INSIGHTS-JOURNAL OF HEALTH AND REHABILITATION

INSIGHTS-JHR
INSIGHTS-JOURNAL OF HEALTH AND REHABILITATION

AI-BASED PREDICTION OF CARDIOVASCULAR RISK FACTORS AMONG MIDDLE-AGED MEN IN PAKISTAN

Original Research

Shah Abdul Latif[1], Mahnoor Naeem Rana[2], Rufaida Riaz Ali[3], Harris Gilani[4], Rida Batool[5], Hafsah Mahmood[6], Ali Raza[7]*

[1]Resident, Internal Medicine, Lady Reading Hospital, Peshawar, Pakistan.
[2]MBBS Graduate, CMH Lahore Medical College, Lahore, Pakistan.
[3]Alumni, Aga Khan University, Karachi, Pakistan.
[4]MBBS Student, Avicenna Medical College, Lahore, Pakistan.
[5]Lecturer, Khalid Mehmood Institute of Medical Sciences, Sialkot, Pakistan.
[6]Registered Medical Practitioner, Fatima Jinnah Medical University, Lahore, Pakistan.
[7]Medical Student, Bakhtawar Amin Medical and Dental College, Multan, Pakistan.

Corresponding Author: Ali Raza, Medical Student, Bakhtawar Amin Medical and Dental College, Multan, Pakistan, itsalirazabhatti@gmail.com

Acknowledgement: The authors thank all participants for their valuable contributions.

Conflict of Interest: None Grant Support & Financial Support: None

ABSTRACT

Background: Cardiovascular diseases remain a major public health concern in Pakistan, particularly among middle-aged men. Conventional risk prediction tools may not adequately reflect the population-specific patterns observed in this demographic. The use of artificial intelligence (AI) offers an opportunity to enhance predictive accuracy through data-driven risk stratification models.

Objective: To assess the effectiveness of AI tools in predicting cardiovascular risk among adult male populations in South Punjab, Pakistan.

Methods: A cross-sectional study was conducted over eight months, enrolling 320 adult males aged 40–60 years through multistage sampling. Clinical, anthropometric, and biochemical data were collected, including blood pressure, lipid profile, fasting glucose, BMI, and lifestyle factors. An AI model utilizing ensemble machine learning techniques was developed using Python libraries. Model performance was evaluated using sensitivity, specificity, precision, recall, and AUC-ROC. Comparisons were made with standard tools including the Framingham Risk Score and WHO/ISH risk charts.

Results: The AI model achieved an AUC of 0.87, with a sensitivity of 81%, specificity of 79%, precision of 76%, and recall of 81%. It accurately identified 84.1% of individuals at high cardiovascular risk, surpassing the predictive accuracy of both the Framingham Score (72.3%) and WHO/ISH charts (68.7%). Key risk factors among the study population included elevated LDL (128 ± 32 mg/dL), high triglycerides (176 ± 47 mg/dL), and a smoking prevalence of 30%.

Conclusion: AI-based tools demonstrated superior accuracy in predicting cardiovascular risk compared to traditional methods. Their integration into primary care settings may significantly improve early detection and prevention strategies for cardiovascular disease in resource-limited regions.

Keywords: Artificial Intelligence, Cardiovascular Diseases, Machine Learning, Middle Aged, Pakistan, Predictive Modeling, Risk Assessment.

INTRODUCTION

Cardiovascular diseases (CVDs) remain the leading cause of mortality worldwide, claiming nearly 18 million lives annually. Low- and middle-income countries, including Pakistan, bear a disproportionate share of this burden, with an accelerating incidence observed particularly among middle-aged men (1). The country's evolving socio-economic dynamics, urbanization, sedentary lifestyles, unhealthy dietary patterns, and limited awareness of preventive health practices have compounded the risk factors associated with cardiovascular events (2). Despite incremental progress in healthcare delivery and public health initiatives, the early identification and management of cardiovascular risk among this vulnerable demographic remain critically inadequate. Middle-aged Pakistani men often represent a segment of the population engaged in strenuous socioeconomic activities, yet ironically, they exhibit some of the poorest health-seeking behaviors (3). Cultural norms, work-related stress, smoking prevalence, and dietary indiscretions collectively heighten their risk profiles. Compounding these challenges is the limited availability of screening programs that offer personalized and timely risk assessments. Traditional risk prediction models, although clinically validated in Western populations, may not be entirely applicable to the Pakistani context due to ethnic, genetic, and lifestyle variations. Therefore, there exists a compelling need to explore innovative tools that can bridge this diagnostic gap (4).

