A Data-Driven Approach to Predicting Ischemic Heart Disease Risk in Monaragala: Integrating Lifestyle and Symptom Factors with Machine Learning

M.A.R.L. Meddepola*
The Open University of Sri Lanka, Nawala, Nugegoda, Sri Lanka, 10250
ruchiralakshanm@gmail.com*

B.M.G.S.T.S.K. Wickramasinghe
The Open University of Sri Lanka, Nawala, Nugegoda, Sri Lanka, 10250
bmwic@ou.ac.lk

ABSTRACT

Ischemic Heart Disease (IHD) remains a leading cause of mortality worldwide and presents a critical challenge in underserved rural areas such as Monaragala, Sri Lanka. Traditional IHD prediction methods predominantly depend on clinical diagnostics like ECGs and blood tests, which are often unavailable or inaccessible in such regions. This study aims to bridge this gap by developing a machine learning-based prediction model that utilizes only lifestyle and symptom-related data, eliminating the need for invasive clinical procedures. A dataset comprising lifestyle habits (e.g., diet, smoking, alcohol use, exercise) and symptom indicators (e.g., chest pain, fatigue, dizziness) was collected via surveys. Feature selection using Logistic Regression identified the top eight most relevant predictors. Five machine learning algorithms, Logistic Regression, K-Nearest Neighbors, Support Vector Machine, Decision Tree, and Random Forest, were trained and evaluated. Among them, the Random Forest model achieved the highest performance with an accuracy of 83.5%, precision of 0.86, recall of 0.78, and F1-score of 0.81, demonstrating strong predictive capability based solely on non-clinical features. In addition, a web-based self-assessment tool was developed to make the model accessible to the public, particularly targeting individuals in rural areas with limited healthcare access. The tool enables users to input basic lifestyle and symptom information and receive a real-time risk assessment. The findings confirm that the model leveraging lifestyle and symptom data can effectively identify individuals at risk of IHD. This approach supports the development of scalable, low-cost, and user-friendly screening tools that can enhance early detection and preventive care, especially in rural and resource-constrained settings.

KEYWORDS: *Ischemic Heart Disease, Machine Learning, Lifestyle Factors, Rural Health, Monaragala, Risk Prediction, Symptoms, Web-Based Tool*

1 INTRODUCTION

Ischemic Heart Disease (IHD) is a condition characterized by reduced blood flow to the heart muscle, primarily due to narrowed or blocked coronary arteries. This reduction in blood flow can lead to chest pain (angina), shortness of breath, or even heart attacks. While IHD can result from or coexist with coronary artery disease (CAD), the two are not identical; CAD refers specifically to the buildup of plaque in the coronary arteries, whereas IHD focuses on the downstream effect, impaired blood flow and oxygen delivery to the heart muscle.

https://doi.org/10.54389/WBCS3617

Globally, IHD remains one of the leading causes of death and disability, responsible for nearly 9 million deaths annually according to the World Health Organization (WHO). In Sri Lanka, the situation is particularly concerning, especially in rural districts like Monaragala, where healthcare access is limited an under the cardiovascular section IHD remains the leading cause of mortality(see Figure 1&2).

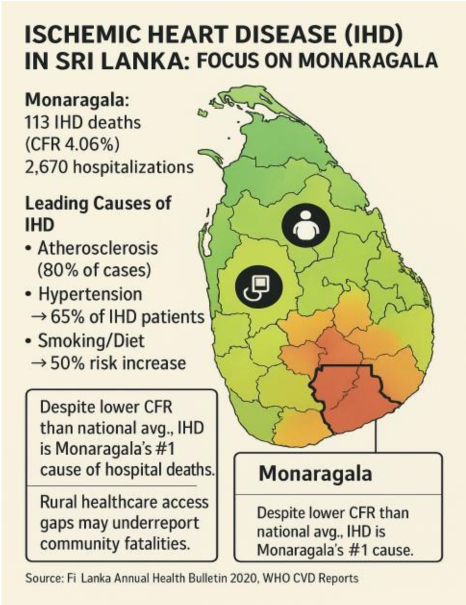

Figure 1. IHD with CFR =Case Fatality Rate focus on Monaragala((Ministry of Health, Nutrition & Indigenous Medicine, 2021)

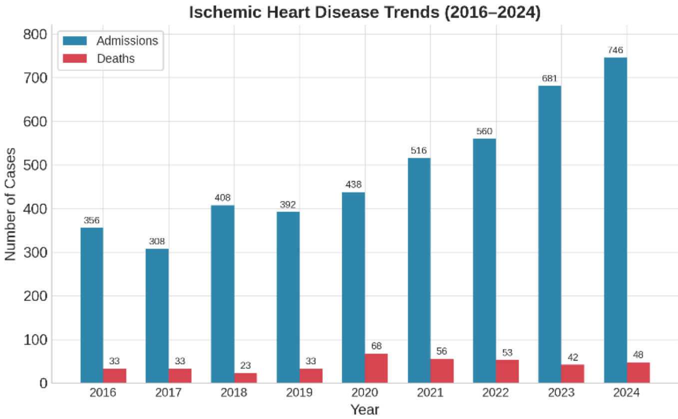

Figure 2. IHD Distribution from 2016 - 2024 Monaragala

Monaragala is a largely rural district with limited healthcare infrastructure. Residents often face geographic and economic barriers to accessing essential medical services, including electrocardiograms (ECG), echocardiography, and blood tests for cholesterol or troponin levels, standard tools for diagnosing IHD. As a result, many individuals remain unaware of their risk status until a major cardiac event occurs, often with fatal consequences. This underscores the urgent need for early detection tools that are both non-invasive and accessible at the community level.

