Athens Journal of Demography & Anthropology –
Volume 1, Issue 3, September 2025 – Pages 193-208

Determinants of Cardiovascular Disease in Sri Lanka: Risk Assessment

By Hansa Jayarathne & Liwan Liyanage±*

Non-communicable diseases (NCDs) have become a leading public health concern, contributing to significant morbidity and mortality worldwide. Among them, cardiovascular diseases (CVDs) are the most prevalent, accounting for 34% of total deaths in Sri Lanka. This research employs a quantitative research approach to analyze the demographic, behavioral, and socioeconomic factors influencing CVD prevalence in Sri Lanka. Using secondary data from the Ministry of Health and the 2021 Non-Communicable Diseases Risk Factor Survey, the study examines key variables to identify the associated risk. The findings indicate that demographic factors, including age and male sex, significantly increase CVD risk, while higher household income appears to have a protective effect. Regional disparities in CVD prevalence highlight the need for targeted interventions in high-risk areas. Lifestyle factors such as consumption of alcohol, tobacco use, physical inactivity, high salt and sugar intake, and diabetes history emerged as major contributors to CVD risk. The predictive model used in this study demonstrated strong capabilities in identifying high-risk individuals, making it a valuable tool for public health planning. The study underscores the importance of early screening, lifestyle modifications, and region-specific policy interventions to mitigate the growing burden of CVDs in Sri Lanka.

Keywords: *Non-Communicable Diseases (NCDs), cardiovascular disease (CVD), Risk Factors, Sri Lanka, Public Health*

Introduction

Non-communicable diseases, also referred to as chronic diseases, are long-term conditions that progress slowly and are not transmissible between individuals (WHO 2021). The primary categories of NCDs include cardiovascular diseases (e.g., heart attacks and strokes), cancers, and diabetes. These diseases have emerged as leading contributors to morbidity and mortality worldwide, presenting a significant public health challenge. Between 1990 and 2015, NCDs accounted for approximately 41 million deaths annually, constituting 71% of global mortality (WHO 2021). Notably, 16 million of these deaths occurred before the age of 70, with 82% concentrated in low- and middle-income countries. Cardiovascular diseases are the most prevalent and responsible for 17.9 million deaths each year, followed by cancers (9.3 million) and diabetes (1.5 million) (WHO 2021).

*PhD Scholar/Casual Academic, School of Computer, Data and Mathematical Sciences, Western Sydney University and Lecturer, University of Colombo, Australia.
±Senior Lecturer, School of Computer, Data and Mathematical Sciences, Western Sydney University, Australia.

https://doi.org/10.30958/ajda.1-3-3 doi=10.30958/ajda.1-3-3

Regions with underdeveloped healthcare infrastructure, such as parts of Africa, exhibit the highest rates of premature deaths due to NCDs. In contrast, developed regions, including North America, Europe, and Australia, demonstrate lower premature mortality rates from these diseases. Sri Lanka, positioned as a middle-income nation, reports intermediate levels of premature mortality due to NCDs. A significant concern is the high rate of premature mortality from NCDs, defined as deaths occurring before the age of 70. The World Health Organization's 2018 country profile indicates that NCDs account for 83% of all deaths in Sri Lanka (WHO 2020)

This surge in NCD prevalence is largely attributed to rapid lifestyle transitions, including changes in dietary habits, reduced physical activity, and increased tobacco use. These modifiable risk factors are strongly linked to the rising NCD burden in the country. Understanding the demographic disparities and factors influencing NCDs in Sri Lanka is crucial for effective health management and policy formulation. This study aims to analyze the trends, patterns, and determinants of NCDs within the Sri Lankan context, and examine the factors affecting NCDs for the management of Health. This type of analysis is essential in developing targeted interventions and allocating resources efficiently to mitigate the impact of NCDs on public health.

Table 1. *Distribution of Deaths by cause in Sri Lanka, highlighting the Significant Impact of NCDs, 2017*

Cause of Death	**Percentage of Total Deaths**
Non-communicable Diseases (NCDs)	83%
- Cardiovascular Diseases	34%
- Cancers	14%
- Diabetes	9%
- Chronic Respiratory Diseases	8%
- Other NCDs	18%
Communicable, Maternal, Perinatal, and Nutritional Conditions	10%
Injuries	7%

Source: World Health Organization, Noncommunicable Diseases (NCD) Country Profiles, 2018.

This data underscores the pressing need for comprehensive strategies to address the NCDs in Sri Lanka, focusing on prevention, early detection, and effective management to reduce the associated morbidity and mortality. Given this context, it is crucial to analyze the trends and patterns of NCD prevalence, along with the underlying determinants influencing their incidence in Sri Lanka.

Literature Review

Sri Lanka's demographic profile, characterized by a life expectancy of approximately 76 years, further underscores the importance of studying NCD trends. Notably, a considerable gender disparity in life expectancy exists, with women outliving men by nearly seven years (Department of Census & Statistics 2016).

