pISSN:2349-5707 Volume:13,Issue 1, January -2026 eISSN:2349-5715

MACHINE LEARNING MODELS FOR EARLY PREDICTION OF CARDIOVASCULAR DISEASES IN PRIMARY CARE SETTINGS

[1]Maxsudov Valijon Gafurjonovich, [1]Normamatov Sardor Faxriddin ugli, [2]BMatxiddinova Nafosat
[1]Associate Professor, Department of Biomedical Engineering, Informatics and Biophysics, Tashkent State Medical University,
[2]Fundamental medicine direction, student of group 103-A, Tashkent State Medical University

Resume. Early prediction of cardiovascular diseases (CVDs) in primary care is crucial for effective prevention and timely intervention. Conventional risk assessment methods often have limited accuracy due to their inability to analyze complex clinical data. This study examines the use of machine learning (ML) models for early prediction of cardiovascular diseases in primary care settings. Demographic data, clinical indicators, laboratory results, and electronic health records were used to train and evaluate multiple ML algorithms, including logistic regression, random forest, and gradient boosting models. Model performance was assessed using accuracy, sensitivity, specificity, and the area under the receiver operating characteristic curve (AUC). The results show that ML-based models outperform traditional risk assessment tools in identifying high-risk patients. The integration of explainable AI methods improves model transparency and supports clinical decision-making. Overall, machine learning approaches demonstrate strong potential as decision support tools for enhancing cardiovascular disease prevention in primary healthcare.

Keywords: machine learning, cardiovascular diseases, early prediction, primary care, clinical decision support, preventive healthcare, explainable artificial intelligence.

Introduction. Cardiovascular diseases (CVDs) remain the leading cause of morbidity and mortality worldwide, placing a substantial burden on healthcare systems, particularly at the primary care level. Primary care settings play a critical role in the early identification and management of cardiovascular risk factors; however, timely detection of individuals at high risk remains challenging. Traditional risk assessment tools, such as score-based models derived from population averages, often fail to account for the complex and nonlinear interactions among demographic, clinical, behavioral, and laboratory variables. As a result, a significant proportion of high-risk patients may remain undetected until the disease progresses to more severe stages. The rapid growth of electronic health records (EHRs) and routinely collected clinical data has created new opportunities for data-driven approaches in preventive cardiology. Machine learning (ML), a key component of artificial intelligence, enables the analysis of large, heterogeneous datasets and the identification of subtle patterns that are not easily captured by conventional statistical methods. In recent years, ML models have demonstrated promising results in disease prediction, risk stratification, and clinical decision support across various medical domains. Despite these advances, the integration of machine learning into primary care practice faces important challenges, including model interpretability, clinical trust, and practical implementation. This study focuses on the application of machine learning models for the early prediction of cardiovascular diseases in primary care settings, aiming to improve risk assessment accuracy and support preventive, patient-centered healthcare strategies.

Literature review. Early prediction of cardiovascular diseases has long been a central objective in preventive medicine, particularly within primary care. Conventional risk assessment models, such as score-based systems derived from epidemiological studies, have been widely used to estimate cardiovascular risk. While these models provide a standardized approach, their

pISSN:2349-5707 Volume:13,Issue 1, January -2026 eISSN:2349-5715

predictive performance is often limited by assumptions of linearity and restricted variable selection. Consequently, they may inadequately represent individual patient risk, especially in diverse populations with complex comorbidities. Recent advancements in machine learning have introduced alternative approaches capable of handling high-dimensional and nonlinear clinical data. Studies have demonstrated that algorithms such as random forests, support vector machines, and gradient boosting methods can improve the accuracy of cardiovascular risk prediction by integrating demographic factors, clinical measurements, laboratory findings, and lifestyle indicators. Deep learning techniques have also been explored, particularly in the analysis of electrocardiograms and imaging data, showing enhanced performance in detecting subclinical cardiovascular abnormalities. The increasing availability of electronic health records has further accelerated research in this field, enabling longitudinal risk modeling and real-world validation of predictive algorithms. However, several studies highlight challenges related to data quality, model generalizability, and potential bias. Additionally, the limited interpretability of complex models has raised concerns among clinicians, emphasizing the need for explainable artificial intelligence approaches. Overall, existing literature supports the potential of machine learning for early cardiovascular disease prediction, while underscoring the importance of transparency, clinical relevance, and integration into primary care workflows.