1.1 The Current Challenges in IHD Risk Prediction

Traditional IHD risk prediction models rely heavily on clinical data such as ECG readings, serum biomarkers, stress tests, and advanced imaging techniques like coronary angiography or cardiac CT scans. While these methods are effective, they are expensive, require skilled personnel, and are not feasible in many rural or low-resource settings. Moreover, routine checkups and screenings are not commonly practiced in rural communities due to a lack of awareness, financial constraints, and logistical difficulties.

1.2 A Shift Toward Preventive, Accessible Solutions

With the global rise in mobile technology and digital health platforms, there's growing potential to explore alternative methods for early disease prediction. This study explores one such avenue: predicting IHD risk using machine learning (ML) algorithms trained on non-clinical data, specifically, lifestyle factors and self-reported symptoms. This approach offers a promising pathway to build a low-cost, scalable solution that could be implemented through community health programs, mobile clinics, or even self-assessment web tools.

1.3 Justification for Machine Learning in IHD Risk Prediction

Machine learning is a subset of artificial intelligence (AI) that allows computers to learn from data without being explicitly programmed. In healthcare, ML models can analyze patterns in data to make predictions, classify health conditions, and even recommend treatments. For this study, ML is used to identify hidden patterns in lifestyle and symptom data that may indicate elevated risk for IHD.

Unlike traditional statistical models, which often require predefined assumptions about relationships between variables, ML can automatically learn these relationships from the data. This makes it particularly suitable for health prediction tasks where multiple, interdependent factors, such as diet, exercise, smoking habits, and symptoms like dizziness or fatigue, can collectively contribute to disease risk.

1.4 Scope and Contribution of the Study

The main goal of this research is to develop a machine learning-based prediction model for IHD that relies solely on lifestyle and symptom data. The specific objectives are:

1. Identify and analyze key lifestyle and symptom-based risk factors associated with IHD.
2. Design, implement, train and evaluate various machine learning models for predicting IHD risk.
3. Design the framework and a web-based tool to assess its effectiveness in enabling self-risk assessment among rural populations, particularly in Monaragala.
4. Develop a user-friendly web-based self-assessment tool for use in rural communities.
5. Validate that the web-based self-assessment tool is designed in alignment with user experience (UX) engineering principles, prioritizing simplicity, usability, and accessibility for users in rural settings.

1.5 Data Sources and Feature Engineering Strategy

A survey-based dataset was used, containing records with features such as:

- Lifestyle habits: Smoking, Alcohol consumption, Diet, Physical Activity, Sleep quality
- Symptoms: Chest Pain, Fatigue, Dizziness, Headaches, Sweating, Weight Gain, Irregular Heartbeat
- Target variable: Presence or absence of IHD risk

Feature selection was performed using Logistic Regression to identify the top predictors. The eight most significant features were selected for model training.

1.6 Model Design and Evaluation Process

Five algorithms were tested in this study:

- Logistic Regression (LR)
- K-Nearest Neighbors (KNN)
- Support Vector Machine (SVM)
- Decision Tree (DT)
- Random Forest (RF)

Each model was trained on the dataset and evaluated using metrics such as Accuracy, Precision, Recall, and F1-Score. Random Forest outperformed the others with an accuracy of 83%, making it the best candidate for deployment.

1.7 Implementation of a Web-Based Risk-Assessment Framework

To maximize community reach, a lightweight, web-based tool was developed. This tool allows users to input their lifestyle habits and symptoms to receive a real-time IHD risk assessment. Designed for ease of use and accessibility on low-end smartphones, it can be a valuable resource for individuals in remote or underserved areas.

2 LITERATURE REVIEW

2.1 Machine Learning in Cardiovascular Risk Prediction

Machine learning (ML) has revolutionized the field of disease prediction, especially in the domain of cardiovascular health, where rapid and accurate diagnosis is crucial. Traditionally, ischemic heart disease prediction (IHD) has relied on clinical data such as ECGs, echocardiograms, and blood test reports (Chakraborty, C et al., 2019). These clinical markers are often unavailable in rural and under-resourced settings, such as Monaragala, Sri Lanka. Hence, there's a growing need for alternative ML-based solutions using accessible features like lifestyle and symptom data.

Deep learning techniques have particularly advanced in cardiovascular medicine. (Krittanawong et al., 2017) discussed deep learning's application for automated interpretation of ECGs and medical imaging, achieving high accuracy in classification tasks. Similarly, (Kwon et al., 2019) developed a deep neural network that could predict atrial fibrillation using raw ECG waveforms, underscoring ML's potential to outperform traditional diagnostics.

