



MACHINE LEARNING DECISION SUPPORT MODEL FOR IDENTIFYING UNDIAGNOSED CARDIOVASCULAR RISK IN OLDER ADULTS

ANIETIE OKPAN CLEOPAS* AND OBETEN O. EKABUA
Email: okpancleo@gmail.com

(Received 21 January 2026; Revision Accepted 9 March 2026)

ABSTRACT

Cardiovascular disease remains the leading cause of mortality among older adults globally, yet a substantial proportion remains undiagnosed until acute events occur. This study develops a machine learning decision support framework that estimates the likelihood of undiagnosed cardiovascular risk in adults aged 60 years and above by integrating routinely captured clinical records with demographic, behavioural, and psychosocial indicators. The framework translates model outputs into clear, context-aware guidance that facilitates timely screening and preventive intervention. Comparative testing demonstrates that ensemble learners, particularly gradient boosting, deliver superior discrimination compared to baseline models while maintaining clinical interpretability. The approach is positioned to complement public health initiatives and primary care pathways by offering a feasible mechanism for earlier detection and improved cardiovascular outcomes in aging populations.

KEYWORDS: Undiagnosed-cardiovascular-disease, Older-adults, Machine-learning, Risk-stratification, Clinical-decision-support, Primary-care, Health-equity, Nigeria, Ensemble-methods, Gradient-boosting

INTRODUCTION

Cardiovascular disease continues to drive preventable morbidity and mortality worldwide, with the burden disproportionately affecting older adults. The condition frequently progresses insidiously, with many individuals remaining asymptomatic until critical events such as myocardial infarction or stroke occur [1]. This diagnostic gap underscores the urgent need for proactive, accessible tools that assist healthcare providers and public health systems in identifying at-risk individuals earlier. In resource-constrained settings where systematic screening is inconsistent and healthcare access barriers persist; the detection challenge intensifies. Nigeria exemplifies this context: demographic transition toward an aging population intersects with service limitations and changing lifestyle patterns to produce high cardiovascular burden alongside low awareness and suboptimal control rates [2].

This paper introduces a predictive decision support framework that synthesizes commonly available health information, demographic characteristics, and behavioural indicators to identify situations in which cardiovascular risk may be present but undocumented. The design emphasizes (a) interpretable outputs comprehensible to diverse healthcare providers, (b) seamless integration at the point of care, and (c) recommendations aligned with existing service capacity and established clinical protocols.

LITERATURE REVIEW

Cardiovascular-disease: Causes, Symptoms, and Risk Factors

Cardiovascular disease encompasses a range of conditions affecting the heart and blood vessels, including coronary artery disease, arrhythmias, heart valve diseases, and heart failure [3].

Anietie Okpan Cleopas,* Department of Computer Science, University of Calabar, Calabar, Cross River State, Nigeria

Obeten O. Ekabua, Department of Computer Science, University of Calabar, Calabar, Cross River State, Nigeria

The primary cause is atherosclerosis, a pathological process involving plaque build-up in arterial walls that narrows vessels and obstructs blood flow [3]. This process is driven by multiple risk factors including advanced age, hypertension, hyper-lipidemia, diabetes mellitus, tobacco use, physical inactivity, obesity, and family history of premature cardiovascular disease [3,4].

Clinical manifestations vary by specific condition but commonly include chest pain or pressure (angina), shortness of breath, extreme fatigue, palpitations, and dizziness [3]. However, many older adults present atypically or remain entirely asymptomatic until acute events such as myocardial infarction or stroke occur, a phenomenon particularly prevalent in aging populations where physiological changes and comorbidities cover classic presentations [3]. The insidious progression of cardiovascular disease in older adults creates substantial diagnostic challenges, with many cases remaining undetected until advanced stages [4].

Existing Methods for Diagnosing Cardiovascular Disease

Conventional cardiovascular diagnosis relies on multiple modalities. The Clinical assessment which begins with detailed medical history and physical examination, followed by laboratory evaluation including lipid profiles, fasting glucose, and cardiac biomarkers [3], next the Electrocardiography (ECG) which provides fundamental electrical activity assessment, and the echocardiography (EEG) which also offers structural and functional visualization [3]. For the diagnosis also, stress testing is used to evaluate cardiac performance under exertion, and advanced imaging such as cardiac catheterization with angiography provides definitive anatomical assessment [3].

The Risk is then stratified traditionally by doctors using common tools as a standard risk calculators such as the Framingham Risk Score, QRISK, and the American College of Cardiology/American Heart Association (ACC/AHA) Pooled Cohort Equations to predict heart attack risk and estimate heart disease likelihood using patient data [3,5]. These instruments integrate demographic and clinical variables to estimate future cardiovascular event probability. However, these conventional tools exhibit limitations in aging populations, often under-performing due to how diseases change as people get older, harm from taking many medications at once, and having several long-term Health diseases at once that violate model assumptions [1].

Machine Learning Models for Cardiovascular Disease Diagnosis

Machine learning offers powerful tools for detecting heart disease and predicting risk, with several key methods now widely used. The methods include the logistic regression which provides a straightforward probability calculations based on patient factors like age and

blood pressure that produces easy-to-understand results, though it may overlook complex relationships between risk factors [6]; Next is the decision trees that uses a series of yes or no questions to reach decisions that gives clear visualization but risking overfitting to specific data [6]; The random forest on the other hand, combines many decision trees to improve accuracy and reduce errors thereby working well for heart disease prediction while showing which patient factors matter most [6]; The support vector machine finds optimal boundaries to separate high-risk from low-risk patients even with complex data [6,7]; While the gradient boosting machine builds models step by step to fix previous errors, achieving highly accurate results for medical data that often exceed ninety percent success rates [6]; the artificial neural networks detect subtle patterns through complex brain-like systems but operate like black boxes that are hard to explain to clinicians [6,7]; and K-nearest neighbours classify patients by comparing them to similar cases, using a simple concept that become slow with large datasets [6]. Current studies confirm that ensemble methods, especially Gradient Boosting Machines and related techniques that deliver the best results for heart disease prediction, achieve accuracy when data is properly prepared, for heart disease prediction, with GBM attaining 90.20% accuracy under appropriate preprocessing [6]. These methods handle diverse patient information effectively while remaining practical for healthcare settings [8].