Historical trends indicate that male and female life expectancy remained comparable until the mid-20th century, after which female longevity increased significantly. This disparity may be attributed to variations in healthcare access, occupational hazards, lifestyle choices, and environmental factors, including pollution and past conflicts (Ghaffa et al. 2004).

It is essential to first comprehend the definition of CVD. "Cardiovascular disease is caused by disorders of the heart and blood vessels, and includes coronary heart disease (heart attacks), cerebrovascular disease (stroke), raised blood pressure (hypertension), peripheral artery disease, rheumatic heart disease, congenital heart disease and heart failure." (World Health Organization 2015).

There are also studies conducted in Sri Lanka to identify the socio-economic factors influencing CVD. A study has been conducted using the micro-level secondary data from Household Income and Expenditure Survey 2016 of Sri Lanka (Abeysekara and Samaraweera 2022). The objective of the study has been to identify the socio-economic factors that affect the prevalence of non-communicable diseases among employed persons. The results revealed that factors of age, being female, and being Indian Tamil persons, positively affected the occurrence of non-communicable diseases; while factors like living in the rural and estate sector, age, being clerical workers, elementary workers, and agriculture workers affected it negatively. Profession-specific policies are further suggested to minimize the negative implications of CVD in Sri Lanka (Abeysekara and Samaraweera 2022).

A cross-sectional community-based study was conducted with 2277 rural adult males aged 20-60 years to detect the periodontal status of male smokers and betel chewers in a rural community in Sri Lanka (Amarasena et al. 2002). Periodontitis is considered as a risk indicator for cardiovascular diseases (Amarasena et al. 2002). The study has identified that oral hygiene and the quantified tobacco use may be considered as risk indicators for periodontitis (Amarasena et al. 2002). The study indirectly indicates how tobacco can be a behavioural risk factor towards the incidence of cardiovascular disease.

In order to understand the individuals' knowledge and awareness on the metabolic factors which affect cardiovascular diseases, a study has been conducted at the Nutrition Clinic of the National Hospital of Sri Lanka (Amarasekara et al. 2015). The study has identified that the participants with high knowledge mean score have significantly lower waist circumference (WC) and showed a trend toward reduced fasting blood glucose levels (Amarasekara et al. 2015). Participants with high practice scores had significantly lower BMI and waist circumference, which signify that better knowledge and practices are associated with decrease in CVD risk markers in these patients (Amarasekara et al. 2015).

Recognizing the complex interplay of factors are essential for formulating effective public health strategies. Literature indicates that demographic, social, and economic and behavioural factors influence CVD outcomes, with significant differences observed in disease prevalence and mortality rates. The analysis of cardiovascular disease history reveals varying prevalence rates across different age groups and genders. In terms of heart disease, the data indicates a consistent increase in prevalence with age for both men and women, with 16.3% of men and

14.8% of women aged 60-69 reporting a history of CVD. To build on this understanding, a rigorous methodological approach is necessary to systematically examine the factors influencing CVD and derive evidence-based insights for public health interventions.

Methodology

The study employs a quantitative research approach to analyze the factors influencing non-communicable diseases (NCDs) specially CVD in Sri Lanka. The methodology section outlines the research design, data sources, analytical techniques, and statistical models used to ensure the validity and reliability of the findings. The study relies on secondary data sources, primarily survey data and health-related records collected at regular intervals, to investigate the prevalence and risk factors associated with selected CVD. By adopting a systematic and ethical research framework, this study aims to generate insights that are both methodologically sound and applicable to a broader public health context.

Data Sources

The data utilized in this study were obtained from the Ministry of Health, Sri Lanka, and the Non-Communicable Diseases Risk Factor Survey conducted in 2021. This survey employed a multi-level cluster sampling method to ensure national representation. The dataset originally contained 133 variables, out of which 15 key variables were selected for analysis. The variables include behavioral, biological, demographic, and socioeconomic, such as alcohol consumption, smoking, diet, physical activity, blood glucose levels, blood pressure, cholesterol levels, BMI, age, sex, ethnicity, education level, and income levels. The selection of these variables was based on their established relevance to NCD risk assessment in global and regional health studies.

Selected Variables and Measurements

Demographic and socio-economic	Province, district, sector, sex, age, education, ethnicity, marital_status, emp_status, household_income
Behavioural	smoking_status, alcohol_status, intake /frequency of fruits, intake /frequency of veg, use of salt, sugar intake, physical activity
Biological/ Metabolic	RPB/hypertension, history of diabetes, raised_cholestrol, history of heart_diseases, systolic_BP, diastolic_BP, heart_rate, BMI, total_cholestrol, FBG

The independent variables were categorized into three main groups: demographic and socioeconomic factors, behavioral factors and biological/metabolic factors. By integrating these diverse factors, the study provides a comprehensive understanding of the multifaceted determinants of CVD prevalence in Sri Lanka.

Statistical Analysis

Study employs both descriptive and inferential statistical techniques to analyze the relationship between risk factors and CVD prevalence. Inferential analyses include chi-square tests, regression modelling, and predictive analytics. Chi-square tests were conducted to examine associations between categorical variables, while binary logistic regression was used to identify significant predictors of CVDs.