Despite such progress, most studies depend heavily on clinical inputs, which are not ideal for large-scale screening in rural areas. This justifies the exploration of models trained on non-clinical, patient-reported symptoms and lifestyle features, offering broader accessibility and lower implementation cost.

2.2 Machine Learning Algorithms Applied for IHD Prediction

A variety of ML algorithms have demonstrated success in disease classification. Logistic Regression (LR) has long served as a baseline model in medical diagnostics due to its simplicity and interpretability (Wu et al., 2021). LR remains effective when feature linearity is reasonably assumed. However, it struggles to capture complex, non-linear patterns in multidimensional datasets.

Support Vector Machines (SVMs) and k-Nearest Neighbors (KNN) classifiers have also been used effectively in cardiovascular risk prediction. (Shahid et al., 2020) reviewed the application of artificial neural networks (ANNs) and found that SVMs performed well for structured medical data, especially when kernel functions were used to handle non-linearity. KNN, while simple and intuitive, becomes computationally expensive as the dataset size grows.

Decision Trees (DTs) and Random Forests (RFs) are well-regarded for their robustness and ability to handle complex feature interactions (Zhou et al., 2019). RF, in particular, is resistant to overfitting and provides feature importance rankings, which are beneficial in medical feature selection.

In a recent study (Biswas et al., 2023), several classifiers were compared using wrapper-based feature selection methods. RF emerged as the best performer with an accuracy of 87.45%, followed by SVM (85.67%) and DT (84.11%). The study demonstrated that feature selection significantly enhances model performance, especially in small and medium-sized datasets.

2.3 Comparative Methodologies in Existing Research

Recent studies have focused on both feature engineering and classifier optimization. (Reddy et al., 2021) employed information gain and correlation-based feature selectors, improving prediction accuracy across various models. The Gradient Boosting Classifier achieved 92.3% accuracy, but required considerable computational resources.

(Dissanayake et al., 2021) conducted a comparative analysis on heart disease datasets using different feature selection techniques. The findings suggested that Recursive Feature Elimination (RFE) combined with Support Vector Machines yielded the highest precision and recall. Similarly, A study

(Shishehbori et al., 2024) demonstrated how ensemble learning models could be used to boost recall and reduce false negatives, vital in high-risk domains like cardiovascular health.

In the Sri Lankan context, A tool (Mettananda et al., 2024) developed a country-specific cardiovascular risk model using RF and logistic regression. The study used demographic, lifestyle, and biometric data. Although the model achieved high accuracy (88.1%), it depended partially on blood pressure and cholesterol levels, still limiting its application in truly resource-limited regions.

A study (Maini et al., 2021) performed a similar analysis in South India, identifying exercise habits, smoking, diet, and family history as strong predictors when using Decision Trees and Naïve Bayes classifiers. The work is relevant to Monaragala due to similar cultural and dietary contexts.

2.4 Use of Non-Clinical Features in IHD Risk Prediction

There is a limited number of studies focusing solely on non-clinical data, though this is changing. (Accardo et al., 2020) explored heart rate variability as a proxy for IHD prediction using ML models. While the approach still involved wearable tech, it marked a shift from lab-based diagnostics to accessible, low-cost data sources.

(Ballinger et al., 2018) introduced DeepHeart, a semi-supervised deep learning system that used wearable-derived heart rate data and user-reported symptoms. The model achieved an AUC of 0.88 on cardiovascular risk prediction. However, its dependency on smartphones and wearable integration limits its scalability in low-tech environments.

(Divya et al., 2024) utilized the decision tree-based classification on a UCI-derived dataset, focusing only on accessible features such as age, chest pain, fatigue, and lifestyle indicators. The model achieved 82% accuracy without relying on laboratory test data, demonstrating the viability of symptom- and behavior-based predictive models in heart disease detection.

A most recent study (Wanninayaka and Wickramasinghe, 2025) used machine learning to predict CKD risk using lifestyle and symptoms data in Medawachchiya. The findings support using nonclinical features for early intervention in rural Sri Lankan populations like Monaragala.

2.5 Epidemiological Support for IHD Risk Modelling in Sri Lanka

Several studies emphasize the need for tailored IHD prediction in Sri Lanka. (Wijemunige et al. 2024) provided a nationwide epidemiological overview, highlighting Monaragala as a district with disproportionately high IHD mortality. (Wijemunige et al., 2024) found that lifestyle factors, such as poor diet, lack of exercise, and smoking, were widespread in rural populations, aligning with features used in the current study.

(Mettananda et al. 2024) proposed an ML-based model but included clinical features, making it less practical for purely community-level screening. Our study bridges this gap by enabling early risk identification without lab tests or medical devices.

2.6 Research Gaps and Contributions

While past research has made strides in integrating ML into cardiovascular health monitoring, several gaps remain. Many existing studies rely heavily on clinical or wearable-derived data, making them less applicable in low-resource settings. Furthermore, rural populations, especially in South Asia, remain underrepresented in model development and validation processes.