Research methodology. This study adopts a quantitative, observational research design to evaluate the effectiveness of machine learning models in the early prediction of cardiovascular diseases within primary care settings. Retrospective clinical data were obtained from electronic health records routinely collected in primary healthcare facilities. The dataset included adult patients with no prior diagnosis of cardiovascular disease at baseline. Relevant features were selected based on clinical significance and data availability, including demographic characteristics, vital signs, lifestyle-related factors, laboratory test results, and comorbid conditions. Data preprocessing involved handling missing values, outlier detection, normalization of continuous variables, and categorical data encoding. The dataset was then divided into training and testing subsets to ensure unbiased model evaluation. Multiple machine learning algorithms were developed and compared, including logistic regression, random forest, and gradient boosting models. Hyperparameter tuning was performed using cross-validation to optimize model performance. Predictive accuracy was assessed using standard evaluation metrics such as accuracy, sensitivity, specificity, and the area under the receiver operating characteristic curve (AUC). To enhance clinical applicability, explainable artificial intelligence techniques were applied to identify key predictors and improve model transparency. Statistical analysis and model development were conducted using validated analytical tools, ensuring methodological rigor and reproducibility of the results.

Statistical analysis. Statistical analysis was conducted to describe the study population and to evaluate the performance of the developed machine learning models. Continuous variables were summarized using mean ± standard deviation or median with interquartile range, depending on data distribution, while categorical variables were presented as frequencies and percentages. Normality of continuous variables was assessed using the Shapiro-Wilk test. Comparisons between groups were performed using the independent t-test or Mann-Whitney U test for continuous variables and the chi-square test for categorical variables, as appropriate. Correlation analysis was applied to explore relationships between key clinical variables. A significance level of $p < 0,05$ was considered statistically significant. For model evaluation, predictive performance was assessed using accuracy, sensitivity, specificity, precision, F1-score, and the area under the receiver operating characteristic curve (AUC). Receiver operating characteristic (ROC) curves were constructed to compare model discrimination ability. Confidence intervals for AUC values were calculated using bootstrapping techniques. All statistical analyses and model evaluations were performed using standard statistical and machine learning software packages, ensuring robustness and reproducibility of the findings.

Conclusion. This study demonstrates that machine learning models have significant potential for the early prediction of cardiovascular diseases in primary care settings. By leveraging routinely collected clinical and demographic data, these models are able to capture complex patterns and interactions that are often overlooked by traditional risk assessment tools. The results indicate that machine learning-based approaches provide improved predictive accuracy and more effective identification of high-risk individuals at early stages of disease development. The integration of explainable artificial intelligence techniques enhances model transparency and supports clinician confidence in data-driven decision-making. Such models can assist primary care physicians in implementing timely preventive interventions, optimizing patient management, and reducing the overall burden of cardiovascular diseases. Despite these promising findings, further prospective validation and real-world implementation studies are required to assess long-term clinical impact and generalizability. Overall, machine learning–driven predictive tools represent a valuable component of future preventive cardiology and primary healthcare systems.

Recommendations. Based on the findings of this study, several recommendations can be made to enhance the application of machine learning for early prediction of cardiovascular diseases in primary care:

Integration into Clinical Workflows: Develop user-friendly interfaces for ML-based predictive tools that can be seamlessly incorporated into electronic health record systems to support routine primary care decision-making.

Focus on Explainable AI: Prioritize the use of interpretable and transparent machine learning models to ensure clinicians understand the basis of predictions, fostering trust and adoption.

Continuous Model Updating: Regularly update models with new patient data to maintain predictive accuracy and account for changes in population health trends.

Prospective Validation: Conduct prospective studies and real-world trials to validate model performance across diverse primary care populations and settings.

Patient-Centered Approaches: Incorporate patient lifestyle, behavioral, and social determinants of health into predictive models to improve individualized risk assessment.

Training and Education: Provide training for primary care clinicians on the use and interpretation of AI-driven tools, emphasizing ethical considerations, data privacy, and clinical relevance.