To further enhance predictive accuracy, five logistic regression models were developed to assess the factors affecting CVD prevalence. The models were evaluated using key performance metrics, including the Area Under the Curve (AUC), accuracy, specificity, sensitivity, precision, F1-score, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). The evaluation criteria enabled the selection of the most robust model for predicting CVD risk based on the identified variables. The findings from these models contribute to a data-driven approach to public health management by highlighting key risk factors and potential intervention strategies for reducing the burden of CVD in Sri Lanka.

Binary logistics regression model belongs to the family of statistical models identified as generalised linear models. In the binary logistics regression model, dependent variable is a dichotomous factor variable, which means it always has two levels only. The dependent variable is the prevalence of cardiovascular disease, whether suffering from the disease or not. Apart from the dependent variable it includes of at least one independent variable that is used to explain or predict values of the dependent variable. In summary, the focus of the binary logistic regression is to predict the probability of the levels of the categorical outcomes (Walsh 2016).

Results

As the first step, Chi-square analysis was conducted to examine the associations between various demographic, behavioral, biological, and environmental factors and the prevalence of CVD. The variables included province, sector, age group, sex, education, ethnicity, marital status, employment status, household income, smoking status, alcohol consumption, frequency of fruit and vegetable intake, salt and sugar consumption, physical activity, diabetes history, and environmental factors such as rainfall, temperature, and air quality. The results, presented in terms of p-values, test statistics, and degrees of freedom (df), indicate significant associations for several variables.

Table 2. *Chi-square Test Results for Factors associated with Cardiovascular Disease Risk*

Variable	P_Value	Statistic	Degrees of Freedom
Province	0.022891676	17.78571	8
Sector	0.505769469	1.363349	2
Age Group	< 0.01	41.7977	3
Sex	< 0.01	90.66605	1
Education	0.167404821	7.803466	5
Ethnicity	0.993052138	0.08963	3
Marital Status	0.686115262	0.753419	2
Employment Status	0.061949046	12.00089	6
Household Income	0.003732808	17.44245	5
Smoking Status	< 0.01	431.8241	2
Alcohol Status	< 0.01	476.7911	2
Frequency of Fruits	0.001158841	10.55484	1
Frequency of Vegetables	< 0.01	337.6583	1
Salt intake	< 0.01	65.31228	4
Sugar intake	< 0.01	724.791	2
Physical activity	< 0.01	852.4881	1
History of Diabetes	< 0.01	297.8708	1
Ever Diabetes	< 0.01	231.5809	1
Rainfall	0.213916798	8.344771	6
Temperature High	0.213916798	8.344771	6
Air Quality	0.022891676	17.78571	8

The results of the Chi-square analysis revealed significant associations between several demographic, behavioral, biological, and environmental factors with the prevalence of CVD. Among demographic variables, province ($\chi^2 = 17.79$, $p = 0.0229$) and age group ($\chi^2 = 41.80$, $p < 0.0001$) showed significant associations with CVD prevalence. Individuals aged 50 years and above exhibited a higher prevalence of CVD compared to younger age groups. Sex was also a highly significant factor ($\chi^2 = 90.67$, $p < 0.0001$), with males (33.56%) displaying a markedly higher prevalence than females (21.05%). However, variables such as sector ($\chi^2 = 1.36$, $p = 0.506$), education level ($\chi^2 = 7.80$, $p = 0.167$), ethnicity ($\chi^2 = 0.09$, $p = 0.993$), and marital status ($\chi^2 = 0.75$, $p = 0.686$) did not exhibit significant associations with CVD prevalence. Among socioeconomic factors, household income demonstrated a statistically significant association ($\chi^2 = 17.44$, $p = 0.0037$), with individuals from lower-income households reporting higher rates of CVD. Employment status approached significance ($\chi^2 = 12.00$, $p = 0.062$), suggesting a potential relationship between occupation and CVD risk.

Lifestyle factors such as smoking status ($\chi^2 = 431.82$, $p < 0.0001$) and alcohol consumption ($\chi^2 = 476.79$, $p < 0.0001$) were among the most strongly associated with

CVD. Current smokers (49.24%) and past smokers (48.40%) exhibited a significantly higher prevalence of CVD compared to those who never smoked (18.40%). Similarly, individuals who currently consume alcohol had the highest CVD prevalence (56.24%) compared to non-drinkers (18.73%). Dietary habits also played a crucial role in CVD prevalence. A lower frequency of fruit consumption (less than three days per week) was significantly associated with CVD ($\chi^2 = 10.55$, $p = 0.0012$), with 28.10% of such individuals having CVD compared to 23.65% of those who consumed fruits more frequently. Vegetable consumption showed an even stronger association ($\chi^2 = 337.66$, $p < 0.0001$), where 56.82% of individuals who consumed vegetables less than three days per week had CVD compared to only 21.79% among those with a higher intake. Salt intake ($\chi^2 = 65.31$, $p < 0.0001$) and sugar intake ($\chi^2 = 724.79$, $p < 0.0001$) were also highly significant. Notably, 91.63% of individuals with high sugar intake had CVD compared to only 20.35% among those with low intake. Physical activity emerged as one of the strongest protective factors against CVD, showing a highly significant association ($\chi^2 = 852.49$, $p < 0.0001$). Individuals who engaged in regular physical activity had a significantly lower prevalence of CVD (11.55%) compared to those with a sedentary lifestyle (50.59%). The history of diabetes ($\chi^2 = 297.87$, $p < 0.0001$) and current diabetes status ($\chi^2 = 231.58$, $p < 0.0001$) were also major predictors of CVD. Over 53.71% of individuals with a history of diabetes had CVD, compared to only 21.87% among those without diabetes.