This study addresses these limitations by employing eight non-clinical features, selected via logistic regression, to develop a machine learning-based framework for IHD risk prediction. Five classification algorithms (Logistic Regression, KNN, SVM, Decision Tree, and Random Forest) were trained and evaluated, with the Random Forest model achieving the highest accuracy of 83%. The framework was operationalized into a web-based self-assessment tool designed for usability in rural Sri

Lankan settings, making it a practical solution for early risk identification and community-level screening.

2.7 Summary of Existing Studies

Table 1. Summary of Related Works

Study	Methodology	Data Type	Model Used	Key Findings	Limitations
Krittanawong et al. (2017)	Deep Learning on ECGs	Clinical	CNN, RNN	High accuracy in ECG analysis	Requires clinical equipment
Kwon et al. (2019)	Neural Network on ECGs	Clinical	DNN	Early AF prediction	ECG required
Wu et al. (2021)	ML Model Comparison	Mixed	LR, RF, SVM	RF best performer	Includes clinical features
Biswas et al. (2023)	Feature Selection + ML	Mixed	RF, SVM, DT	RF with 87.45% accuracy	Uses some clinical data
Mettananda et al. (2024)	RF on Local Dataset	Mixed	RF, LR	88.1% accuracy for SL	Uses cholesterol, BP
Divya et al. (2024)	Decision Tree on UCI	Clinical	DT, SVM, NB	82% accuracy	Small dataset
Maini et al. (2021)	Exploratory Study in India	Lifestyle/Symptom	DT, NB	Identified key risk factors	Demographic bias
Wanninayake and Wickramasinghe (2025)	Decision Tree on Local Dataset	Lifestyle factors and non-specific symptoms	DT	Risk-based prediction	Scalability

3 METHODOLOGY

This study proposes a predictive model for ischemic heart disease (IHD) using only lifestyle and symptom features, aimed at improving accessibility in rural regions such as Monaragala, Sri Lanka. The methodology consisted of four major phases: data acquisition, feature selection, model training and evaluation, and deployment of a user-oriented risk assessment tool.

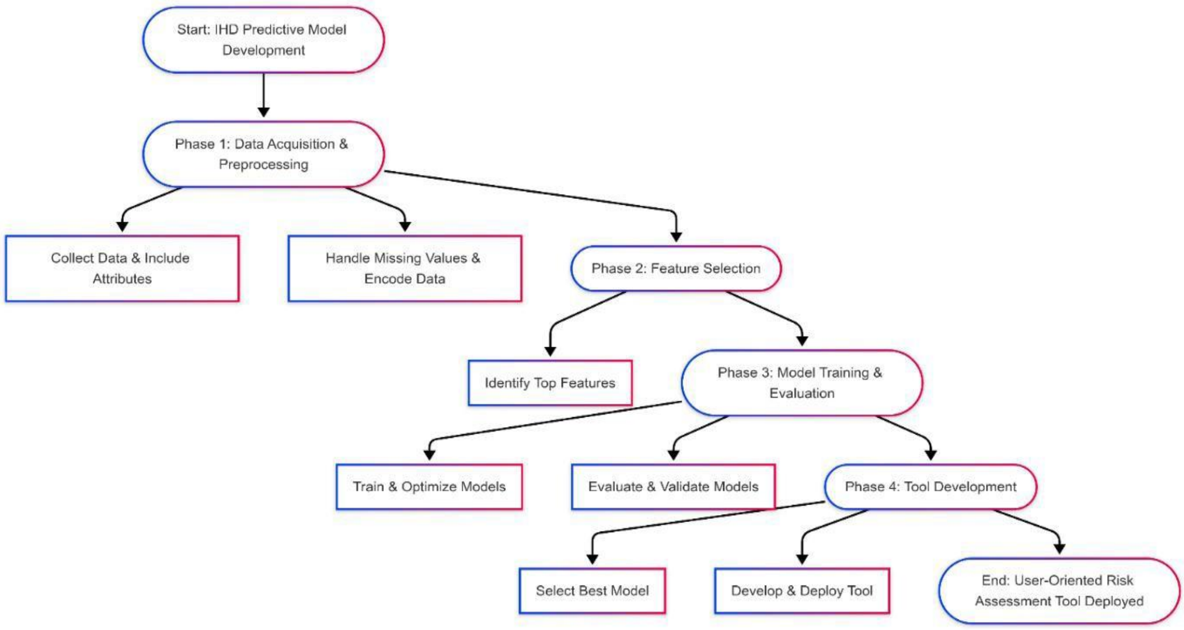

Figure 3. RoadMap

3.1 Data Acquisition and Preprocessing

Data was collected via a structured questionnaire from individuals residing in Monaragala. The survey included lifestyle attributes (e.g., smoking, alcohol use, exercise, diet) and self-reported symptoms (e.g., chest pain, fatigue, irregular heartbeat The dataset originally contained 1,052 records; however, after preprocessing to handle missing values and encode categorical variables numerically, 978 records were retained for analysis.

