Category		
Green ($< 10\%$)	69 (62.7)	83 (75.5)
Yellow (10- $< 20\%$)	19 (17.3)	22 (20.0)
Orange (20- $< 30\%$)	15 (13.6)	4 (3.6)
Red (30- $< 40\%$)	2 (1.8)	0 (0.0)
Deep Red ($\geq 40\%$)	5 (4.5)	1 (0.9)
Risk Category (Grouped)		
Mild ($< 10\%$)	69 (62.7)	83 (75.5)
Moderate to Severe ($\geq 10\%$)	41 (37.3)	27 (24.6)

Figures in parenthesis indicate percentage

Table 3: Agreement analysis between WHO ISH risk chart with and without Cholesterol

Without Cholesterol	With cholesterol		Total
	Mild ($< 10\%$)	Moderate to Severe ($\geq 10\%$)	
Mild ($< 10\%$)	64	19	83
Moderate-Severe ($\geq 10\%$)	5	22	27
Total	69	41	110

Note: Observed agreement = 86 (78.2%); agreement expected by chance = 62.1 (56.5%); Kappa = 0.499 (moderate agreement); SE = 0.085; 95% CI = 0.332-0.666.

The cholesterol-based chart categorized more individuals into high-risk strata, including 4.5% in the very high-risk ($\geq 40\%$) category compared to 0.9% by the non-cholesterol chart. Overall, inclusion of cholesterol led to an absolute increase of 12.7% in participants labeled at higher CVD risk, suggesting possible overestimation.

Agreement analysis between the two charts is shown in Table 3. Out of 110 participants, 86 (78.2%) had concordant risk classification, while 24 were discordant. Cohen's Kappa statistic was 0.499 (95% CI: 0.332-0.666), reflecting moderate agreement. This suggests that although broadly comparable, the non-cholesterol chart may underestimate risk relative to the cholesterol-based chart.

Age, diabetes, fasting blood sugar, and total cholesterol were significantly associated with higher CVD risk. The prevalence of moderate-severe risk increased from 17.8% in the 40-50-year group to 88.9% in those > 70 years ($\chi^2=22.1$, $p<0.001$). Participants with history of diabetes (75.0%), abnormal fasting blood sugar (45.3%), and elevated cholesterol (70.3%) had markedly higher risk (all $p<0.001$). Postprandial hyperglycemia showed a non-significant trend toward increased risk ($p=0.06$). Gender and smoking were not associated with risk (Table 3).

The CART model identified fasting blood sugar (FBS) as the primary splitting variable (threshold 87 mg/dL). Individuals with FBS ≤ 87 mg/dL were further categorized by age: those ≤ 67 years predominantly classified as "Mild" risk (98.1%), while individuals > 67 years showed significantly higher risk, with 63.6% falling into the "Moderate to Severe" category (Figure 3).

Among individuals with FBS > 87 mg/dL, total cholesterol (TC) was the next key determinant. If TC exceeded 5.603 mmol/L, all individuals (100%) were classified as "Moderate to Severe" risk. For those with TC ≤ 5.603 mmol/L, age again played a discriminatory role: individuals ≤ 63.5 years underwent an additional FBS-based split, while those > 63.5 years were uniformly categorized as "Moderate to Severe" risk.