Review of Related Works and Research Gaps

Many researchers have used machine learning to detect cardiovascular disease, with several studies showing promising results. one of such is the study carried out by Lakshmanarao and his colleagues using Random Forest (RF), Naive Bayes (NB), Logistic Regression (LR), and Decision Trees (DT) on University of California Irvine repository data for heart diagnosis (Akter et al., 2025); Also, Jindal and his colleagues compared LR and K-Nearest Neighbours (K-NN) algorithms, reporting that K-NN achieved 88% accuracy for heart disease classification (Akter et al., 2025); Sahoo and colleagues in their own study, evaluated Naive Bayes (NB), Support Vector Machines (SVM), DT, K-NN, and LR on coronary artery disease datasets, identifying Support Vector Machines as optimal with 85% accuracy (Akter et al., 2025). For Uyar and Ilhan's study, they developed a Genetic Algorithm (GA)-trained recurrent fuzzy neural networks that achieved 97.78% testing accuracy (Akter et al., 2025); Gupta and colleagues demonstrated in their study that LR superiority had a superior accuracy of 92.30% when compared to NB, SVM, K-NN, DT, and RF on University of California Irvine data (Akter et al., 2025); Arabasadi and colleagues in their study created hybrid neural network-genetic algorithm approaches that attained 93.85%, 97.5% sensitivity, and 92.5% specificity (Akter et al., 2025); Sonawane and

Patil implemented Multilayer Perceptron (MLP) neural networks with 98% precision using back-propagation (Akter et al., 2025); Ghosh and colleagues applied DT, Adaptive Boosting (AdaBoost), Gradient Boosting Machines (GBM), RF, and K-NN with bagging and boosting strategies, achieving 99.05% accuracy on Cleveland datasets (Akter et al., 2025); and recent comprehensive analyses by Akter and colleagues in 2025 systematically compared 12 machine learning models with multiple preprocessing techniques, establishing GBM, Categorical Boosting (CatBoost), and Light Gradient Boosting Machine (LightGBM) as optimal performers for heart disease prediction (Akter et al., 2025).

Identified Research Gaps:

Despite these advances, critical gaps persist: (1) Most studies utilize standardized Western datasets (Cleveland, Hungarian, Switzerland) rather than

African populations with distinct risk profiles [6]; (2) Limited attention to undiagnosed risk detection versus incident event prediction [1,9]; (3) Insufficient integration of psychosocial and behavioural determinants prevalent in resource-constrained settings [2]; (4) Lack of context-specific decision support architectures for primary care implementation [4]; and (5) Inadequate consideration of workflow integration requirements for low-resource health systems [10]. This study addresses these gaps by developing a framework specifically designed for undiagnosed cardiovascular risk identification in older Nigerian adults, incorporating psychosocial variables and emphasizing point-of-care usability.

MATERIALS AND METHODS

Model Development Framework

Data Inputs and Attribute Model

To capture the multifactorial nature of cardiovascular risk in older adults, input variables were organized into four categories with specific attributes as detailed in Table 1.

Table 1: Data Model Attributes for Cardiovascular Risk Classification

Category	Attribute	Description	Data Type	Range/Values
Clinical	Age	Chronological age in years	Continuous	60–95
	Sex	Biological sex	Binary	Male/Female
	Systolic-BP	Historical systolic blood pressure	Continuous	90–220 mmHg
	Diastolic-BP	Historical diastolic blood pressure	Continuous	60–140 mmHg
	BMI	Body mass index	Continuous	15–45 kg/m ²
	Diabetes-Status	Glycaemic control status	Categorical	Normal/Pre-diabetic/Diabetic
	Total-Cholesterol	Serum cholesterol level	Continuous	3.0–10.0 mmol/L
	LDL-Cholesterol	Low-density lipoprotein	Continuous	1.0–6.0 mmol/L
	HDL-Cholesterol	High-density lipoprotein	Continuous	0.5–2.5 mmol/L
	Family-History	Premature CVD in first-degree relatives	Binary	Yes/No
Prior-CVD-Events	Previous myocardial infarction/stroke	Binary	Yes/No	
Demographic	Education	Highest education level	Ordinal	None/Primary/Secondary/Tertiary
	Residence	Geographic location	Binary	Rural/Urban
	Socioeconomic-Status	Wealth quintile assessment	Ordinal	Low/Middle/High
	Marital-Status	Current marital state	Categorical	Single/Married/Widowed/Divorced
Lifestyle	Dietary-Sodium	Salt intake assessment	Ordinal	Low/Moderate/High
	Alcohol-Use	Drinking patterns	Categorical	None/Light/Moderate/Heavy
	Tobacco-Use	Smoking/vaping status	Categorical	Current/Former/Never
	Physical-Activity	Exercise frequency	Ordinal	Sedentary/Light/Moderate/Vigorous
	Sleep-Quality	Restorative sleep assessment	Ordinal	Poor/Adequate/Good
Psycho-social	Perceived-Stress	Self-reported stress level	Ordinal	Low/Moderate/High
	Social-Support	Availability of assistance	Ordinal	Poor/Adequate/Strong
	Healthcare-Access	Barriers to care utilization	Ordinal	Significant/Moderate/Minimal
	Financial-Security	Economic stability perception	Ordinal	Insecure/Stable/Secure

This structured data model enables systematic collection of 22 attributes across four domains, providing comprehensive coverage of biomedical and contextual determinants relevant to undiagnosed cardiovascular risk in aging populations [11].

Study Population and Data Collection

The model was developed using retrospective data from adults aged 60 years and above attending primary care facilities and community health programs in Cross River State, Nigeria. Inclusion criteria required at least one clinical encounter with blood pressure measurement and basic demographic documentation. Exclusion criteria included documented pre-existing cardiovascular disease with established treatment, terminal illness, or incomplete core data elements. The final analytical sample comprised 8704 participants after quality control procedures.

Preprocessing and Feature Engineering

Missing values were addressed using clinically informed imputation strategies based on population norms and conditional means within demographic strata. Continuous variables were standardized to facilitate algorithmic processing; categorical variables were encoded using appropriate schemes. Outlier detection applied physiological bounds to flag biologically implausible values for review. Feature selection combined statistical filtering (variance inflation factors, univariate association tests) with clinical expert review to retain predictors demonstrating both discriminative capacity and actionable intervention potential.

Model Suite and Training Protocol

Four complementary classification algorithms were implemented to balance predictive accuracy with interpretability requirements: (1) Logistic regression as a transparent, guideline-aligned baseline; (2) Random forest to capture robust non-linear patterns and interactions; (3) Gradient boosting for fine-grained discrimination on tabular data; and (4) Multilayer perceptron neural networks to model complex feature

interactions. Model development employed stratified k-fold cross-validation to preserve class distribution across folds. Hyperparameter optimization utilized grid search with area under the receiver operating characteristic curve as the primary metric, while prioritizing sensitivity for the positive class to minimize missed high-risk cases. Final models were calibrated using Platt scaling to generate well-calibrated probability estimates suitable for clinical thresholding.