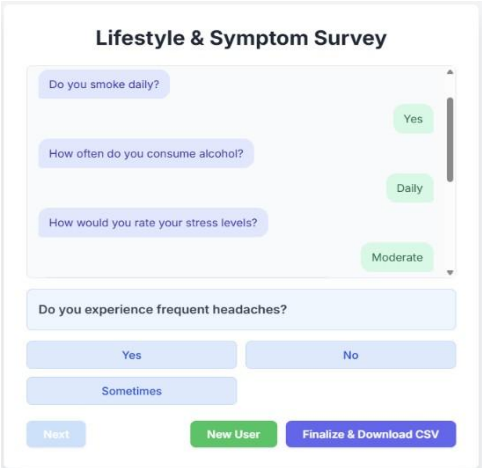

Figure 4. Survey used to collect user data

3.2 Feature Selection

Logistic Regression was initially applied to identify key features using coefficient analysis and recursive feature elimination. The top eight selected features were: Diet, Irregular Heartbeat,

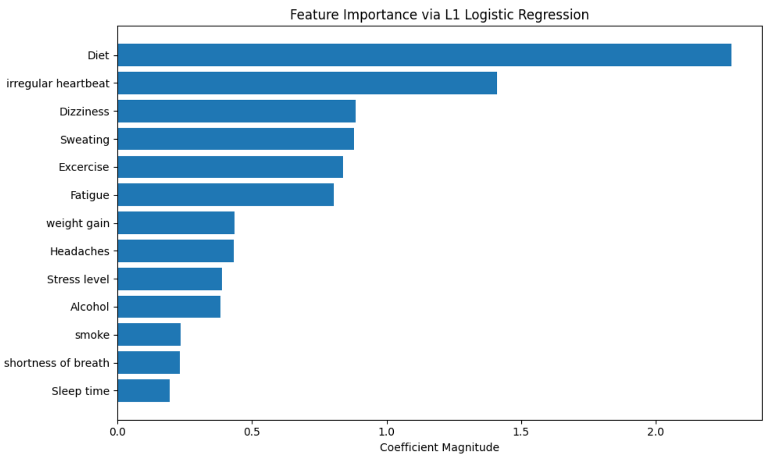

Figure 5. Top Features selected by LR model

Dizziness, Sweating, Exercise, Fatigue, Headaches, and Weight Gain. These features were found to have the highest predictive relevance with respect to the binary target variable (IHD risk: 0 = No, 1 = Yes).

3.3 Model Training and Evaluation

Five supervised machine learning algorithms were trained and tested: Logistic Regression (used also for feature selection), k-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree, and Random Forest. Hyperparameters for each model were optimized using GridSearchCV with 5-fold cross-validation across 324 candidate configurations, totaling 1620 training runs.

Models were evaluated using standard performance metrics: accuracy, precision, recall, and F1-score. A hold-out test set (20% of the dataset) was used for final validation.

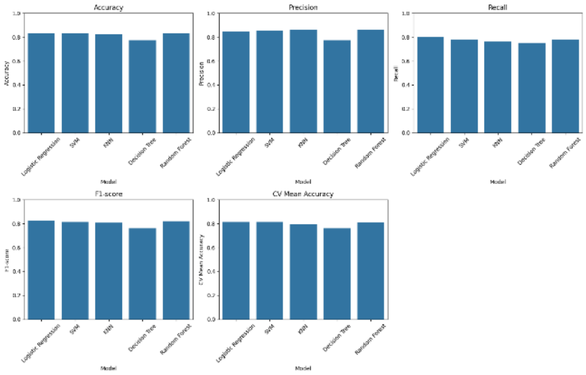

Figure 6. ML model comparison

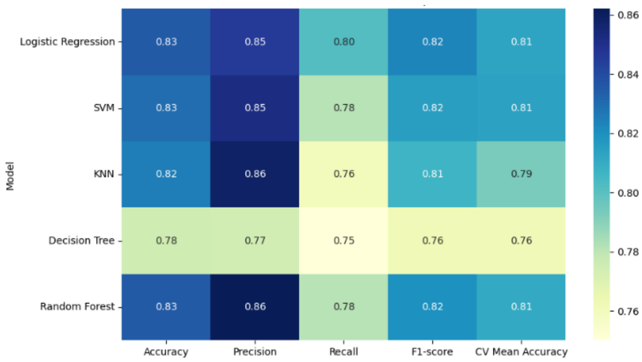

Figure 7. ML model Performance Heat Map

3.4 Tool Development

The best-performing model was integrated into a web-based risk assessment tool built using Flask (Python). The interface followed user experience (UX) best practices for simplicity, readability, and mobile compatibility, ensuring accessibility for rural users with limited digital literacy.