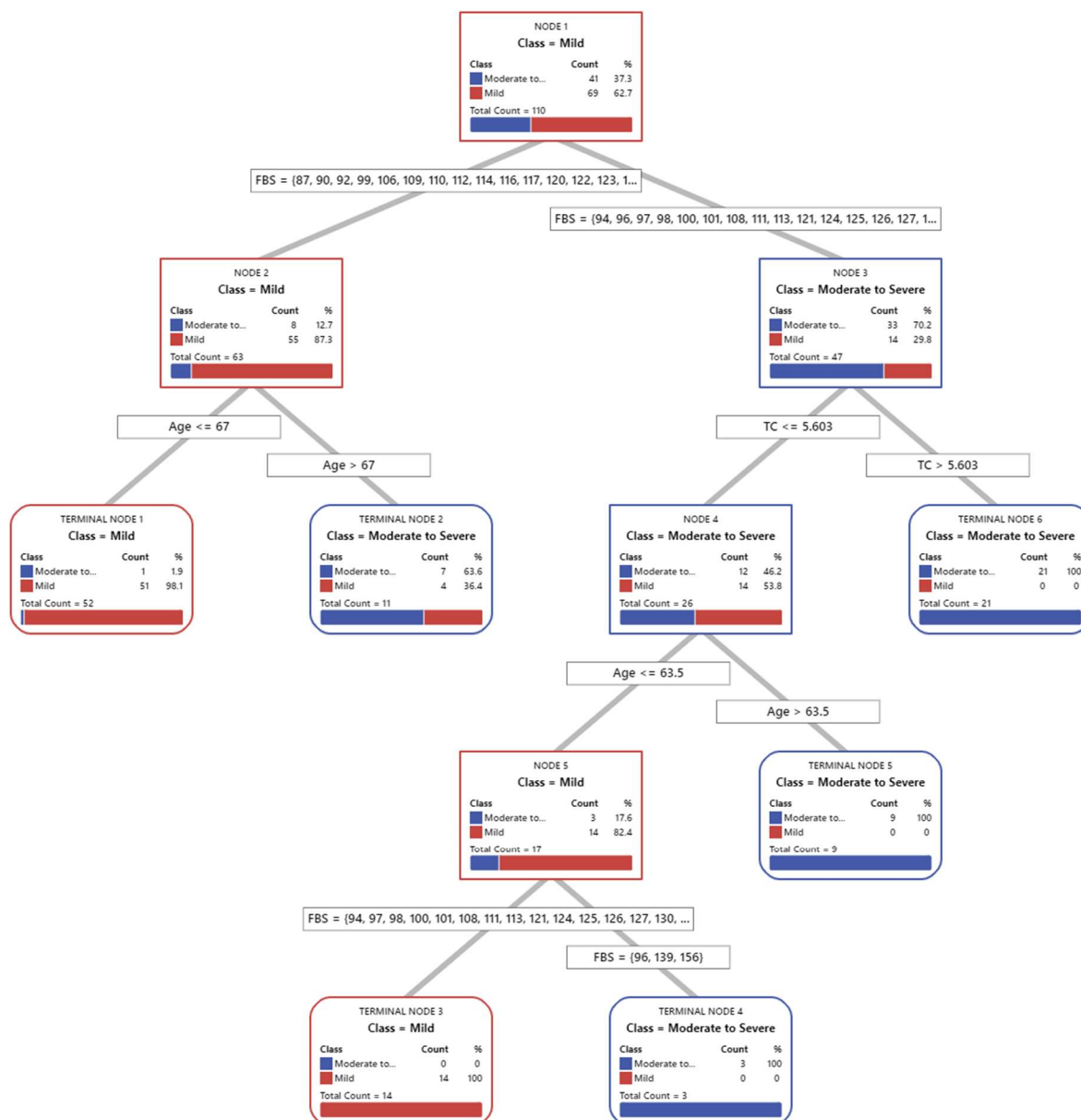
The confusion matrix in Table 5 details the model's predictive accuracy, showing a high overall correctness of 95.5% on the training set and 71.8% on the independent test set. Specifically, the model accurately predicted 97.6% of the moderate-to-severe cases in the training set, although this dropped to 61% on the test set.

Table 6 provides key performance metrics in which the true positive rate or sensitivity for moderate-to-severe CVD risk, was 97.6% in training but decreased to 61% in testing, indicating some overfitting. Conversely, the true negative rate or specificity for mild CVD risk, was robust at 94.2% in training and remained relatively high at 78.3% in the test set.