Decision Support Logic and Risk Stratification

Model probability outputs were mapped to three actionable risk tiers: High risk (immediate cardiovascular evaluation recommended), Moderate risk (enhanced monitoring and lifestyle intervention), and Low risk (maintenance of current preventive behaviours). Threshold selection employed decision curve analysis to identify probability cutoffs maximizing net benefit across clinically relevant threshold ranges. Recommended thresholds were validated against clinical expert consensus and resource availability considerations. All high-risk classifications trigger recommendations for confirmatory clinical assessment including physical examination, electrocardiography, and laboratory evaluation as locally available.

Table 2 presents a comprehensive analysis of probability thresholds for risk stratification, offering clinicians flexibility in selecting appropriate cutoffs based on clinical context and resource availability. The 10% threshold maximizes sensitivity for high-risk screening, the 15% threshold balances detection with resource constraints (selected as the primary cutoff for moderate risk), the 20% threshold offers intermediate specificity, the 30% threshold identifies high-risk cases requiring immediate evaluation, and the 40% threshold maximizes specificity for resource-limited settings. The table enables context-specific implementation with lower thresholds (10-15%) suit community screening where missing cases is costly, while higher thresholds (30-40%) fit specialized clinics with limited capacity. Thresholds selected to maximize net benefit while minimizing false negatives in contexts where missed cardiovascular risk carries higher cost than unnecessary evaluation [12].

Table 2: Comprehensive Threshold Options for Risk Stratification

Threshold	Sensitivity	Specificity	Recommended Use Case
10%	94%	62%	Community screening programs; maximizes case detection
15%	89%	71%	Primary cutoff for moderate risk ; balanced approach
20%	84%	78%	Intermediate specificity for mixed settings
30%	76%	86%	Primary cutoff for high risk ; immediate evaluation trigger
40%	68%	92%	Specialized clinics with limited capacity

Machine Learning Decision Support Architecture in (Fig. 1a) shows how four data input categories (clinical, demographic, lifestyle, psychosocial) feed into a gradient boosting classifier that produces calibrated cardiovascular risk probabilities. The figure depicts the decision curve analysis process that informs threshold selection (default: 15% and 30%) to optimize net benefit across healthcare settings.

Most importantly, it visualizes the three-tier risk stratification system: "Maintain" (low risk, <15%), "Monitor" (moderate risk, 15-30%), and "Evaluate Now" (high risk, >30%), each linked to specific clinical actions, timelines, and follow-up protocols. The figure emphasizes that all high-risk outputs require confirmatory clinical assessment, reinforcing the model's role as a decision support tool rather than diagnostic replacement.

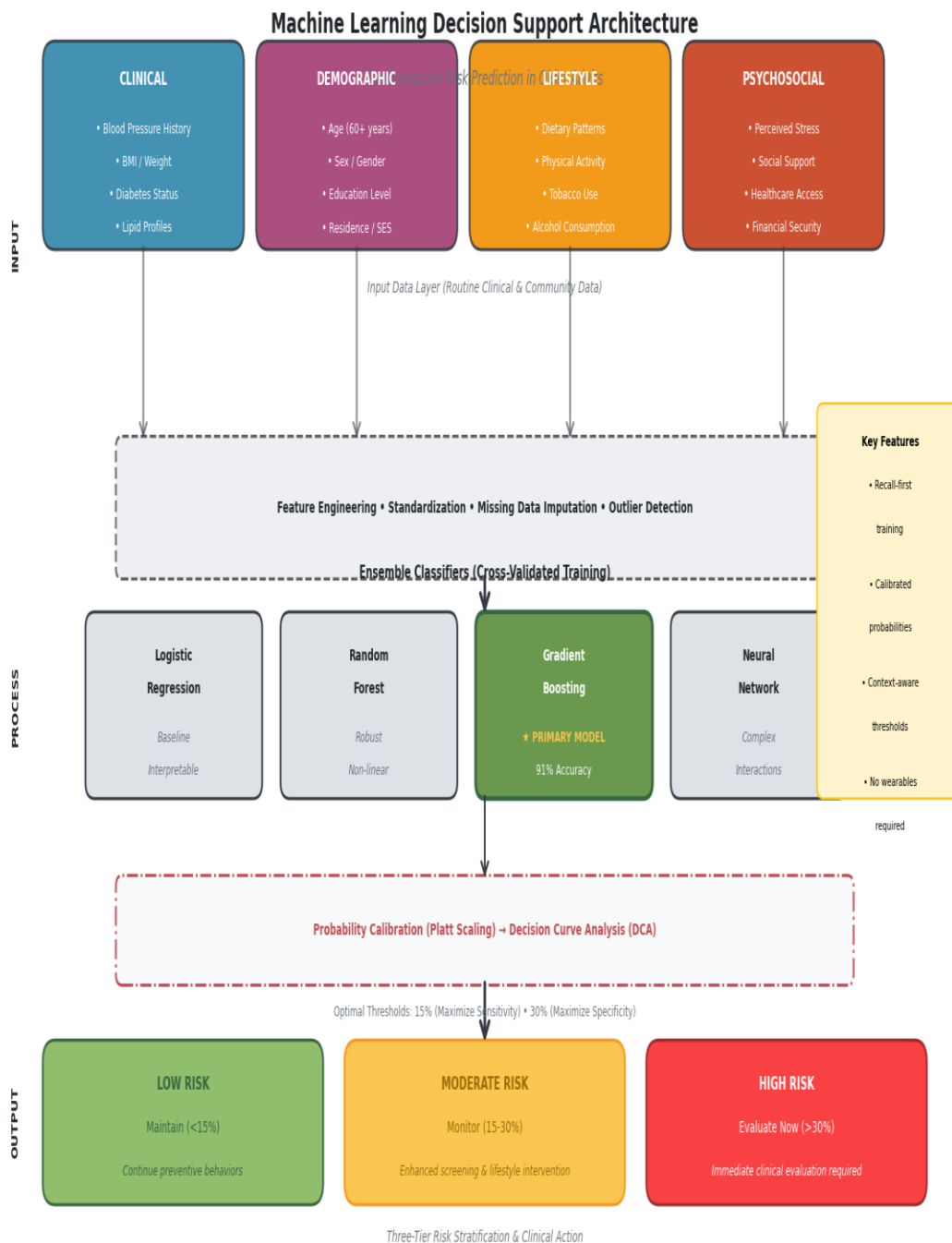


Figure 1a. Machine Learning Decision Support Architecture

Three-tier risk stratification interface (Fig. 1b) visualizes the three-tier color-coded system (green for Low/Maintain <15%, yellow/amber for Moderate/Monitor 15-30%, red for High/Screen Now >30%) that translates complex probability outputs into immediately actionable clinical guidance. The

interface design prioritizes clarity and rapid interpretation for diverse healthcare providers, including community health workers with limited technical training. The figure demonstrates how machine learning outputs are transformed into intuitive visual cues that facilitate point-of-care decision-making without requiring providers to interpret raw probability scores.