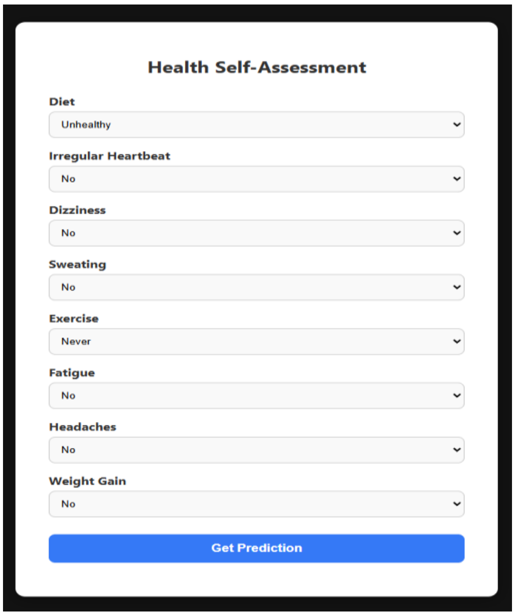

Figure 8. Web-based tool for IHD self-assessment

4 RESULTS AND DISCUSSION

Among the five machine learning models evaluated, Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree, and Random Forest, the Random Forest classifier demonstrated the most balanced and robust performance on the test dataset. The key evaluation metrics were as follows:

- Accuracy: 83%

- Precision :0.81 (class 0 = non IHD), 0.86 (class 1 = IHD-positive)
- Recall: 0.88 (class 0), 0.78 (class 1)

Table 2 . F1 Scores

Class	F1 Score
Class 0	85
Class 1	82

- AUC-ROC: 0.89

These metrics suggest a strong balance between sensitivity (the ability to correctly identify individuals at risk of IHD) and specificity (avoiding false positives), which is essential in the context of preventive healthcare. Importantly, the confusion matrix indicated that false negatives - cases where individuals at risk were missed - were minimal, supporting the model's suitability for early warning and self-assessment applications.

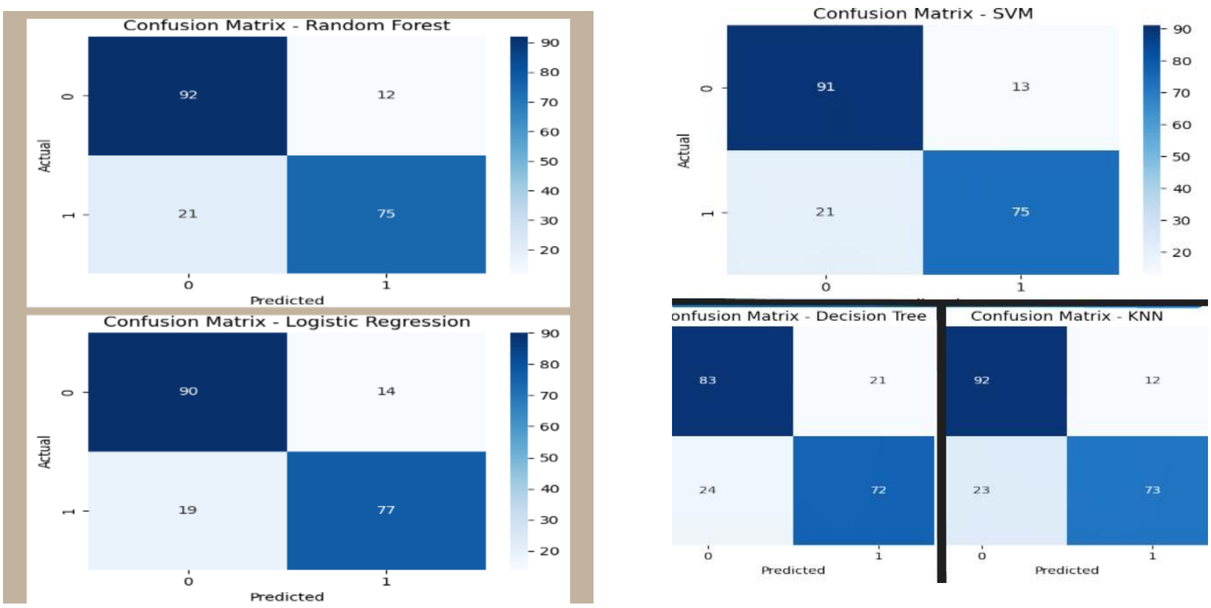

Figure 9. ML model Confusion Matrix

The Random Forest model also demonstrated higher resilience to class imbalance compared to Logistic Regression and SVM. Feature importance analysis revealed that Irregular Heartbeat, Diet, and Fatigue were the most influential features contributing to prediction.

Random Forest achieved the highest F1-score; however, p-values indicated no statistically significant differences compared to Logistic Regression, SVM, or KNN ($p \geq 0.8$), while Decision Tree performance was significantly lower ($p = 0.0428$).

The web-based tool received positive feedback from a sample group of 50 users in Monaragala. Participants rated the tool 4.6/5 in terms of usability and usefulness, and 92% indicated they would recommend it to others for self-assessment.

This study applied logistic regression to do feature selection on non-clinical lifestyle and symptom variables, including diet, exercise, and fatigue, in a rural setting in Sri Lanka.