Regarding environmental factors, air quality ($\chi^2 = 17.79$, $p = 0.0229$) demonstrated a significant relationship with CVD, indicating a potential impact of pollution on cardiovascular health. However, variables such as rainfall ($\chi^2 = 8.34$, $p = 0.214$) and temperature variations ($\chi^2 = 8.34$, $p = 0.214$) did not show significant associations.

This study found significant associations between CVD prevalence and several demographic, behavioral, and biological factors. Age, sex, household income, smoking, alcohol consumption, dietary habits (fruit, vegetable, salt, and sugar intake), physical activity, and diabetes history were strongly linked to CVD prevalence. Environmental factors such as air quality also showed a significant impact, whereas other climatic factors like temperature and rainfall were not significant. These findings emphasize the importance of lifestyle modifications and public health interventions in mitigating CVD risk.

Principal Component Analysis (PCA) is an unsupervised learning technique that have been used to reduce dimensionality of data and identify key components influencing disease risk. PCA has been used with the numeric variables of the study such as systolic BP, diastolic BP, heart rate, waist circumference, FBG (fasting blood glucose), BMI, and total_Chol (total cholesterol). The analysis has obtained biplots for PCA to identify the structures and patterns between the variables. PCA has been conducted in the study to obtain a robust understanding on the factors affecting CVDs in Sri Lanka.

The results show that the first principal component (PC1) accounts for 22.3% of the total variance, suggesting it captures the most dominant pattern within the dataset likely driven by key metabolic indicators such as blood pressure, glucose levels, and BMI. The second (PC2) component explain an additional 17.7%, respectively, indicating that other clusters of related risk factors, possibly linked to lipid profiles or behavioral characteristics, also contribute significantly to the overall data structure. Collectively, the first two components explain over 40% of the total variance, underscoring their capacity to represent the majority of the variability in the original dataset with far fewer dimensions.

To further understand and predict CVD prevalence, five logistic regression models were evaluated for their ability to predict CVD prevalence.

Model 1	Analysed using the full dataset with all the independent variables. Model output indicated that independent variables age, sector, education, wealth group and waist circumference were not statistically significant.
Model 2	Analysed using the full dataset but only with the significant independent variables from the Model 1.
Model 3	Analysed using the full dataset but only with the significant variables from the chi-square test of independence. Model output indicated that independent variable waist circumference was not statistically significant
Model 4	Analysed using the split dataset with all the independent variables. Model output indicated that independent variables age, sector, wealth group, hypertension, DBP and waist circumference were not statistically significant.
Model 5	Analysed using the split dataset but only with the significant independent variables from the Model 5.

The models were assessed using key performance metrics, including Area Under the Curve (AUC), accuracy, specificity, sensitivity, precision, F1-score, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). The table presents the performance metrics for five logistic regression models predicting the prevalence of CVD.

Table 3. *Performance Metrics of Logistic Regression Models for Cardiovascular Disease Prediction*

Metric	Model 1	Model 2	Model 3	Model 4	Model 5
AUC	0.913541042	0.913407111	0.958954063	0.957736589	0.957006
Misclassification Rate	0.136580087	0.136363636	0.106520768	0.09672619	0.09747
Accuracy	0.863419913	0.863636364	0.893479232	0.90327381	0.90253
Specificity	0.937260677	0.93814433	0.94541679	0.941346154	0.940385
Sensitivity	0.65877551	0.657142857	0.735320687	0.773026316	0.773026
Precision	0.791176471	0.793103448	0.815631263	0.793918919	0.791246
F1 Score	0.718930958	0.71875	0.773396675	0.783333333	0.78203
AIC	2980.071388	2975.456597	2055.772485	1480.195307	1480.897
BIC	3224.721087	3200.791846	2280.015098	1691.947597	1686.599

AUC (Area Under the Curve) measures the model's ability to distinguish between classes. Model 3, 4, and 5 (~0.957) performed the best. Misclassification rate is lowest in Model 4 (0.0967) and Model 5 (0.0975). Accuracy is highest in Model 4 (90.33%) and Model 5 (90.25%). Sensitivity (true positive rate) is crucial for detecting CVD cases. Model 4 & 5 (77.3%) performed the best among 5 models. Specificity (true negative rate) is slightly lower in Models 4 and 5

(~94.1%) but still high. Precision (positive predictive value) is slightly higher in Model 3 (81.5%) than Model 4 (79.3%), but Model 4 balances it better with the highest F1-score (78.3%), meaning it optimally balances precision and recall. Model 4 has the lowest AIC (1480.19) and BIC (1691.95), meaning it is the best-fitting model while penalizing complexity.