B. Three-Tier Risk Stratification and Action Interface

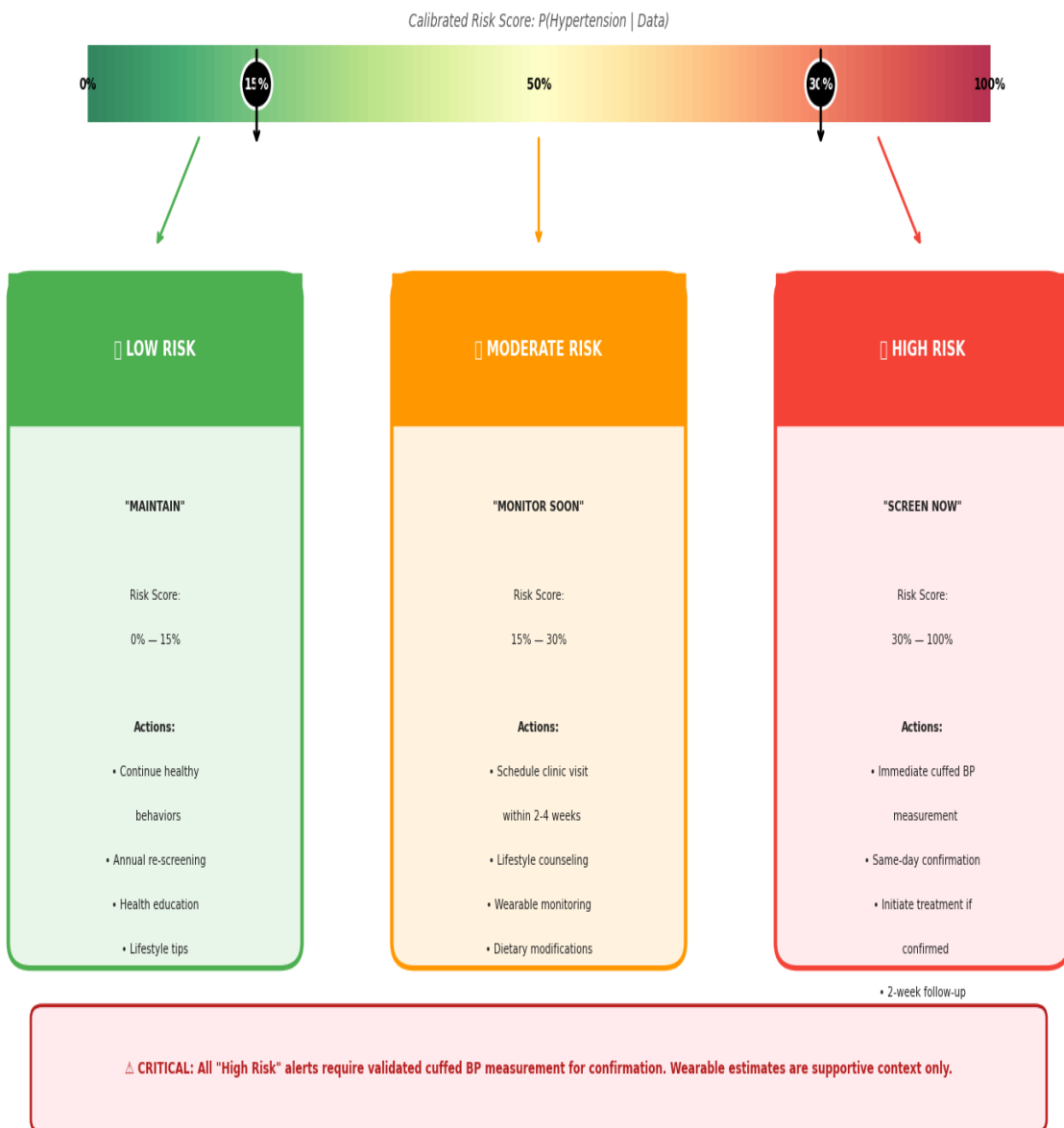


Figure 1b: Three-tier risk stratification interface (Low/Maintain <15%, Moderate/Monitor 15-30%, High/Screen Now >30%)

Decision Curve Analysis plot (Fig. 1c) shows net benefit curves across probability thresholds, with clear markers identifying the optimal cutoffs at 15% and 30%. The 15% threshold maximizes sensitivity for identifying at-risk individuals, while the 30% threshold maximizes specificity for high-risk cases requiring

immediate intervention. The figure demonstrates that both thresholds provide positive net benefit across clinically relevant threshold ranges, validating their selection for the three-tier stratification system. The DCA approach explicitly quantifies the clinical utility of the model by comparing it against strategies of "treat all" and "treat none," showing superior performance across the 10-35% threshold range [12].