Table 4: Association of Risk factors with CVD Risk categories

Variable	Mild (n=69) (%)	Moderate-Severe (n=41) (%)	χ^2 value	p-value
Age (years)				
40-50	37 (82.2)	8 (17.8)	22.1	<0.001*
51-60	23 (63.9)	13 (36.1)		
61-70	8 (40.0)	12 (60.0)		
>70	1 (11.1)	8 (88.9)		
Gender				
Female	21 (65.6)	11 (34.4)	0.162	0.687
Male	48 (61.5)	30 (38.5)		
Smoker	10 (52.6)	9 (47.4)	1.001	0.317
Fasting Blood sugar level				
Normal <126 mg/dl	22 (91.7)	2 (8.3)	10.8	<0.001
Abnormal >126mg/dl	47 (54.7)	39 (45.3)		
Post prandial Blood sugar level				
Normal	28 (75.7)	9 (24.3)	3.57	0.059
Abnormal	41 (56.2)	32 (43.8)		
H/O Diabetes (DM)	7 (25.0)	21 (75.0)	22.867	<0.001*
Total cholesterol				
Normal (≤ 5.16 mmol/L)	58 (79.5)	15 (20.5)	25.96	<0.001*
Abnormal (>5.16 mmol/L)	11 (29.7)	26 (70.3)		

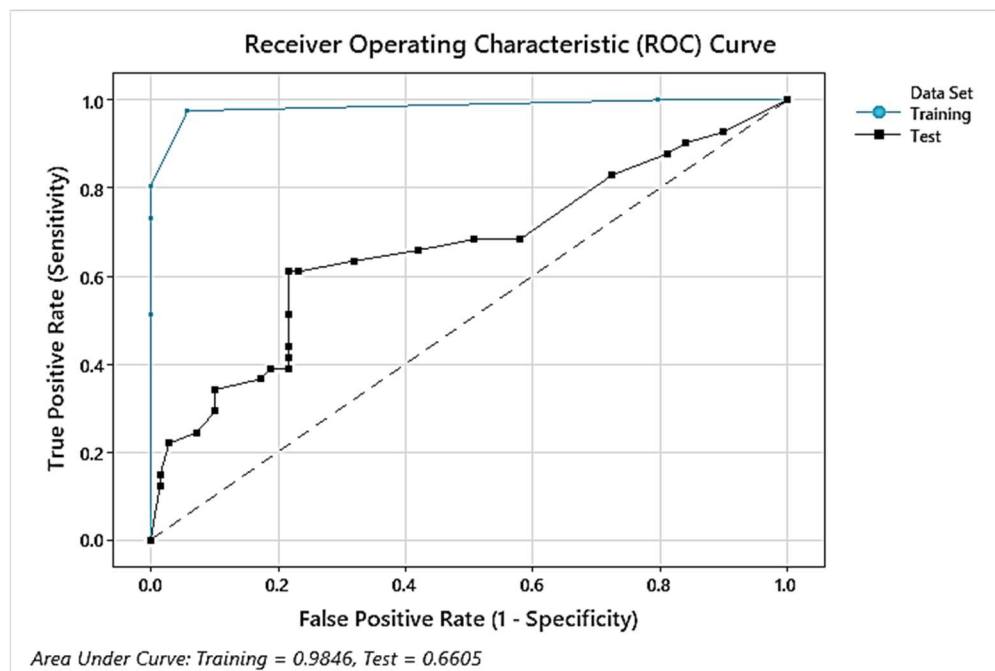
Mild = <10% 10-year CVD risk; Moderate-Severe = $\geq 10\%$ 10-year CVD risk (WHO/ISH chart with cholesterol).



Red: Mild Risk category, Blue: Moderate to High-risk category
Figure 3: CART Model for Cardiovascular Risk Prediction