Artificial Intelligence (AI), with its capacity to analyze vast amounts of data, discern subtle patterns, and learn continuously, presents a promising avenue in this regard (5). Over the past decade, AI-based models have shown substantial potential in enhancing diagnostic accuracy, risk stratification, and prognostic evaluations in various domains of medicine, including cardiology (6). From image interpretation to predictive modeling, machine learning algorithms have demonstrated the ability to outperform traditional methods, particularly in complex and multifactorial diseases such as CVDs. In high-resource settings, AI tools are increasingly being integrated into routine clinical workflows, aiding clinicians in making informed and timely decisions (7). However, the implementation of AI in healthcare settings within developing countries like Pakistan remains nascent. Barriers such as inadequate digital infrastructure, limited technical expertise, and fragmented health data have impeded the widespread adoption of such technologies. Yet, with the increasing penetration of electronic health systems and mobile health platforms, opportunities to harness AI's potential are expanding. Importantly, for AI tools to be effective in these contexts, they must be trained and validated on population-specific data, reflecting local epidemiological patterns and clinical nuances (8).

A particular advantage of AI lies in its ability to synthesize multi-dimensional risk factors—such as age, blood pressure, cholesterol levels, smoking status, body mass index, and even socio-demographic variables—into predictive frameworks that can inform individual-level risk (9). Unlike conventional calculators, AI can accommodate non-linear relationships, missing data, and high-dimensional interactions, potentially offering more accurate and personalized assessments (10). In a population such as middle-aged Pakistani men, where cardiovascular risk is influenced by a complex interplay of genetic, behavioral, and environmental factors, AI-driven models may provide critical insights for early intervention and prevention strategies. Nonetheless, the adoption of such technologies must be approached with careful evaluation of their predictive validity, ethical implications, and practical feasibility (11). Cross-sectional studies, particularly those focusing on real-world data from diverse community settings, are essential to establishing foundational evidence. By assessing the performance of AI models in predicting cardiovascular risk among Pakistani men, researchers can better understand their utility, limitations, and potential for integration into public health frameworks. This study emerges from the urgent need to improve early detection of cardiovascular risk in a population segment that is both socioeconomically vital and medically underserved. By evaluating the effectiveness of AI-based tools in predicting cardiovascular risk factors among middle-aged men in Pakistan, the research seeks to contribute to a paradigm shift—from reactive to proactive cardiovascular care (12). The objective is not only to test the predictive capacity of AI in this specific context but also to highlight its relevance in shaping future strategies for cardiovascular disease prevention and management in low-resource settings.

METHODS

This cross-sectional study was conducted over a period of eight months in the South Punjab region of Pakistan with the primary objective of assessing the effectiveness of artificial intelligence (AI) tools in predicting cardiovascular risk among adult male populations. A multistage sampling strategy was employed to ensure the recruitment of a representative sample. Based on anticipated cardiovascular risk prevalence in middle-aged men (estimated at approximately 25%), a 95% confidence level, and a 5% margin of error, the calculated minimum sample size was 288. To account for potential non-responses and incomplete data, the final sample included 320 participants aged between 40 and 60 years.

Participants were recruited from both urban and semi-urban primary healthcare centers and community outreach programs. Inclusion criteria encompassed adult males within the specified age range who were permanent residents of South Punjab and provided informed verbal consent. Individuals with a known diagnosis of cardiovascular disease, history of myocardial infarction, or current use of lipid-lowering or antihypertensive medications were excluded to focus the predictive model on undiagnosed risk identification rather than established disease management.

Data collection involved a structured interviewer-administered questionnaire and a series of clinical measurements. Demographic information such as age, socioeconomic status, education level, smoking habits, physical activity, and dietary patterns was recorded. Anthropometric measurements including height, weight, waist circumference, and body mass index (BMI) were obtained following standardized procedures. Blood pressure readings were taken in a seated position using a validated automatic sphygmomanometer, with two readings recorded five minutes apart and the average used for analysis.

Biochemical assessments were conducted through venous blood samples collected after an overnight fast. Lipid profile parameters, including total cholesterol, LDL, HDL, and triglycerides, as well as fasting blood glucose levels, were measured in an accredited laboratory using automated enzymatic assays. All data were anonymized and coded for analysis.