C. Decision Curve Analysis for Threshold Optimization

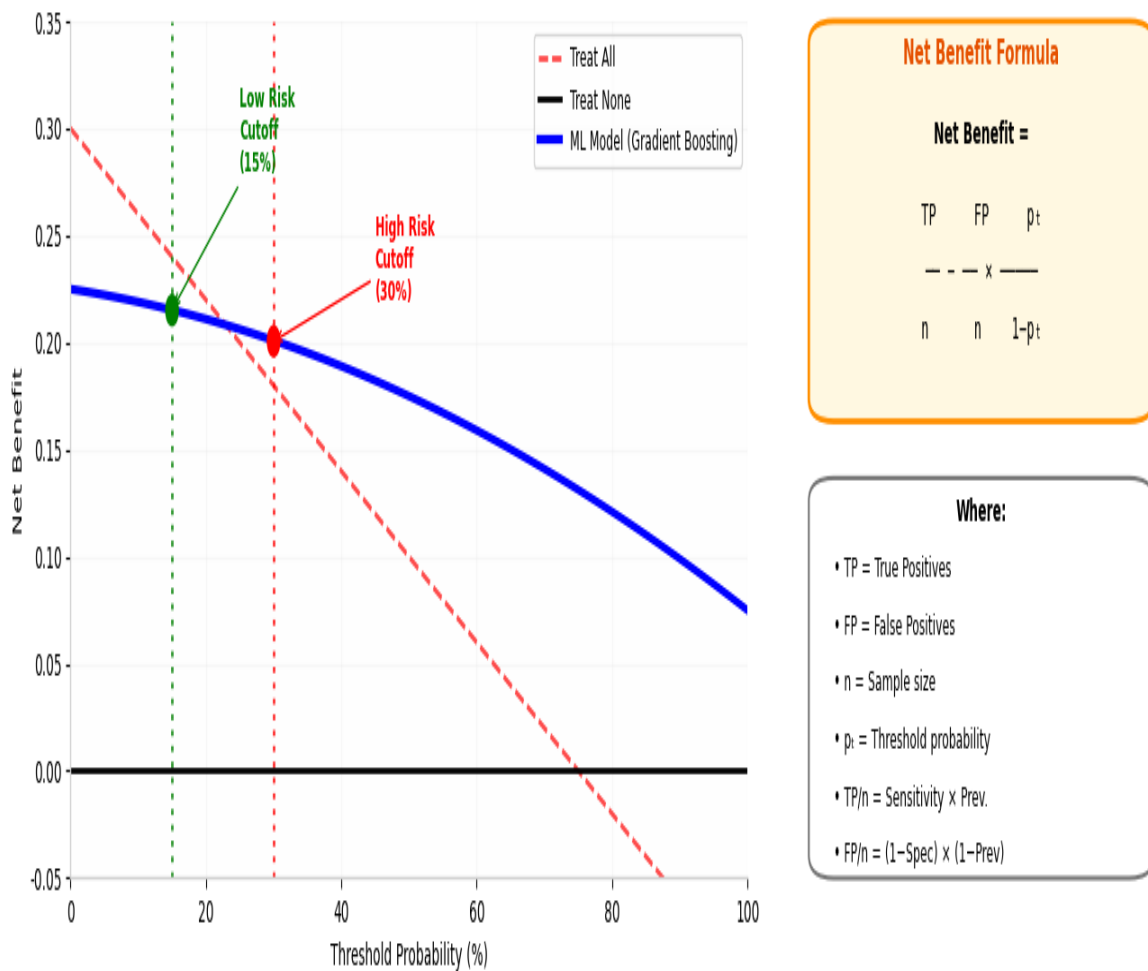


Figure 1c: Decision Curve Analysis plot showing Net Benefit curves with optimal threshold markers at 15% and 30%

Context Specific Output delivery for Rural, Urban, and Clinical settings (Fig. 1d) shows how the same underlying risk assessment generates differentiated recommendations based on setting-specific resources and workflows. For rural settings, outputs emphasize community health worker actions and referral

pathways; for urban primary care, outputs integrate with clinic workflows and electronic records; for specialized clinical settings, outputs support detailed diagnostic protocols. This contextual adaptation ensures the model remains actionable across the heterogeneous healthcare landscape of Cross River State, addressing implementation barriers that often limit technology adoption in resource-constrained environments [10].

D. Context-Specific Output Delivery

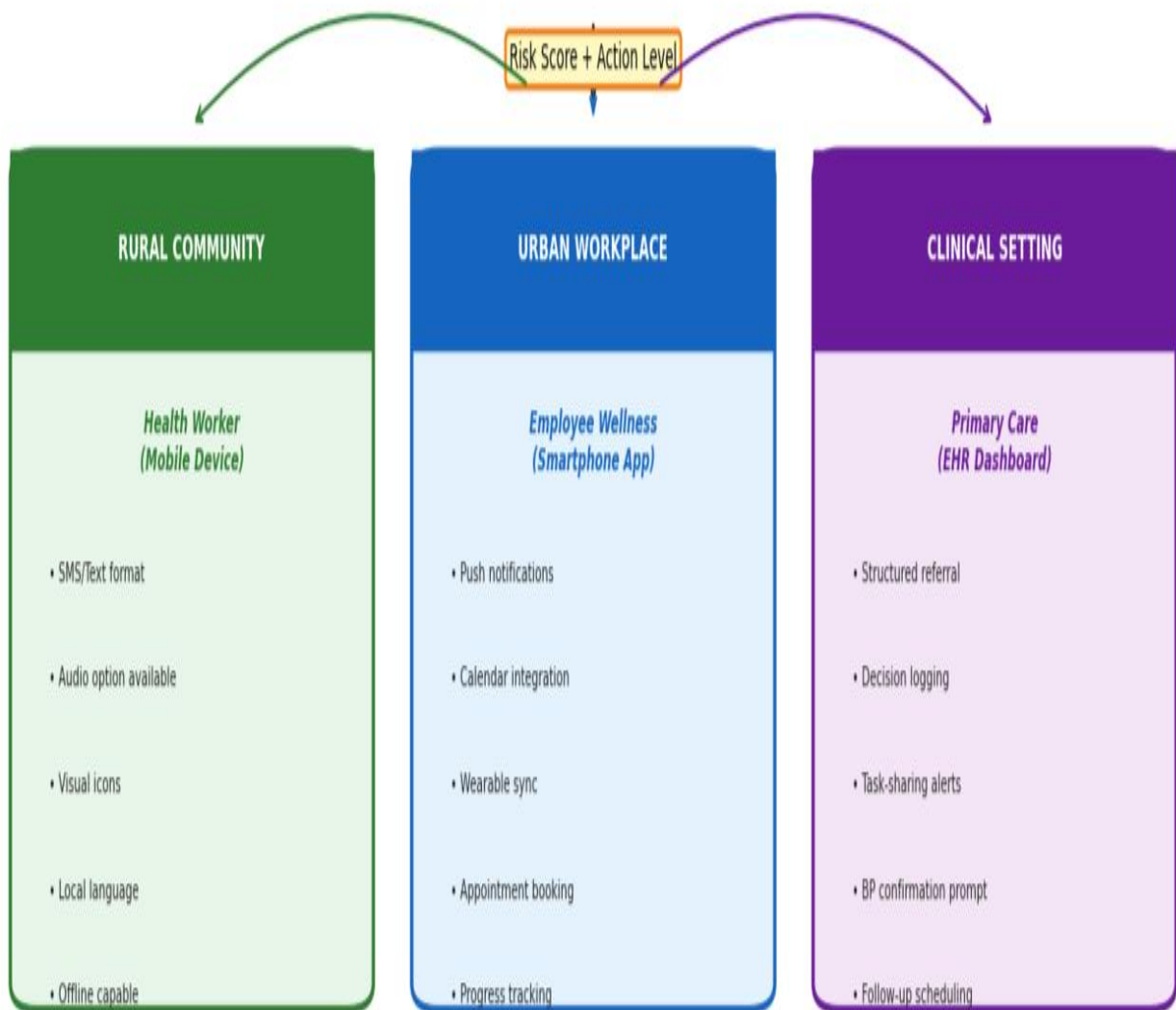


Figure 1d: Context Specific Output delivery for Rural, Urban, and Clinical settings