Table 5: The CART Model Analysis for CVD Risk

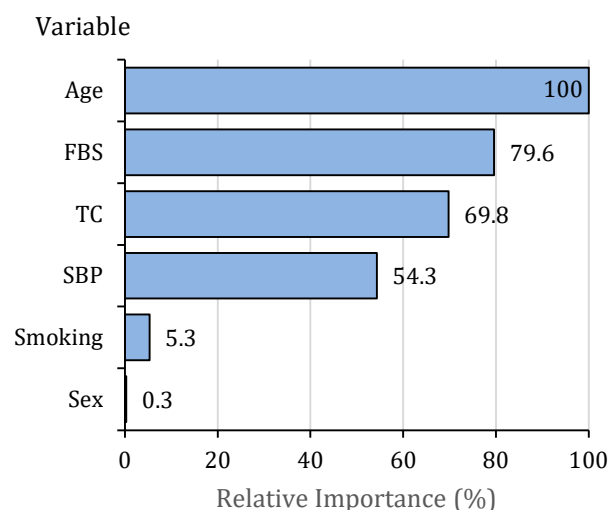
Category of CVD	Predicted class (Training set)			Predicted class (Test set)		
	Moderate to severe	Mild	% Correct	Moderate to severe	Mild	% Correct
Moderate to Severe Event	41	1	97.60%	25	16	61.00%
Mild Event	69	65	94.20%	15	54	78.30%
Total	110	66	95.50%	40	70	71.80%

**Figure 4: Receiver Operating Characteristic (ROC) Curve****Table 6: Performance Matrix of CART Model for CVD Risk**

Statistics	Training (%)	Test (%)
True positive rate (sensitivity or power)	97.6	61
False positive rate (type I error)	5.8	21.7
False negative rate (type II error)	2.4	39
True negative rate (specificity)	94.2	78.3

As depicted in figure no. 4, training data set shows an excellent fit with an Area Under Curve (AUC) of 0.9846, while the test data set's curve indicates a much lower performance with an AUC of 0.6605.

As depicted in Figure 5, Age emerged as the most influential predictor (100%), followed by FBS (79.6%) and TC (69.8%). SBP also contributed significantly (54.3%), while smoking (5.3%) and sex (0.3%) had minimal impact.

**Figure 5: Relative Variable importance for CVD risk**

DISCUSSION

Our study highlights a significant burden of cardiovascular disease (CVD) risk factors among adults in rural North Maharashtra in middle-aged males 54.97 ± 10.75 years. This demographic is highly susceptible to CVD, as validated by prior studies using risk prediction tools.^{19,20}

The prevalence of modifiable risk factors is substantial. A 62.7% had SBP >130 mmHg, indicating widespread hypertension. It is consistent with earlier studies identifying hypertension as a major CVD contributor for CVD among Indian population.^{11,21} Additionally, 33.6% had abnormal cholesterol levels, and 25.5% had diabetes, reinforcing the clustering of metabolic risk factors. These findings align with previous research emphasizing the compounded impact of multiple risk factors on CVD.^{7,8} Moreover, a signifi-

cant portion of the participants had multiple cardiovascular risk factors, highlighting the relevance of targeted prevention strategies in this population.

A CVD risk prediction chart from the WHO/ISH that included total cholesterol (TC) classified 37.3% of individuals as having a moderate-to-high risk for a CVD event within the next decade. When cholesterol data was excluded, this figure dropped to 24.5%, underscoring the importance of lipid profiles for accurate risk estimation. Similar studies in rural India reported that moderate-to-high-risk prevalence of CVD ranging from 17% to 44.4%, influenced by demographic and lifestyle variations.^{6,8,22} The differences in reported risk levels may stem from different study population characteristics and the choice of WHO sub-region charts.⁶

Our findings reinforce the practicality of WHO/ISH risk charts in resource-limited settings like rural North Maharashtra.^{2,6-8,21} The SEAR-D region-specific charts offer a feasible risk stratification tool.²⁰ However, limitations in these charts encompasses the omission of key risk factors relevant to South Asian populations, such as abdominal obesity, family history of CVD, high salt intake, and treatment status of non-communicable diseases. Concerns have also been raised about their accuracy in certain populations.^{8,11,23} Thus, determining the most suitable risk assessment model for Indian populations remains an area of ongoing research.¹¹

Agreement analysis between WHO/ISH charts with and without cholesterol showed moderate consistency ($\kappa = 0.499$), with 78.18% classification agreement, with 0.085 SE (κ), and 0.332 to 0.666 95% CI. This suggests that while both charts are useful, incorporating cholesterol may improve cardiovascular risk assessment, particularly for individuals in intermediate-risk categories.