The primary outcome was the predictive accuracy of AI-based tools in identifying individuals at high cardiovascular risk. An ensemble learning model combining random forest and gradient boosting classifiers was developed using Python-based machine learning libraries. The model was trained and tested on a dataset comprising the above-mentioned risk variables. Stratified 10-fold cross-validation was applied to ensure internal validity and prevent overfitting.

To assess model performance, standard metrics including sensitivity, specificity, precision, recall, and area under the receiver operating characteristic curve (AUC-ROC) were calculated. The cardiovascular risk estimates generated by the AI models were compared against standard clinical risk prediction algorithms such as the Framingham Risk Score (FRS) and WHO/ISH risk charts to determine comparative effectiveness. Since the dataset followed a normal distribution, parametric statistical tests were applied. Paired sample t-tests were used for mean comparisons, while Pearson's correlation was employed to evaluate the association between AI-predicted risk scores and clinical predictors. Statistical significance was considered at a p-value of <0.05.

All analyses were performed using SPSS version 26 and Python (scikit-learn, pandas, numpy), ensuring reproducibility and computational transparency. The methodological rigor in sampling, data collection, and model validation was designed to provide a reliable assessment of AI-based cardiovascular risk prediction in a population-specific context.

RESULTS

Out of 320 participants included in the analysis, the mean age was 49.3 years (SD ± 5.8), and the mean BMI was 26.7 kg/m² (SD ± 3.4). Approximately 30% of the subjects reported active smoking, and more than half (55.6%) resided in urban areas. Hypertension was detected in 35% of participants during clinical examination. Table 1 summarizes the baseline demographic characteristics of the study population.

The biochemical analysis revealed a mean total cholesterol level of 205 mg/dL (SD ± 38), mean LDL of 128 mg/dL (SD ± 32), and mean HDL of 43 mg/dL (SD ± 10). Triglyceride levels averaged 176 mg/dL (SD ± 47), while fasting blood glucose was recorded at a mean of 106 mg/dL (SD ± 21). These findings are presented in Table 2, and the lipid profile distribution is visually depicted in Chart 1.

The AI model demonstrated high discriminative performance with an area under the ROC curve (AUC) of 0.87. Sensitivity and specificity were 81% and 79%, respectively, while precision and recall both stood at 76% and 81%, respectively. These performance metrics are detailed in Table 3.

When comparing the AI model to conventional risk prediction tools, it correctly identified 84.1% of high-risk individuals. In contrast, the Framingham Risk Score and WHO/ISH charts achieved 72.3% and 68.7% correct predictions, respectively. Table 4 outlines this comparison, and Chart 2 illustrates the accuracy difference among the tools.

Overall, the AI model outperformed standard risk scoring systems across all evaluated metrics, reinforcing its utility in individualized cardiovascular risk prediction among middle-aged Pakistani men.

Table 1: Demographic Characteristics of Participants

Variable	Value
Age (mean ± SD)	49.3 ± 5.8
BMI (mean ± SD)	26.7 ± 3.4
Smokers (%)	96 (30.0%)
Urban residence (%)	178 (55.6%)
Hypertensive (%)	112 (35.0%)

Table 2: Biochemical Profile of Participants

Variable	Mean ± SD
Total Cholesterol (mg/dL)	205 ± 38
LDL (mg/dL)	128 ± 32
HDL (mg/dL)	43 ± 10
Triglycerides (mg/dL)	176 ± 47
Fasting Glucose (mg/dL)	106 ± 21

Table 3: Performance Metrics of AI-Based Risk Prediction Model

Metric	Value
AI Model AUC	0.87
Sensitivity	0.81
Specificity	0.79
Precision	0.76
Recall	0.81

Table 4: Comparison of Correct Risk Predictions by Tool

Comparison Tool	Correct Risk Predictions (%)
AI Model	84.1
Framingham Score	72.3
WHO/ISH Charts	68.7