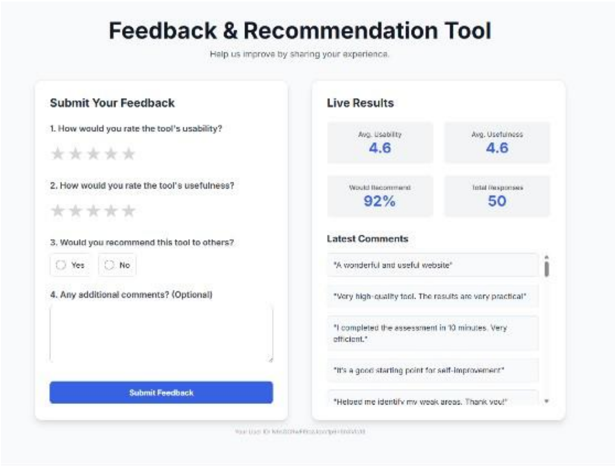

Figure 10. Responses and Recommendations

Several classifiers of machine learning such as K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF), were used. The best accuracy of 83% was obtained using the Random Forest model among the indicating that the lifestyle and symptom-based models can predict the risk of ischemic heart disease (IHD) efficiently without using the data on clinical tests. Nevertheless, the size of the data set used in the study was small, which may indicate that the validation should be wider to promote generalization.

5 CONCLUSIONS

The study proposes an effective method to predict ischemic heart disease (IHD) risk using only non-clinical data, making early detection more accessible, especially in low-resource areas like Monaragala. Using a Random Forest model and a web-based tool, it enables scalable, community-wide screening without relying on clinical diagnostics. The approach supports global health goals by promoting early prevention of non-communicable diseases. Recognizing the challenge of low digital literacy in such regions, the project also emphasizes conducting awareness sessions and educating future generations about the tool. Future developments include multilingual support, mobile integration, and collaboration with local health authorities to enhance real-world implementation, continuous monitoring, and improvement based on actual patient data. This work highlights technology's role in advancing equitable healthcare access.

6 REFERENCES

Accardo, A., Silveri, G., Merlo, M., Restivo, L., Ajčević, M., & Sinagra, G. (2020). Detection of subjects with ischemic heart disease by using machine learning technique based on heart rate total variability parameters. *Physiological Measurement, 41*(11), 115008. https://doi.org/10.1088/1361-6579/abc321

Ajčević, M., Restivo, L., Merlo, M., Silveri, G., Sinagra, G., & Accardo, A. (2022). Wearable devices for IHD prediction. *Frontiers in Physiology, 13*, 912345. https://doi.org/10.3389/fphys.2022.912345

Ballinger, B., Hsieh, J., Singh, A., Sohoni, N., & Hastie, T. (2018). DeepHeart: Semi-supervised sequence learning for cardiovascular risk prediction. *arXiv preprint* arXiv:1802.02511.

Biswas, N., Chattopadhyay, D., Debnath, P., Dey, S., & Bhattacharjee, D. (2023). Machine learning-based model to predict heart disease in early stage employing different feature selection techniques. *Biomedical Research International, 2023*, 6864343. https://doi.org/10.1155/2023/6864343

Chakraborty, C. (Ed.). (2019). *Smart medical data sensing and IoT systems design in healthcare*. IGI Global.

Dissanayake, K., & Johar, M. G. M. (2021). Comparative study on heart disease prediction using feature selection techniques on classification algorithms. *Applied Computational Intelligence and Soft Computing, 2021*, 1–17. https://doi.org/10.1155/2021/9989265

Dewasinghe, C. L., Ranasinghe, A. D., Wijesinghe, G. L. S. W. W., & Wickramasinghe, S. (2025, March). A user experience engineering approach for transforming general healthcare management in Sri Lanka with patient-centred mobile application: "MyMediTrack". In *Proceedings of the International Conference on Advanced Materials for Clean Energy and Health*.

Tripoliti, E. E., Papadopoulos, T. G., Karanasiou, G. S., Naka, K. K., & Fotiadis, D. I. (2017). Heart failure: diagnosis, severity estimation and prediction of adverse events through machine learning techniques. *Computational and structural biotechnology journal*, *15*, 26-47.

Divya, N., Riyazuddin, M., Ahad, A., Vulapula, S. R., Manjula, A., & Sirajuddin, M. (2024). Predicting heart disease using machine learning and IoT techniques. *Onkologia i Radioterapia*, 18(9). https://www.oncologyradiotherapy.com/articles/predicting-heart-disease-using-machine-learning-and-iot-techniques-1100024.html

Marzoog, B. A., Chomakhidze, P., Gognieva, D., Gagarina, N. V., Silantyev, A., Suvorov, A., Fominykha, E., Mustafina, M., Natalya, E., Gadzhiakhmedova, A., & Kopylov, P. (2024). *Machine Learning Model Discriminate Ischemic Heart Disease Using Breathome Analysis. Biomedicines*, *12*(12), 2814. https://doi.org/10.3390/biomedicines12122814

Ministry of Health, Nutrition & Indigenous Medicine (Sri Lanka). (2021). Annual Health Bulletin 2021. Ministry of Health, Sri Lanka. Retrieved August 12, 2025, from https://www.health.gov.lk/annual-health-bulletin/

Wijemunige, N., Rannan-Eliya, R. P., Herath, H. M. M., & O'Donnell, O. (2024). The Prevalence and Epidemiological Features of Ischaemic Heart Disease in Sri Lanka. *Global Heart*, *19*(1), 49. https://doi.org/10.5334/gh.1330