Excluding cholesterol underestimated CVD risk in a substantial proportion of individuals, echoing findings from previous studies demonstrating poor agreement between these charts.¹⁶ Some studies report 80.3% agreement with a κ of 0.429, particularly noting discrepancies in higher-risk categories.²⁴ Research from other groups indicate that excluding cholesterol led to a greater proportion of individuals being categorized as low risk.²⁵ These results strengthen the recommendation to include cholesterol in risk assessments especially in rural setting whenever possible for accurate CVD risk classification.^{26,27}

In this study, age, diabetes, fasting blood sugar, and total cholesterol all demonstrated strong associations with moderate to severe CVD risk, while gender and smoking status were not statistically significant. This pattern is consistent with other published research, which has similarly found stronger links between metabolic and non-modifiable factors and CVD risk, with weaker or non-significant associations for gender and smoking in some populations. The lack of association with smoking in our study

may be attributed to the lower number of smokers in the studied population or the presence of unmeasured confounders. Additionally, the predominance of smokeless tobacco uses in rural India, as highlighted in national surveys, may further explain this observation.²⁶ While some studies report gender differences in CVD risk^{3,6,27}, smoking remains a well-established risk factor globally.^{2,6,23,20} The figure 5 highlights the relative importance of variables such as higher FBS, elevated TC, and older age which significantly increase the likelihood of severe cardiovascular risk, recognizing their importance in early screening, targeted interventions, and prevention strategies for high-risk populations.

Performing CART analysis using the same parameters as the cholesterol based WHO/ISH risk charts offered a novel, data-driven approach to cardiovascular risk stratification. The CART model effectively identified fasting blood sugar (FBS), total cholesterol (TC), and age as key predictors. Individuals with lower FBS and younger age were predominantly classified in the "Mild" risk group, whereas higher FBS, TC, and advanced age led to consistent categorization under "Moderate to Severe" risk. Unlike the WHO/ISH charts, which provide standardized estimates, the CART model captured complex, non-linear interactions and allowed hierarchical risk interpretation. This enhances the clinical utility of risk prediction by offering greater interpretability and precision.

CART identified fasting blood sugar (FBS), TC, and age as primary risk predictors. Its hierarchical classification revealed that individuals with FBS >87 mg/had significantly higher risk, especially if TC exceeded 5.603 mmol/L. Age also played a crucial role, with those >67 years classified predominantly in the moderate to severe risk category. This hierarchical decision-making framework provides a clear, clinically interpretable structure for risk stratification in CVD prediction. This aligns with prior studies recognizing age as a strong non-modifiable CVD predictor.^{1-3,18,22,28}

Our study found that the CART model performed well on the training data, with an AUC of 0.98, but its performance dropped significantly on the independent test set, yielding an AUC of 0.6605. In comparison, advanced machine learning models have demonstrated superior predictive power in larger populations. A study done in China, showed a random forest model achieving a superior AUC of 0.787 for cardiovascular disease (CVD) prediction, significantly outperforming a multivariate regression model which had an AUC of 0.7143.²⁹ The overfitting observed in the CART model, as indicated by the performance discrepancy, was likely due to our relatively small dataset and its high variability. Despite this limitation, the CART algorithm remains a valuable tool because of its strong interpretability and ability to stratify risk in a clinically meaningful way.

This community-based study focused on a rural population in North Maharashtra, a region largely underrepresented in cardiovascular disease (CVD) research. By employing both the WHO/ISH risk prediction charts and the CART model, the study provided a comprehensive comparison between conventional and machine learning-based approaches. This dual-method evaluation offered context-specific insights which adds to CVD risk stratification and helpful for policy makers for future targeted prevention strategies especially in rural settings.

LIMITATIONS

The WHO/ISH risk charts used may not fully capture important risk factors specific to South Asian populations, such as abdominal obesity and family history, which could influence the accuracy of risk predictions. Additionally, questions remain about the general calibration and applicability of these charts across diverse populations. The cross-sectional design, relatively small sample size, and sampling from a single rural health training center limit the representativeness of the findings and introduce potential selection bias. Furthermore, reliance on self-reported data for smoking and diabetes history may have introduced recall bias, affecting data validity.