Model 4 was selected as the final model for further analysis, as it demonstrated the best trade-off between sensitivity (77.3%) and specificity (94.1%). It also exhibited the highest accuracy (90.33%) and the lowest misclassification rate (9.67%), along with the lowest AIC (1480.19) and BIC (1691.95), indicating an optimal balance between predictive power and model complexity.

Predictors of Cardiovascular Disease

A binary logistic regression analysis was conducted using Model 4 to identify significant predictors of CVD. The dependent variable was CVD prevalence, and the predictor variables included demographic, socioeconomic, lifestyle, dietary, and clinical factors.

Table 4. *Logistic Regression Estimates of Cardiovascular Disease Risk Factors in Sri Lanka*

	Estimate	Std..Error	z.value	Pr...z..
(Intercept)	-1.874485328	0.724513474	-2.58723322	0.009675008
Central	-0.471617202	0.292369783	-1.613084626	0.106726124
Southern	-0.139765155	0.242260508	-0.576920921	0.563992877
Northen	-0.473376249	0.330773038	-1.431121024	0.152395533
Eastern	-0.290112934	0.294378716	-0.985509204	0.324374013
Northwestern	-0.848135206	0.260376734	-3.257338671	0.001124622
North central	0.722139572	0.275188753	2.624160921	0.008686273
Uva	1.368931725	0.287463467	4.762106775	< 0.01
Sabaragamuwa	0.745308859	0.251622201	2.962015493	0.003056324
35-49	0.489891898	0.199036381	2.461318357	0.013842747
50-64	0.66998265	0.197459913	3.393005909	0.000691301
65+	0.638255614	0.279807162	2.281055312	0.022545174
Mae	0.837632409	0.141243854	5.930398987	< 0.01
10,001- 23,500	-0.36617207	0.285145221	-1.284159941	0.199085998
23,501- 36,500	-0.310233864	0.276869122	-1.120507273	0.262497655
36,501- 52,000	-0.309786284	0.275231121	-1.125549624	0.260356201
52,001- 81,500	-0.830938228	0.305958139	-2.715855935	0.006610468
more than 81501	-0.849100577	0.330639038	-2.56805906	0.010226973
Current smoker	-1.951588485	0.20634896	-9.457709316	< 0.01
Past smoker	-0.094266268	0.244796036	-0.385080861	0.700177508
Current	-2.443246634	0.198136039	-12.33115715	< 0.01
Past	-2.610379837	0.239162426	-10.91467367	< 0.01
More than 3 days	-0.254361117	0.155378695	-1.637039862	0.101622141

frequency_veg2	-2.037223734	0.198254442	-10.27580374	< 0.01
Salt lower	0.635108942	0.204492627	3.105779172	0.001897784
Salt higher	1.395494258	0.208853253	6.68169749	< 0.01
Salt highest	1.258815672	0.273822743	4.597191814	< 0.01
higher	0.700417142	1.233573899	0.56779504	0.570174148
Sugar higher	2.821357543	0.242820374	11.61911373	< 0.01
Sugar highest	5.3081139	0.496531328	10.69039071	< 0.01
Physical inactive	-3.48370584	0.177958729	-19.57592003	< 0.01
History of diabetes(positive)	1.526827546	0.18437341	8.281169951	< 0.01
Systolic blood pressure	0.022805312	0.00282255	8.079684034	< 0.01
Waist circumference	0.017241432	0.005629487	3.062700519	0.002193495
Total cholesterol	0.003387117	0.001577735	2.14682249	0.031807416

Demographic and Socioeconomic Factors

Age was found to be a significant predictor, with individuals aged 35–49 (OR = 1.63, p = 0.0138), 50–64 (OR = 1.95, p = 0.0007), and 65+ (OR = 1.89, p = 0.0225) exhibiting higher odds of CVD compared to the reference category. Males had significantly higher odds of developing CVD (OR = 2.31, p < 0.0001) than females.

Geographically, individuals in the North Western Province had significantly lower odds of CVD (OR = 0.43, p = 0.0011), whereas those in the North Central Province (OR = 2.06, p = 0.0087), Uva Province (OR = 3.92, p < 0.0001), and Sabaragamuwa Province (OR = 2.11, p = 0.003) exhibited significantly higher odds. Higher household income was associated with lower odds of CVD, particularly for individuals earning more than LKR 52,000 per month (OR = 0.43, p = 0.0066) and LKR 81,500 per month (OR = 0.43, p = 0.0102), suggesting economic advantages contribute to better cardiovascular health outcomes.