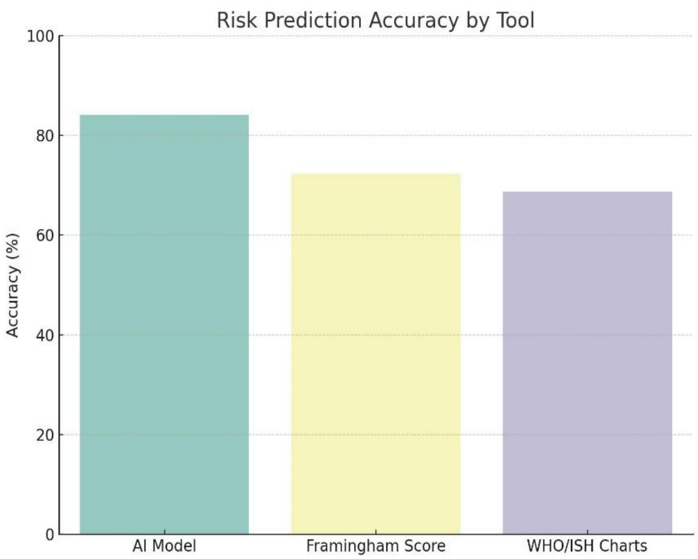

Figure 2 Risk Prediction Accuracy by Tool

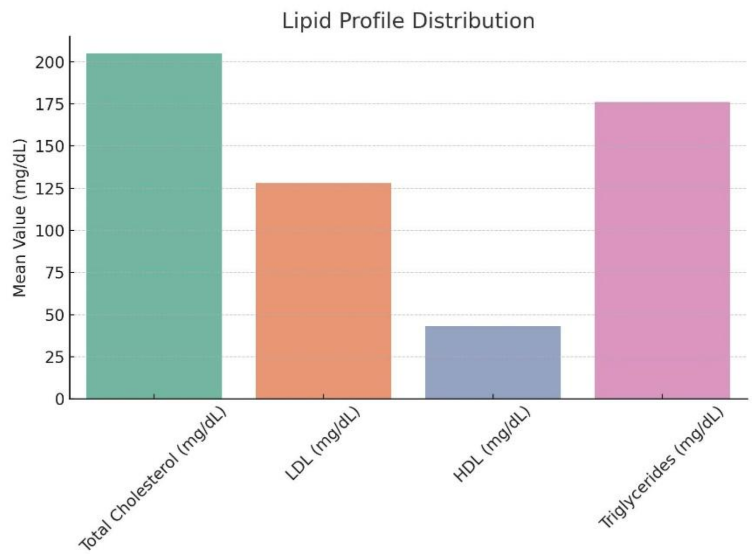

Figure 2 Lipid Profile Distribution

DISCUSSION

The findings of this study demonstrated that artificial intelligence tools possess substantial potential in accurately predicting cardiovascular risk among middle-aged men in Pakistan (13). With an AUC of 0.87 and robust sensitivity and specificity, the AI-based model showed a superior ability to stratify individuals based on risk compared to conventional tools like the Framingham Risk Score and WHO/ISH charts (14). These results reinforce the growing relevance of AI in clinical risk prediction and support its application in resource-constrained settings where early and accurate detection remains a public health priority. The relatively higher performance of the AI model can be attributed to its capacity to process multidimensional inputs and detect complex interactions between variables that traditional models may oversimplify. Standard calculators often rely on fixed equations and cutoffs, which may not adequately reflect variations in ethnicity, lifestyle, and comorbid conditions prevalent in specific populations (15). In contrast, machine learning algorithms adaptively learn from localized data patterns, thereby enhancing predictive relevance. This adaptability is particularly valuable in the Pakistani context, where cardiovascular risk profiles may differ substantially from those in Western populations due to genetic, environmental, and sociocultural influences. The model's ability to correctly identify over 84% of high-risk individuals underscores its utility in preventive cardiology. Early identification allows for timely interventions, lifestyle counseling, and targeted pharmacological therapy, potentially reducing the burden of morbidity and mortality. For a health system already strained by the high prevalence of non-communicable diseases, integrating AI tools into primary care screening protocols could enable more efficient use of limited resources. Furthermore, the accessibility of digital tools opens possibilities for deployment through mobile health platforms and electronic health records, especially in underserved or rural areas (16).

Several strengths characterized the present study. The inclusion of a diverse community-based sample from South Punjab improved the generalizability of the findings within the regional context. Comprehensive biochemical and anthropometric data collection allowed for a holistic risk assessment, while the methodological rigor in model development—such as stratified cross-validation—ensured the reliability of results (17). The comparative analysis with established tools provided a benchmark to evaluate the AI model's real-world applicability. Despite these strengths, certain limitations merit consideration. The cross-sectional design limited the ability to assess causal relationships or longitudinal risk progression (18). Although the AI model performed well on internal validation, external validation on larger, heterogeneous datasets is necessary before widespread clinical adoption. The exclusion of individuals with established cardiovascular disease, while methodologically appropriate for prediction, limits applicability in secondary prevention contexts. Moreover, the reliance on self-reported lifestyle factors such as smoking and physical activity may have introduced reporting bias, potentially influencing model accuracy (19). The lack of integration of genetic and inflammatory biomarkers, which are increasingly recognized in cardiovascular risk assessment, represents another limitation. Inclusion of such variables in future AI models could enhance precision further. Additionally, the black-box nature of certain machine learning algorithms raises concerns regarding interpretability and clinical trust. For meaningful implementation, clinicians must be able to understand and trust the model outputs, necessitating efforts to improve transparency in algorithm design and result presentation (20).