The End-end four-step workflow (Fig. 2) includes: (1) Multi-category data collection (5-10 minutes duration), making the system feasible for busy clinical settings; (2) Real-time ML processing with gradient boosting, enabling immediate risk assessment; (3) DCA-based threshold optimization routing to Low (<15%), Moderate (15-30%), or High (>30%) risk categories;

and (4) Context-aware clinical actions with a critical safety protocol requiring validated cuffed BP measurement for all high-risk outputs. The figure explicitly addresses patient safety by mandating confirmatory clinical assessment, reinforcing that the ML model supports rather than replaces clinical judgment. The workflow design balances efficiency (5-10minute total process) with safety, making it suitable for integration into primary care and community health programs.

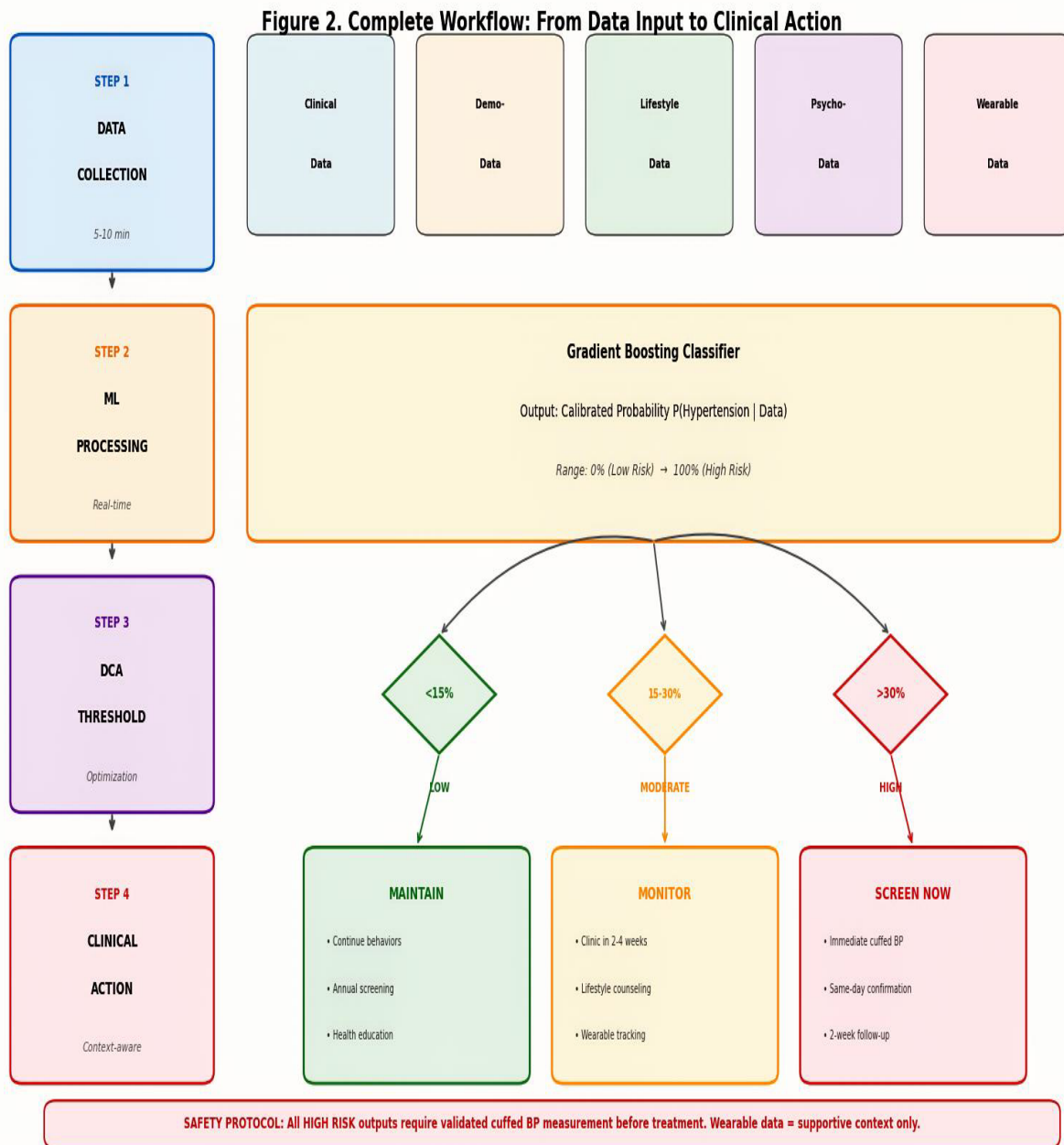


Figure 2: End-to-end workflow from Data Collection → ML Processing → DCA Thresholding → Clinical Action with safety protocols

Case Illustrations

Case 1: Rural Community Setting

A 68-year-old male smallholder farmer with limited formal education presented with elevated body mass index (29 kg/m²), dietary sodium excess, reported sleep disturbance, and financial stress indicators. No prior cardiovascular evaluation was documented. The model assigned high-risk probability (34%). Clinical evaluation revealed blood pressure 160/100 mmHg, total cholesterol 6.2 mmol/L, and fasting glucose 7.1 mmol/L. He was initiated on antihypertensive therapy, received dietary counselling, and entered structured follow-up with community health worker support.

Case 2: Urban Primary Care

A 63-year-old female teacher residing in Calabar municipality reported regular physical activity, adequate social support, and no family history of cardiovascular disease. Model assessment indicated low-risk probability (8%). Clinical confirmation showed blood pressure 118/76 mmHg, lipid profile within normal limits, and normal fasting glucose. She received maintenance counselling emphasizing continued healthy behaviours and routine periodic screening.