Policy and Guidelines Development: Collaborate with healthcare policymakers to establish guidelines for the safe and effective implementation of AI-based predictive models in primary care.

These recommendations aim to maximize the clinical utility of machine learning models, improve early detection of cardiovascular diseases, and support preventive healthcare strategies.

References.

1. Taylan O, Alkabaa A S, Alqabbaa H S, Pamukçu E, Leiva V. Early Prediction in Classification of Cardiovascular Diseases with Machine Learning, Neuro-Fuzzy and Statistical Methods. Biology. 2023;12(1):117. DOI:10.3390/biology12010117. This study compares multiple ML and neuro-fuzzy methods for early CVD prediction and demonstrates high classification performance of combined approaches.

2. Ogunpola A, Saeed F, Basurra S, Albarrak A M, Qasem S N. Machine Learning-Based Predictive Models for Detection of Cardiovascular Diseases. Diagnostics. 2024;14(2):144. Open Access. This work reviews recent ML models for detecting CVD using diverse clinical features and highlights performance metrics of modern approaches.

3. Machine learning based prediction models for cardiovascular disease risk using electronic health records data: systematic review and meta-analysis. PubMed. This systematic

pISSN:2349-5707 Volume:13,Issue 1, January -2026 eISSN:2349-5715
review examines ML models using EHR data and compares their predictive performance against traditional risk scores, showing superior AUC values in many ML models.

4. Early Prediction in Classification of Cardiovascular Diseases with Machine Learning, Neuro-Fuzzy and Statistical Methods – PubMed. This article proposes ML methods including support vector regression, M5Tree, and neural networks for CVD prediction and highlights their role in improving diagnostic accuracy.

5. Interpretable machine learning model for cardiovascular disease risk prediction: a feature decomposition-based study. BMC Public Health. 2025;25:3639. This research constructs and validates interpretable ML models for CVD risk, using SHAP for feature importance to support clinical interpretability.

6. Noori Mohammad Ali S, Ahmed N M. Comparing Some Machine Learning Models for Cardiovascular Disease. Journal of Pioneering Medical Sciences. 2025;14(4):60-66. This paper compares performance of several ML classifiers for CVD prediction, useful for understanding strengths and limitations across models.

7. Effective Heart Disease Prediction Using Machine Learning Techniques. This study evaluates several supervised ML algorithms and reports that models like k-nearest neighbor can achieve high predictive accuracy, illustrating ML utility in early detection.

8. Bazarbayev, M. I., Bozarov, U. A., Maxsudov, V. G., & Ermetov, E. Y. (2023). Application of differential equations in the field of medicine. International Journal of Engineering Mathematics (Online), 5(1).

9. Maxsudov, V. G., Bazarbayev, M. I., Ermetov, E. Y., & Norbutayeva, M. Q. (2020). Types of physical education and the technologies of organization of matters in the modern education system. European Journal of Research and Reflection in Educational Sciences Vol, 8(9).

10. Махсудов, В. Г. (2017). Гармоник тебранишларни инновацион технологиялар асосида ўрганиш («Кейс-стади»,«Ассесмент»,«Венн диаграммаси» мисолида). Современное образование (Узбекистан), (7), 11-16.

11. Maxsudov, V. G. (2018). Improvement of the methodological basics of training of the section «Mechanical oscillations» in higher educational institutions (Doctoral dissertation, Dissertation.–Tashkent: 2018. https://scholar. google. com/citations).

12. Zuparov, I. B., Ibragimova, M. N., Norbutayeva, M. K., Otaxonov, P. E., Normamatov, S. F., Safarov, U. Q., & Maxsudov, V. G. (2023). Modern directions and perspectives of using medical information systems. Switzerland: Innovations in technology and science education, 1218-1233.

13. Maxsudov, V. G., Ermetov, E. Y., & Jo, Z. R. rayeva. Types of physical education and the technologies of organization of matters in the modern education system. Fan, ta‘lim va amaliyot integratsiyasi 2022. Vol. 4. P29-34.

14. Maxsudov, V. G. (2018). Improvement of the methodological basics of training of the section «Mechanical oscillations» in higher educational institutions (Doctoral dissertation, Dissertation.–Tashkent).