CONCLUSION

The present study underscores a substantial burden of cardiovascular risk among adults in rural North Maharashtra. Age, fasting blood sugar, and total cholesterol were the most significant predictors of moderate to severe 10-year CVD risk. Inclusion of cholesterol in risk assessment improved accuracy compared to non-cholesterol-based charts. CART analysis provided additional insights into non-linear predictor interactions, enhancing interpretability and individualized risk estimation. Strengthening community-based screening, lifestyle modification, and targeted prevention initiatives is imperative to mitigate the rising CVD risk in rural populations.

PUBLIC HEALTH IMPLICATION

Incorporating cholesterol testing and data-driven analytical models such as CART into primary care settings can substantially improve early detection and risk stratification of cardiovascular disease in resource-limited rural regions of India.

RECOMMENDATIONS

Future research should involve multicentric trials with larger and more diverse populations to improve generalizability. It is also recommended to perform external validation of classification and regression tree (CART) models using independent cohorts to assess their predictive accuracy in various settings.

Individual Authors' Contributions: The study was conceptualized and methodologically designed by **AG**, who also prepared the original draft and managed the overall project. **JP** was responsible for project data collection and performed the CART analysis. **RP** contributed to resource arrangements and data interpretation. **SP** handled data validation and software management. **AK-G** supervised and monitored the laboratory investigations, ensuring the quality of research procedures. **MS** undertook the primary data collection, supporting the core data acquisition process of the study.

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. Geethu S, Jadhav J, Ranganath TS. Cardiovascular disorder risk assessment among police personnel in Bengaluru City, India, using World Health Organization/International Society of Hypertension risk prediction chart. *Cureus*. 2023 Nov 6;15(11):e48378. DOI: <https://doi.org/10.7759/cureus.48378> PMID: 38060764 PMCID: PMC10699544
2. Mohamed SM, Anandaraj R, Sivasubramanian V. Assessment of ten-year risk of cardiovascular event using WHO/ISH risk prediction chart among adults in a tertiary care hospital in Puducherry, India. *J Med Sci Res*. 2021;9(2):96-100.
3. Norman G, George CE, Krishnamurthy A, Mukherjee D. Burden of cardiovascular risk factors of a rural population in South India using the WHO multivariable risk prediction algorithm. *Int J Med Sci Public Health*. 2014;3(6):764-768. DOI: <https://doi.org/10.5455/ijmsph.2014.180320141>
4. Rezaei F, Seif M, Fattahi MR, Gandomkar A, Hasanzadeh J. Estimation of 10-Year Risk of Cardiovascular Diseases Using WHO Risk Prediction Charts: A Population-Based Study in Southern Iran. *Iran J Public Health*. 2022 Jul;51(7):1667-1676. DOI: <https://doi.org/10.18502/ijph.v51i7.10101> PMID: 36248300 PMCID: PMC9529739
5. Anjana RM, Unnikrishnan R, Deepa M, Pradeepa R, Tandon N, Das AK, et al. Metabolic non-communicable disease health report of India: the ICMR-INDIAB national cross-sectional study (ICMR-INDIAB-17). *Lancet Diabetes Endocrinol*. 2023;11(7):474-489. DOI: [https://doi.org/10.1016/S2213-8587\(23\)00119-5](https://doi.org/10.1016/S2213-8587(23)00119-5) PMID: 37301218
6. Bansal P, Chaudhary A, Wander P, Satija M, Sharma S, Girdhar S, et al. Cardiovascular Risk Assessment Using WHO/ISH Risk Prediction Charts In a Rural Area of North India. *J Res Med Dent Sci*. 2016;4(2):167-172. DOI: <https://doi.org/10.5455/jrmds.20164210>
7. Ananda Selva Das P, Dubey M, Kaur R, Salve HR, Varghese C, Nongkynrih B. WHO Non-Lab-Based CVD Risk Assessment: A Reliable Measure in a North Indian Population. *Global Heart*. 2022;17(1):64. DOI: <https://doi.org/10.5334/gh.1148> PMID: 36199565 PMCID: PMC9438460
8. Ghorpade AG, Shrivastava SR, Kar SS, Sarkar S, Majgi SM, Roy G. Estimation of the cardiovascular risk using World Health Organization/International Society of Hypertension (WHO/ISH) risk prediction charts in a rural population of South India.