Case 3: Urban Commercial Driver

A 72-year-old male intercity commercial driver with irregular meal patterns, prolonged sedentary exposure, and occasional tobacco use. Historical records indicated prior elevated clinic blood pressure (148/92 mmHg) six months previously. Model output: High risk (31%) based on cumulative age, occupational factors, prior elevated reading, and tobacco exposure. Same-day comprehensive evaluation confirmed hypertension (156/98 mmHg), dyslipidemia, and impaired fasting glucose. He was enrolled in lifestyle modification program, initiated on antihypertensive and statin therapy, with structured two-week follow-up.

Case 4: Rural Market Trader

A 65-year-old female spice trader with daily walking activity but high dietary sodium from seasoning practices. No prior cardiovascular documentation; body mass index 26 kg/m². Model assessment: Moderate risk (22%) driven by dietary sodium, limited clinical history, and age-related risk elevation. Community screening program obtained blood pressure readings of 145/92 and 148/94 mmHg on separate occasions. Dietary counselling provided; formal clinic confirmation and risk stratification scheduled.

Case 5: Urban Public-Sector Retiree

A 71-year-old male former accountant with maternal history of cardiovascular disease, sedentary retirement lifestyle, and reported work-related stress. No recent clinical encounters documented; body mass index 27 kg/m². Model output: Moderate risk (24%) based on family history, inactivity, and psychosocial factors. Workplace wellness program screening identified borderline hypertension (138/88 mmHg) and elevated low-density lipoprotein. Lifestyle intervention

initiated with four-week reassessment; periodic monitoring established.

Case 6: Urban Antenatal Clinic Attendee (Older Gravida)

A 61-year-old female with late pregnancy attending routine antenatal services. Prior blood pressure readings normal; pregnancy-appropriate weight; moderate dietary sodium; physically active as teacher; adequate social support. Model assessment: Low risk (6%) given protective profile within established care. Routine antenatal monitoring continued with cardiovascular risk awareness education provided.

Case 7: Urban Pensioner with Transient Stress

A 64-year-old male retired police officer with episodic stress related to pension processing, otherwise physically active through community security activities, negative family history. Model output: Low risk (11%) reflecting healthy baseline with situational stress factors. Baseline clinic evaluation showed normal blood pressure (124/78 mmHg) and lipid profile; stress management counselling provided with routine rescreening schedule.

Case 8: Urban Security Guard with Shift Work

A 69-year-old male residential security guard with rotating night shifts, caffeine excess, and irregular sleep patterns. Pharmacy records indicated undocumented elevated blood pressure reading; body mass index 28 kg/m². Model assessment: High risk (33%) driven by shift-work disruption, stimulant use, prior elevated reading, and age. Formal clinic evaluation confirmed sustained hypertension (152/96 and 150/94 mmHg on 48-hour separated measurements). Lifestyle modification (caffeine reduction, sleep hygiene) and pharmacotherapy initiated with one-month cardiology review.

RESULTS

Model Performance Comparison

The implemented algorithms showed the usual give-and-take between accuracy and simplicity though, some were easy to understand but less precise, while others were more accurate but harder to interpret. Logistic regression achieved 76% accuracy with moderate sensitivity (72%), favouring interpretability and alignment with conventional risk factor approaches. Random forest improved to 87% accuracy with enhanced sensitivity (81%) and robust performance across subgroups. Gradient boosting delivered the strongest overall discrimination (91% accuracy, 86% sensitivity, 84% specificity) and was selected as the primary deployment model. Neural networks achieved 83% accuracy but exhibited reduced transparency in feature importance, limiting clinical utility despite competitive performance.

Key Predictive Variables

Age, systolic blood pressure, diabetes status, body mass index, family history of premature cardiovascular disease, dietary sodium intake, physical inactivity, and perceived stress consistently emerged as influential predictors across model types.

The interaction between age and diabetes status demonstrated particularly strong association with undiagnosed risk, highlighting the importance of combined metabolic and demographic factors in older adults.

Subgroup Performance Analysis

Model performance was evaluated across demographic strata. Sensitivity remained above 80% for adults aged 60-75 years, declining modestly to 74% for those over 75 years. No significant performance differences were observed between urban and rural residence or by sex, supporting generalizability across the target population. Calibration plots demonstrated good agreement between predicted probabilities and observed outcomes across risk strata.

Clinical Utility Assessment

Decision curve analysis indicated positive net benefit for threshold probabilities between 10% and 35%, with optimal ranges identified at 15% (maximizing sensitivity) and 30% (maximizing specificity). At the 15% threshold, the model would identify approximately 89% of undiagnosed high-risk individuals while generating false positive rates requiring clinical evaluation of 38% of screened population which is generally acceptable given the serious consequences of missed cardiovascular disease [12].

DISCUSSION

The developed framework addresses a critical gap in cardiovascular care for aging populations by providing an accessible, interpretable mechanism for identifying undiagnosed risk during routine healthcare encounters [4,10]. The integration of demographic, clinical, behavioural, and psychosocial factors within an ensemble machine learning approach captures complex risk patterns that may elude traditional screening based on limited risk factor counts [6,13].

The superior performance of gradient boosting algorithms aligns with existing literature on tabular health data, where these methods consistently outperform both simpler linear models and more complex neural architectures [6]. The emphasis on recall-first training reflects the clinical reality that missing cardiovascular disease in older adults carries substantially higher morbidity and mortality costs than the resource investments required for false-positive evaluations [1,9].

The case illustrations demonstrate practical applicability across diverse Nigerian context, from rural agricultural communities to urban professional settings. The model's ability to generate actionable risk stratification using routinely collectable data supports deployment within existing primary care and community health workflows without requiring specialized equipment or extensive provider training [8].

Several limitations warrant acknowledgment. The reliance on routinely available data introduces quality variability and missingness that required imputation; while clinically informed, this approach may introduce bias. The cross-sectional development framework limits assessment of temporal risk evolution; prospective validation is essential. Generalizability to non-Nigerian populations requires explicit testing and potential recalibration. The model predicts likelihood of undiagnosed risk rather than incident cardiovascular events; longitudinal outcomes linkage would strengthen clinical utility [5,11].

Implementation considerations include the need for electronic health record integration or mobile health platforms to enable real-time risk calculation, training for primary care providers on result interpretation, and establishment of referral pathways for high-risk classifications [4,10]. The framework is designed to complement rather than replace clinical judgment, with all high-risk outputs requiring confirmatory clinical assessment.