Behavioral and Lifestyle Factors

Smoking and alcohol consumption showed a positive association with CVD prevalence. Current smokers (OR = 0.14, p < 0.0001) and past smokers (OR = 0.91, p = 0.7002) had lower odds of CVD, as did individuals who consumed alcohol currently (OR = 0.09, p < 0.0001) or in the past (OR = 0.07, p < 0.0001). These findings may be influenced by confounding variables, such as age distribution and underlying health conditions.

Physical inactivity was a strong predictor of CVD, with sedentary individuals having significantly higher odds (OR = 32.52, p < 0.0001). This underscores the protective role of regular physical activity in cardiovascular health.

High dietary salt and sugar intake were strongly associated with increased CVD risk. Moderate salt intake (OR = 1.89, p = 0.0019), high salt intake (OR = 4.04, p < 0.0001), and very high salt intake (OR = 3.52, p < 0.0001) were all significant risk factors. Similarly, moderate sugar intake (OR = 16.80, p < 0.0001) and high sugar intake (OR = 201.27, p < 0.0001) substantially increased CVD odds. Conversely, higher vegetable consumption was protective ($OR = 0.13$, $p < 0.0001$).

Biological/Clinical Factors

A history of diabetes was a significant predictor of CVD (OR = 4.60, $p < 0.0001$), as were elevated systolic blood pressure (OR = 1.02, $p < 0.0001$), waist circumference (OR = 1.02, $p = 0.0022$), and total cholesterol levels (OR = 1.003, $p = 0.0318$), indicating that metabolic health parameters play a critical role in CVD risk.

The confusion matrix for Model 4 indicated high classification accuracy, with 979 true negatives (TN), 69 false negatives (FN), 61 false positives (FP), and 235 true positives (TP). The model achieved an accuracy of 90.3%, specificity of 94.1%, sensitivity of 77.3%, and a precision of 79.4%, ensuring a balanced identification of both CVD-positive and negative cases. The Receiver Operating Characteristic (ROC) curve analysis confirmed the model's high discriminative ability, with an AUC of 0.958, reinforcing its effectiveness in distinguishing between individuals with and without CVD.

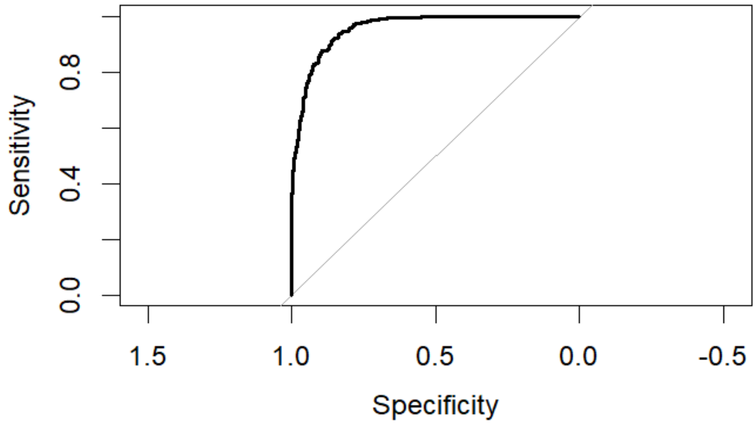
Model Performance Metrics using Test Data

The binary logistic regression model demonstrated strong predictive performance on the testing dataset. An overall accuracy of 82.45% indicates that the model correctly classified a large majority of the instances. The precision (83.46%) and sensitivity (83.25%) values suggest that the model was both reliable in correctly identifying individuals with CVD and effective at minimizing false negatives. The F1 Score (83.35%), which balances precision and recall, further supports the model's robustness in identifying true CVD cases. Most notably, the AUC of 0.8995 signifies excellent discriminatory power, indicating the model's strong ability to distinguish between individuals with and without CVD. These findings validate the model's suitability for predictive risk stratification in public health contexts and support its integration into early screening and intervention strategies.

Table 5. *Performance of RF on Test Dataset for CVD Prediction*

Metric	Value
Accuracy	82.45%
Misclassification Rate	17.55%
Precision	83.46%
Sensitivity (Recall)	83.25%
F1 Score	83.35%
AUC (Area Under Curve)	0.8995

Figure 1. *Receiver Operating Characteristic (ROC) Curve of the Outcome*



The results of this study highlight significant demographic, behavioral, and socioeconomic determinants of CVD prevalence in Sri Lanka, emphasizing the interplay between lifestyle choices and disease risk. These findings provide a foundation for a broader discussion on targeted public health interventions and policy recommendations aimed at reducing the national burden of cardiovascular diseases.

Limitations and Future Research

While our models demonstrated high predictive accuracy, several limitations merit discussion. The use of self-reported data introduces potential bias, particularly in smoking, alcohol use, and physical activity. Cross-sectional data also limits causal inference. Environmental variables, although included, showed limited significance possibly due to coarse measurement scales or missing mediators.

Future research should explore longitudinal designs to assess CVD progression and intervention impact. Incorporating biomarkers, genetic data, and real-time behavioral tracking (e.g., through mobile health apps) could enhance model precision. Moreover, extending prediction models to include other NCDs such as chronic respiratory diseases and cancers could offer a more holistic risk assessment platform.