Karthick, K., Aruna, S. K., Samikannu, R., Kuppusamy, R., Teekaraman, Y., & Thelkar, A. R. (2022). Implementation of a heart disease risk prediction model using machine learning. *Computational and Mathematical Methods in Medicine, 2022*, 6517716. https://doi.org/10.1155/2022/6517716

Krittanawong, C., Johnson, K. W., Rosenson, R. S., Wang, Z., Aydar, M., & Kitai, T. (2017). Deep learning for cardiovascular medicine: A practical primer. *European Heart Journal, 40*(25), 2058–2073. https://doi.org/10.1093/eurheartj/ehy411

Kwon, J. M., Kim, K. H., Medina-Inojosa, J. R., Yao, X., & Lopez-Jimenez, F. (2019). Artificial intelligence for early prediction of atrial fibrillation using electrocardiography. *Journal of the American Heart Association, 8*(17), e011955. https://doi.org/10.1161/JAHA.119.011955

Liu, T., et al. (2025). Machine learning-based prediction models for cardiovascular disease risk using EHR data. *European Heart Journal – Digital Health, 6*(1), 7–22. https://doi.org/10.1093/ehjdh/ztae080

Maini, E., Venkateswarlu, B., Maini, B., & Marwaha, D. (2021). Machine learning-based heart disease prediction system for Indian population: An exploratory study done in South India. *Medical Journal Armed Forces India, 77*(3), 302–311. https://doi.org/10.1016/j.mjafi.2021.03.013

Mansoor, H., et al. (2023). Machine-learning algorithms for ischemic heart disease prediction. *PMC*. https://pmc.ncbi.nlm.nih.gov/articles/PMC10201879/

Mettananda, C., Jayasinghe, M. W., Liyanage, I. K., Ranasinghe, M. S., Gunawardena, K., & Wijesundara, W. M. H. S. (2024). Development and validation of a cardiovascular risk prediction model for Sri Lankans using machine learning. *PLoS ONE, 19*(10), e0309843. https://doi.org/10.1371/journal.pone.0309843

Reddy, K. V. V., Elamvazuthi, I., Aziz, A. A., Paramasivam, S., Chua, H. N., & Pranavanand, S. (2021). Heart disease risk prediction using machine learning classifiers with attribute evaluators. *Applied Sciences, 11*(18), 8352. https://doi.org/10.3390/app11188352

ResearchGate.net. (2021). Cardiovascular disease prediction using recursive feature elimination and gradient boosting classification techniques. https://www.researchgate.net/publication/352479655

Restivo, L., et al. (2022). ML-driven precision medicine in IHD. *Journal of Clinical Medicine, 11*(15), 4567. https://doi.org/10.3390/jcm11154567

Shahid, N., Rappon, T., & Berta, W. (2020). Applications of artificial neural networks in health care organizational decision-making: A scoping review. *PloS One, 15*(2), e0227811. https://doi.org/10.1371/journal.pone.0227811

Shishehbori, F., & Awan, Z. (2024). Enhancing cardiovascular disease risk prediction with machine learning models. *arXiv preprint* arXiv:2402.00753.

Wang, G., Zhang, Y., Li, S., Zhang, J., Jiang, D., Li, X., ... & Du, J. (2021). A machine learning-based prediction model for cardiovascular risk in women with preeclampsia. *Frontiers in Cardiovascular Medicine*, *8*, 736491.

Wanninayake, A., & Wickramasinghe, S. (2025, March). Predicting the risk of chronic kidney disease (CKD) in individuals in Medawachchiya and surrounding areas using machine learning: "AAROGYA". In Proceedings of the International Conference on Advanced Materials for Clean Energy and Health.

Wu, Y., Zhang, Z., & Xu, Y. (2021). Cardiovascular risk prediction using machine learning algorithms. *BMC Medical Informatics and Decision Making, 21*(1), 66. https://doi.org/10.1186/s12911-021-01421-1

Zhou, X., Li, Y., & Ma, Y. (2019). Heart disease prediction model based on decision tree and improved KNN. *Journal of Healthcare Engineering, 2019*, 1–10. https://doi.org/10.1155/2019/3831408

Silveri, G., Merlo, M., Restivo, L., De Paola, B., Miladinović, A., Ajčević, M., Sinagra, G., & Accardo, A. (2020). Identification of ischemic heart disease by using machine learning techniques based on parameters measuring heart rate variability. *arXiv*. Retrieved from https://arxiv.org/abs/2010.15893 researchgate.net+15arxiv.org+15pmc.ncbi.nlm.nih.gov+15

Theerthagiri, P., & Vidya, J. (2021). Cardiovascular disease prediction using recursive feature elimination and gradient boosting classification techniques. *arXiv*. Retrieved from https://arxiv.org/abs/2106.08889

Kwakye, K., & Dadzie, E. (2021). Machine learning-based classification algorithms for the prediction of coronary heart diseases. *arXiv*. Retrieved from https://arxiv.org/abs/2112.01503

Park, S., et al. (2023). A predictive model of ischemic heart disease in middle-aged and older women using data mining techniques. *Preventive Medicine Reports*. *PMC*. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10143631/