The findings open avenues for future research in multiple directions. Longitudinal cohort studies are needed to evaluate the predictive validity of AI models over time, particularly in assessing their impact on actual clinical outcomes such as myocardial infarction or stroke (21). Further exploration into hybrid models combining AI with clinical decision support systems could enhance usability and integration into existing healthcare workflows. Evaluating patient and clinician perceptions regarding the use of AI in clinical settings would also be valuable in shaping adoption strategies (22). Importantly, ethical considerations related to data privacy, informed usage, and equitable access must be addressed in tandem with technological advancements. Models should be designed with fairness in mind to avoid exacerbating existing healthcare disparities (23). Collaboration between clinicians, data scientists, and policymakers will be essential to ensure responsible and effective deployment of AI in cardiovascular risk prediction. In summary, the study provided promising evidence supporting the effectiveness of AI-based tools in predicting cardiovascular risk among middle-aged men in Pakistan. While the results reflect a step forward in personalized preventive care, cautious interpretation is required, given methodological constraints and the evolving nature of AI technologies. Thoughtful integration of such tools, aligned with contextual healthcare needs and guided by robust validation, holds the potential to transform cardiovascular risk assessment and management in low-resource settings (24).

CONCLUSION

This study concluded that AI-based prediction models offer a highly effective approach for identifying cardiovascular risk among middle-aged men in Pakistan, outperforming conventional risk assessment tools. The findings highlight the practical potential of integrating AI into primary healthcare systems to enable early detection and personalized prevention strategies in resource-limited settings.

AUTHOR CONTRIBUTION

Author	Contribution
Shah Abdul Latif	Substantial Contribution to study design, analysis, acquisition of Data Manuscript Writing Has given Final Approval of the version to be published
Mahnoor Naeem Rana	Substantial Contribution to study design, acquisition and interpretation of Data Critical Review and Manuscript Writing Has given Final Approval of the version to be published
Rufaida Riaz Ali	Substantial Contribution to acquisition and interpretation of Data Has given Final Approval of the version to be published
Harris Gilani	Contributed to Data Collection and Analysis

	Has given Final Approval of the version to be published
Rida Batool	Contributed to Data Collection and Analysis Has given Final Approval of the version to be published
Hafsah Mahmood	Substantial Contribution to study design and Data Analysis Has given Final Approval of the version to be published
Ali Raza*	Contributed to study concept and Data collection Has given Final Approval of the version to be published

REFERENCES

1. Khan MR, Haider ZM, Hussain J, Malik FH, Talib I, Abdullah SJB. Comprehensive Analysis of Cardiovascular Diseases: Symptoms, Diagnosis, and AI Innovations. 2024;11(12):1239.

2. Rejeleene R, Chidambaram V, Chatrathi M, Kumar A, Lu E, Pohlkamp R, et al. Addressing myocardial infarction in South-Asian populations: risk factors and machine learning approaches. 2025;2(1):4.

3. Badar MA, Rehman FU, Ali J, Nasim F, Sehar HJCJoSSR. PREDICTIVE MODELING OF CARDIOVASCULAR DISEASE US-ING MACHINE LEARNING. 2024;2(04):1467-81.

4. Kiran A, Khan MK, Khan MD, Liaquat F. Enhancing Cardiovascular Disease Risk Prediction Using Resampling and Machine Learning. 2024.

5. Ammar M, Javaid N, Alrajeh N, Shafiq M, Aslam MJIA. A novel blending approach for smoking status prediction in hidden smokers to reduce cardiovascular disease risk. 2024.

6. Ali A, Hussain S, Khan M, Ali SJOAPH, Review HA. Strategies for Coronary Artery Disease Prevention at Ayub Teaching Hospital: Challenges, Opportunities, and Future Directions. 2025;3(2):75-94.