Discussion

The findings of this study highlight significant associations between demographic, behavioral, biological and CVD prevalence. These results align with existing literature that underscores the multifactorial nature of CVD risk.

Age and sex were found to be strong predictors of CVD, with older individuals and males showing significantly higher prevalence rates. This is consistent with previous studies that establish aging as a primary risk factor due to vascular changes, oxidative stress, and cumulative exposure to risk factors over time (Benjamin et al. 2019). Similarly, the higher prevalence in males aligns with prior research indicating sex-related differences in CVD risk, likely due to hormonal variations and differences in risk behavior such as smoking and alcohol consumption (Virani et al. 2021).

Socioeconomic status, particularly household income, was significantly associated with CVD prevalence, with lower-income groups exhibiting higher rates. This finding aligns with previous studies demonstrating that economic disadvantage correlates with higher exposure to risk factors such as poor diet, limited healthcare access, and higher psychosocial stress (Marmot, 2018). Additionally, geographical variations in CVD prevalence, as seen in this study, may be attributed to differences in healthcare infrastructure, lifestyle patterns, and environmental conditions (Yang et al. 2022).

Lifestyle behaviors, including smoking and alcohol consumption, were strongly linked to CVD risk. Smoking has been extensively documented as a major contributor to CVD through mechanisms such as endothelial dysfunction, increased oxidative stress, and inflammation (Banks et al. 2019). Alcohol consumption also showed a significant association, supporting prior findings that excessive alcohol intake contributes to hypertension, cardiomyopathy, and arrhythmias (Rehm et al., 2017). However, some studies suggest moderate alcohol consumption may have protective effects, which warrants further investigation into drinking patterns and their impact on cardiovascular health (O'Keefe et al. 2018).

The model results highlight marked gender differences in CVD prevalence, consistent with existing global literature indicating higher cardiovascular risk among males compared to females (Lloyd-Jones et al. 2010). The analysis also suggests a complex relationship between alcohol use and CVD outcomes. Although current alcohol consumption is associated with elevated CVD risk in the Sri Lankan population, literature indicates that light-to-moderate alcohol consumption may be cardioprotective under certain circumstances (Ronksley et al. 2011). Given the cultural and behavioral context of alcohol use in Sri Lanka, such protective effects must be interpreted cautiously.

Dietary habits, particularly low fruit and vegetable intake, high salt consumption, and excessive sugar intake, were significant contributors to CVD risk. Numerous studies confirm that diets rich in fruits and vegetables reduce CVD risk by providing essential micronutrients, antioxidants, and fiber that promote cardiovascular health (Aune et al. 2017). High salt intake is a well-known risk factor for hypertension, which in turn elevates CVD risk (Mozaffarian et al. 2018). Similarly, excessive sugar intake, particularly from processed foods and sugary beverages, has been linked to obesity, metabolic syndrome, and insulin resistance, all of which contribute to CVD (Te Morenga et al. 2013).

Physical activity emerged as a crucial protective factor against CVD. The strong association between physical inactivity and CVD observed in this study is well supported by previous research demonstrating that regular exercise enhances cardiovascular function, reduces blood pressure, and improves lipid profiles (Lee et al., 2012). This underscores the importance of promoting physical activity as a key public health intervention to mitigate CVD risk.

Diabetes history was another major predictor of CVD, reinforcing the well-established link between diabetes and cardiovascular complications. Diabetes contributes to CVD through mechanisms such as hyperglycemia-induced vascular damage, increased inflammation, and dyslipidemia (Einarson et al. 2018). Effective management of diabetes is thus essential in reducing CVD burden.

Overall, these findings emphasize the importance of targeted public health strategies to address modifiable risk factors such as unhealthy diets, smoking, and physical inactivity. Further longitudinal research and interventional studies are necessary to establish causality and explore effective measures to mitigate CVD risk at both individual and population levels.

Conclusion

Demographic factors such as age and male sex significantly increased CVD risk, while higher household income appeared protective. Regional disparities in CVD prevalence were observed, with some provinces exhibiting significantly higher or lower risk levels. Lifestyle factors, particularly physical inactivity, high salt and sugar intake, and a history of diabetes, were major risk factors. Smoking and alcohol consumption showed a positive association with CVD prevalence. The model demonstrated strong predictive capabilities, making it a valuable tool for identifying high-risk individuals and informing public health interventions in Sri Lanka.

The findings emphasize the importance of early screening and intervention strategies, particularly in high-risk regions. Strengthening healthcare infrastructure and promoting community-based health initiatives could help mitigate the growing CVD burden. Further implementing educational programs focusing on lifestyle modifications, including physical activity, reduced salt and sugar intake, and diabetes prevention. Strengthening local health services and promote grassroots-level interventions to encourage healthier lifestyles and enforcing stricter regulations on processed foods high in salt and sugar and promote workplace wellness programs to reduce sedentary behavior would aid prevent the condition in future.