- Int J Health Policy Manag. 2015;4(8):531-536. DOI: <https://doi.org/10.15171/ijhpm.2015.88>
9. Kadiyala P, Renuka M, Kulkarni P, Narayanamurthy MR. Prevalence of risk factors and 10 year risk estimation of cardiovascular diseases among rural population of Mysuru, Karnataka. *Int J Community Med Public Health*. 2019;6(3):1178-1185. DOI: <https://doi.org/10.18203/2394-6040.ijcmph20190607>
 10. Premanandh K, Shankar R. Predicting 10-year cardiovascular risk using WHO/ISH risk prediction chart among urban population in Salem. *Int J Community Med Public Health*. 2018;5(12):5228-5234. DOI: <https://doi.org/10.18203/2394-6040.ijcmph20184795>
 11. Sasikumar M, Marconi SD, Dharmaraj A, Mehta K, Das M, Goel S. Prevalence of risk factors and estimation of 10-year risk for cardiovascular diseases among male adult population of Tamil Nadu India-an insight from the National Family Health Survey-5. *Indian Heart J*. 2023;75(4):251-257. DOI: <https://doi.org/10.1016/j.ihj.2023.06.003> PMID: 37336261
 12. WHO CVD Risk Chart Working Group. World Health Organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions. *Lancet Glob Health*. 2019 Oct;7(10):e1332-e1345. DOI: [https://doi.org/10.1016/S2214-109X\(19\)30318-3](https://doi.org/10.1016/S2214-109X(19)30318-3) Erratum in: *Lancet Glob Health*. 2023 Feb;11(2):e196. DOI: [https://doi.org/10.1016/S2214-109X\(22\)00522-8](https://doi.org/10.1016/S2214-109X(22)00522-8) PMID: 31488387 PMCID: PMC7025029
 13. Momayyezi M, Sefidkar R, Fallahzadeh H. Agreement between ten-years cardiovascular disease risk assessment tools: An application to Iranian population in Shahedieh Cohort Study. *Heliyon*. 2023;9(10):e20396. DOI: <https://doi.org/10.1016/j.heliyon.2023.e20396> PMID: 37810856 PMCID: PMC10556586
 14. Selvarajah S, Kaur G, Haniff J, Cheong KC, Hiong TG, Van Der Graaf Y, et al. Comparison of the Framingham Risk Score, SCORE and WHO/ISH cardiovascular risk prediction models in an Asian population. *Int J Cardiol*. 2014 Sep;176(1):211-218. DOI: <https://doi.org/10.1016/j.ijcard.2014.07.066>
 15. Thulani UB, Mettananda KCD, Warnakulasuriya DTD, Peiris TSG, Kasturiratne KTAA, Ranawaka UK, et al. Validation of the World Health Organization/International Society of Hypertension (WHO/ISH) cardiovascular risk predictions in Sri Lankans based on findings from a prospective cohort study. *PLoS One*. 2021 Jun 7;16(6):e0252267. DOI: <https://doi.org/10.1371/journal.pone.0252267> PMID: 34097699
 16. Babatunde OA, Olarewaju SO, Adeomi AA, Akande JO, Bashorun A, Umeokonkwo CD, et al. 10-year risk for cardiovascular diseases using WHO prediction chart: findings from the civil servants in South-western Nigeria. *BMC Cardiovasc Disord*. 2020 Dec;20(1):154. DOI: <https://doi.org/10.1186/s12872-020-01438-9> PMID: 32234017 PMCID: PMC7110661
 17. World Health Organization. WHO cardiovascular disease risk chart working group. World Health Organization CVD risk charts. Geneva: WHO; <https://www.who.int/news/item/02-09-2019-who-updates-cardiovascular-risk-charts> . [Accessed 23 June 2025]