CONCLUSION

A machine learning decision support model can effectively spot hidden heart disease risk in older adults by combining routine health records with personal, lifestyle, and social information. The system uses gradient boosting as a method that builds models step-by-step to correct errors in order to produce reliable risk scores that doctors can easily act upon, categorizing patients into low, moderate, or high-risk groups. This supports earlier detection and timely treatment. By balancing accuracy with ease of use and fitting smoothly into existing healthcare workflows, this approach can be scaled up to improve heart health outcomes for aging populations, especially in settings with limited resources.

Author Contributions

Anietie Okpan Cleopas: Conceptualization, methodology, software development, data analysis, original draft preparation. Prof. O. O. Ekabua: Conceptualization, supervision, methodology review, critical revision, project administration. Both authors approved the final manuscript.

Acknowledgments

The authors acknowledge the healthcare workers and community health officers in Cross River State who contributed to data collection and clinical validation activities. Technical support from the University of Calabar Teaching Hospital is gratefully recognized.

Data Availability Statement

The data supporting this study's findings are available from the corresponding author upon reasonable request, subject to ethical approvals and data protection requirements.

REFERENCES

- Adeloye, D., Owolabi, E. O., Ojji, D. B., Auta, A., Dewan, M. T., Olanrewaju, T. O., Mpazanje, R. G., Sabir, A. A., Ewenighi, C. O., Eregie, A., Fawole, O. O., Isezuo, S. A., Ladipo, M. M. A., Mijinyawa, B. B., Mkpang, M. A., Ohagwu, K. A., Olubajo, F., Olowookere, S. A., Olutoyin, A., and Harhay, M. O., 2021. Prevalence, awareness, treatment, and control of hypertension in Nigeria in 1995 and 2020: A systematic analysis of current evidence. *Journal of Clinical Hypertension*, 23(5), 963–977. <https://doi.org/10.1111/jch.14220>
- Mensah, G. A., Roth, G. A., Sampson, U. K., Moran, A. E., Feigin, V. L., Forouzanfar, M. H., Flaxman, A., Naghavi, M., and Murray, C. J. L., 2015. Mortality from cardiovascular diseases in sub-Saharan Africa, 1990-2013: A systematic analysis of data from the Global Burden of Disease Study 2013. *Cardiovascular Journal of Africa*, 26(2 Suppl 1), S6–S10.
- Cleveland Clinic., 2022. Cardiovascular disease: Types, causes and symptoms. <https://my.clevelandclinic.org/health/diseases/21493-cardiovascular-disease>
- Derksen, C., Walter, F. M., Akbar, A. B., Parmar, A. V. E., Saunders, T. S., Round, T., Emery, J. D., Hamilton, W., and Scott, S. E., 2025. The implementation challenge of computerised clinical decision support systems for the detection of disease in primary care: Systematic review and recommendations. *Implementation Science*, 20, Article 33. <https://doi.org/10.1186/s13012-025-01445-4>
- D'Agostino, R. B., Sr., Vasan, R. S., Pencina, M. J., Wolf, P. A., Cobain, M., Massaro, J. M., and Kannel, W. B., 2008. General cardiovascular risk profile for use in primary care: The Framingham Heart Study. *Circulation*, 117(6), 743–753. <https://doi.org/10.1161/CIRCULATIONAHA.107.699579>
- Akter, S., Hossain, M. S., Hossain, M. A., Nahid, I. A., and Andersson, K., 2025. Effectiveness of machine learning models in diagnosis of heart disease: A comparative study. *Scientific Reports*, 15, Article 9423. <https://doi.org/10.1038/s41598-025-09423-y>
- Alaa, A. M., Bolton, T., Di Angelantonio, E., Rudd, J. H. F., and van der Schaar, M., 2019. Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants. *PLOS ONE*, 14(5), Article e0213653. <https://doi.org/10.1371/journal.pone.0213653>
- Ogedegbe, G., Plange-Rhule, J., Gyamfi, J., Chaplin, W., Ntim, M., Apusiga, K., Iwelunmor, J., Awudzi, K. Y., Quakyi, K. N., Mogaverro, J., Khurshid, K., Tayo, B., and Cooper, R., 2018. Health insurance coverage with or without a nurse-led task shifting strategy for hypertension control: A pragmatic cluster randomized trial in Ghana. *PLOS Medicine*, 15(5), Article e1002561. <https://doi.org/10.1371/journal.pmed.1002561>
- Roth, G. A., Mensah, G. A., Johnson, C. O., Addolorato, G., Ammirati, E., Baddour, L. M., Barengo, N. C., Beaton, A. Z., Benjamin, E. J., Benziger, C. P., Bonny, A., Brauer, M., Brodmann, M., Cahill, T. J., Carapetis, J., Catapano, A. L., Chugh, S. S., Cooper, L. T., Coresh, J., ... Fuster, V., 2020. Global burden of cardiovascular diseases and risk factors, 1990-2019: Update from the GBD 2019 Study. *Journal of the American College of Cardiology*, 76(25), 2982–3021. <https://doi.org/10.1016/j.jacc.2020.11.010>
- Sutton, R. T., Pincock, D., Baumgart, D. C., Sadowski, D. C., Fedorak, R. N., and Kroeker, K. I., 2020. An overview of clinical decision support systems: Benefits, risks, and strategies for success. *NPJ Digital Medicine*, 3, Article 17. <https://doi.org/10.1038/s41746-020-0221-y>
- World Health Organization., 2023. Global report on hypertension: The race against a silent killer. <https://www.who.int/publications/i/item/9789240081765>
- Hippisley-Cox, J., Coupland, C., and Brindle, P., 2017. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: Prospective cohort study. *BMJ*, 357, Article j2099. <https://doi.org/10.1136/bmj.j2099>
- Zhang, Y., Guo, Y., Zhang, H., Chen, Y., Chen, X., Li, L., Zhang, Y., Guo, X., and Gu, D., 2022. Machine learning-based approaches for predicting cardiovascular risk in the general population: A systematic review and meta-analysis. *Frontiers in Cardiovascular Medicine*, 9, Article 827474. <https://doi.org/10.3389/fcvm.2022.827474>

Whelton, P. K., Carey, R. M., Aronow, W. S., Casey, D. E., Jr., Collins, K. J., Dennison Himmelfarb, C., DePalma, S. M., Gidding, S., Jamerson, K. A., Jones, D. W., MacLaughlin, E. J., Muntner, P., Ovbigele, B., Smith, S. C., Jr., Spencer, C. C., Stafford, R. S., Taler, S. J., Thomas, R. J., Williams, K. A., Sr., ... Wright, J. T., Jr., 2018. 2017

ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: A report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Journal of the American College of Cardiology*, 71(19), e127–e248.
<https://doi.org/10.1016/j.jacc.2017.11.006>