References

Abeysekara DS, Samaraweera GRSRC (2022) Hard Pill to Swallow: Socio-economic Factors behind Non-communicable Diseases of Employed Persons in Sri Lanka. *Sri Lanka Journal of Social Sciences and Humanities*, *2*(1), 55. https://doi.org/10.4038/slj ssh.v2i1.56.

Amarasena N, Ekanayaka ANI, Herath L, Miyazaki H (2002) Tobacco use and oral hygiene as risk indicators for periodontitis. *Community Dentistry and Oral Epidemiology*, *30*(2), 115–123. https://doi.org/10.1034/j.1600-0528.2002.300205.x

Amarasekara P, de Silva A, Swarnamali H, Senarath U, Katulanda P (2015) Knowledge, Attitudes, and Practices on Lifestyle and Cardiovascular Risk Factors Among Metabolic Syndrome Patients in an Urban Tertiary Care Institute in Sri Lanka. *Asia*

Pacific Journal of Public Health, *28*(1_suppl), 32S40S. https://doi.org/10.1177/1010539515612123

Aune D, Giovannucci E, Boffetta P, Fadnes LT, Keum NN, Norat T ... & Tonstad S (2017) Fruit and vegetable intake and the risk of cardiovascular disease, total cancer, and all-cause mortality—a systematic review and dose-response meta-analysis of prospective studies. *International Journal of Epidemiology, 46*(3), 1029-1056.

Banks E, Joshy G, Korda RJ, Stavreski B, Soga K, Egger S, Lopez AD (2019) Tobacco smoking and risk of 36 cardiovascular disease subtypes: fatal and non-fatal outcomes in a large prospective Australian study. *BMC Medicine, 17*(1), 1-12.

Benjamin EJ, Muntner P, Alonso A, Bittencourt MS, Callaway CW, Carson AP, Virani SS (2019) Heart disease and stroke statistics 2019 update: a report from the American Heart Association. *Circulation, 139*(10), e56-e528.

Brook RD, Rajagopalan S, Pope CA, Brook JR, Bhatnagar A, Diez-Roux AV, Peters A (2010) Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the American Heart Association. *Circulation, 121*(21), 2331-2378.

Ghaffar A, Reddy KS, Singhi M (2004) Burden of non-communicable diseases in South Asia. *Clinical Review,* 328, 807-810.

Einarson TR, Acs A, Ludwig C, Panton UH (2018) Prevalence of cardiovascular disease in type 2 diabetes: a systematic literature review of scientific evidence from across the world in 2007–2017. *Cardiovascular Diabetology, 17*(1), 1-19.

Lee IM, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT (2012) Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *The Lancet, 380*(9838), 219-229.

Marmot M (2018) Social determinants of health inequalities. *The Lancet, 365*(9464), 1099-1104.

Mozaffarian D, Fahimi S, Singh GM, Micha R, Khatibzadeh S, Engell R, Powles J (2018) Global sodium consumption and death from cardiovascular causes. *New England Journal of Medicine, 371*(7), 624-634.

Ministry of National Policies and Economic Affairs & Ministry of Health, Nutrition and Indigenous Medicine (2017) *Sri Lanka Demographic and Health Survey.* Department of Census and Statistics. http://www.statistics.gov.lk/Health/StaticalInformation

O'Keefe JH, Bhatti SK, Bajwa A, DiNicolantonio JJ, Lavie CJ (2018) Alcohol and cardiovascular health: the dose makes the poison… or the remedy. *Mayo Clinic Proceedings, 93*(8), 1104-1112.

Rehm J, Baliunas D, Borges GLG, Graham K, Irving H, Kehoe T, ... Taylor B (2017) The relation between different dimensions of alcohol consumption and burden of disease: an overview. *Addiction, 105*(5), 817-843.

Ronksley PE, Brien SE, Turner BJ, Mukamal KJ, Ghali WA (2011) Association of alcohol consumption with selected cardiovascular disease outcomes: a systematic review and meta-analysis. *BMJ,* 342, d671. https://doi.org/10.1136/bmj.d671

Virani SS, Alonso A, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, ... American Heart Association Council on Epidemiology and Prevention (2021) Heart disease and stroke statistics—2021 update: a report from the American Heart Association. *Circulation, 143*(8), e254-e743.

World Health Organization (2018) *Sri Lanka - Risk of pre mature death due to NCD's* (Rep.). World Health Organization - Noncommunicable Diseases (NCD) Country Profiles.

World Health Organization (2021) *Non-communicable diseases*. https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases

World Health Organization (2020) Non-communicable diseases. https://www.who.int/data/ gho/data/indicators/indicator-details/GHO/gho-ghe-ncd-deaths-in-thousands

World Health Organization (2020) STEP Approach to Chronic Disease Risk Factor Surveillance; No Communicable Disease Risk Factor Survey. Geneva: World Health Organization; 2020. Available from: https://www.who.int/ncds/surveillance/steps/steps-report-2015-sri-lanka.pdf. [Last accessed on 2020 Jun 25].