 18. Mohan V, Deepa R, Rani SS. Prevalence of coronary artery disease and its relationship to lipids in a selected population in South India: The Chennai Urban Population Study (CUPS No. 5). *J Am Coll Cardiol*. 2001;38(3):682-687. DOI: [https://doi.org/10.1016/S0735-1097\(01\)01415-2](https://doi.org/10.1016/S0735-1097(01)01415-2) PMID: 11527617
 19. Gupta R, Gupta VP, Sarna M, Bhatnagar S, Thanvi J, Sharma V, Singh AK, Gupta JB, Kaul V. Prevalence of coronary heart disease and risk factors in an urban Indian population: Jaipur Heart Watch-2. *Indian Heart J*. 2002 Jan-Feb;54(1):59-66. PMID: 11999090.
 20. Kamili M, Dar I, Ali G, Wazir H, Hussain S. Prevalence of coronary heart disease in Kashmiris. *Indian Heart J*. 2007 Jan-Feb;59(1):44-49. PMID: 19098334.
 21. Deori TJ, Agarwal M, Masood J, Sharma S, Ansari A. Estimation of cardiovascular risk in a rural population of Lucknow district using WHO/ISH risk prediction charts. *J Family Med Prim Care*. 2020 Sep 30;9(9):4853-4860. DOI: https://doi.org/10.4103/jfmpc.jfmpc_646_20 PMID: 33209812
 22. Islam JY, Zaman MM, Moniruzzaman M, Ara Shakoor S, Hossain AHME. Estimation of total cardiovascular risk using the 2019 WHO CVD prediction charts and comparison of population-level costs based on alternative drug therapy guidelines: a population-based study of adults in Bangladesh. *BMJ Open*. 2020 Jul 19;10(7):e035842. DOI: <https://doi.org/10.1136/bmjopen-2019-035842> PMID: 32690512
 23. Raghu A, Praveen D, Peiris D, Tarassenko L, Clifford G. Implications of cardiovascular disease risk assessment using the WHO/ISH risk prediction charts in rural India. *PLoS One*. 2015 Aug 19;10(8):e0133618. DOI: <https://doi.org/10.1371/journal.pone.0133618> PMID: 26287807 PMCID: PMC4545825
 24. Otgontuya D, Oum S, Buckley BS, Bonita R. Assessment of total cardiovascular risk using WHO/ISH risk prediction charts in three low and middle income countries in Asia. *BMC Public Health*. 2013 Dec;13(1):539. DOI: <https://doi.org/10.1186/1471-2458-13-539> PMID: 23734670 PMCID: PMC3679976
 25. Khanal MK, Ahmed MSAM, Moniruzzaman M, Banik PC, Dhungana RR, Bhandari P, et al. Total cardiovascular risk for next 10 years among rural population of Nepal using WHO/ISH risk prediction chart. *BMC Res Notes*. 2017 Dec;10(1):120. DOI: <https://doi.org/10.1186/s13104-017-2436-9>
 26. Rajdeep PS, Shigwan SR, Gera M. Prevalence of smoking in rural and urban areas in India: systematic review. *Int J Health Sci*. 2022;6(S3):6606-6616. DOI: <https://doi.org/10.53730/ijhs.v6nS3.7472>
 27. Gaikwad A, Khan Y. Evaluation of Discordance between 10 year Cardiovascular Risk Scores in Indian Patients presenting with Myocardial infarction. *Cardiology and Cardiovascular Medicine* 3 (2019):360-368. DOI: <https://doi.org/10.26502/fccm.92920085>
 28. Sitaula D, Dhakal A, Mandal SK, Bhattarai N, Silwal A, Adhikari P, et al. Estimation of 10-year cardiovascular risk among adult population in western Nepal using nonlaboratory-based WHO/ISH chart: a cross-sectional study. *Health Sci Rep*. 2023 Oct;6(10):e1614. DOI: <https://doi.org/10.1002/hsr2.1614> PMID: 37818312 PMCID: PMC10560824
 29. Yang L, Wu H, Jin X, Zheng P, Hu S, Xu X, et al. Study of cardiovascular disease prediction model based on random forest in eastern China. *Sci Rep*. 2020 Mar 23;10(1):5245. DOI: <https://doi.org/10.1038/s41598-020-62133-5>