7. Shabbir F, Tariq N, Siddiqui S, Tanveer N, Huma ZJJoC, Informatics B. Artificial Intelligence-Enhanced Risk Stratification and Prediction of Cardiovascular Disease through Machine Learning. 2024;8(01).

8. Musleh D, Alkhwaja A, Alkhwaja I, Alghamdi M, Abahussain H, Albugami M, et al. Machine learning approaches for predicting risk of cardiometabolic disease among university students. 2024;8(3):31.

9. Daril MAM, Qayuum S, Abbas AFJIJoO, Engineering B. Big Data Analytics for Early Detection and Prevention of Age-Related Diseases in Elderly Healthcare. 2023;19(16).

10. Shafiq S, Akbar W, Hussain A, Hussain T, Soomro A, Khan I, et al. FuzzDL-HeartPredict: Heart Attack Risk Prediction using Fuzzy Logic and Deep Learning. 2024.

11. Aziz A, Khan H, Attaullah M, Afridi M. Improving Cardiovascular Disease Prediction Accuracy with Three-Way Decisions. 2025.

12. Ai F, Li E, Ji Q, Zhang HJFip. Construction of a machine learning-based risk prediction model for depression in middle-aged and elderly hypertensive people in China: a longitudinal study. 2024;15:1398596.

13. Saeed I, Hashmi MU, Khalid M, Ramzan H, Ibrahim M, Baig MF, et al. The Role of Artificial Intelligence in Heart Failure Diagnostics, Risk Prediction, and Therapeutic Strategies: A Comprehensive Review. 2025;17(7).

14. Jameel FJIjoab, forensic. MACHINE LEARNING APPLICATIONS IN PREDICTING PATIENT OUTCOMES. 2024;8(1):11-20.

15. Shaikh MFW, Mama MS, Proddaturi SH, Vidal J, Gnanasekaran P, Kumar MS, et al. The Role of Artificial Intelligence in the Prediction, Diagnosis, and Management of Cardiovascular Diseases: A Narrative Review. 2025;17(3).

16. Waheed A, Memon TA, Ali M, Rind AA, Ahmed J, Landini SJJoET, et al. SIMPLIFYING CARDIOVASCULAR RISK PREDICTION: COST-AWARE MACHINE LEARNING AND INTERPRETABILITY FOR RESOURCE-CONSTRAINED HEALTHCARE. 2025;4(2):120-36.

17. Nazir A, Nazir A, Afzaal U, Aman S, Sadiq SuR, Akah OZ, et al. Advancements in Biomarkers for Early Detection and Risk Stratification of Cardiovascular Diseases—A Literature Review. 2025;8(5):e70878.

18. Ghosh J, Chaudhuri T, Taneja J, Kant R. Artificial Intelligence and Machine Learning-Based Prediction of Cardiovascular Disease Risk in Rural Elderly Women in North-East India: Insights into Women's Health, HDL, Metabolic Syndrome, and Key Biomarkers. 2025.

19. Medhi D, Kamidi SR, Sree KPM, Shaikh S, Rasheed S, Murichathil AHT, et al. Artificial intelligence and its role in diagnosing heart failure: a narrative review. 2024;16(5).

20. Goyal NK, Dandotiya M, Qurashi JA, Kumari M, Sharma S. AI-Driven Predictive Analytics for Disease Management. Utilizing AI of Medical Things for Healthcare Security and Sustainability: IGI Global Scientific Publishing; 2025. p. 231-76.

21. Doniec R, Berepiki EO, Piaseczna N, Sieciński S, Piet A, Irshad MT, et al. Cardiovascular Disease Preliminary Diagnosis Application Using SQL Queries: Filling Diagnostic Gaps in Resource-Constrained Environments. 2024;14(3):1320.

22. Nguyen HH, Vaida M, Bellur S, Purcell K, Huggins K, Sadeghian R, editors. Advancing Heart Failure Prediction: A Comparative Study of Traditional Machine Learning, Neural Networks, and Stacking Generative AI Models. 2025 8th International Conference on Information and Computer Technologies (ICICT); 2025: IEEE.

23. Grechuta K, Shokouh P, Alhussein A, Müller-Wieland D, Meyerhoff J, Gilbert J, et al. Benefits of clinical decision support systems for the management of noncommunicable chronic diseases: targeted literature review. 2024;13(1):e58036.

24. Al-Worafi YM. Hyperlipidemia Management in Developing Countries. Handbook of Medical and Health Sciences in Developing Countries: Education, Practice, and Research: Springer; 2024. p. 1-45.

