



RADemics

Machine Learning–Based Predictive Analytics for Stroke and Cardiovascular Risk Assessment

Sachin S. Chourasia, Marngam Gameh,
Lecturer

DHOTE BANDHU SCIENCE COLLEGE, SIKKIM
MANIPAL INSTITUTE OF MEDICAL SCIENCES

Machine Learning–Based Predictive Analytics for Stroke and Cardiovascular Risk Assessment

¹Sachin S. Chourasia, Associate Professor, Department of Chemistry, Dhote Bandhu Science College, Gondia-441614. chousac1409@gmail.com

²Marngam Gameh, Lecturer, Department of Allied Health Professions, Sikkim Manipal Institute of Medical Sciences, Gangtok, Sikkim India. Marngam.g@smims.smu.edu.in

Abstract

Cardiovascular diseases and stroke remain leading causes of morbidity and mortality worldwide, necessitating accurate and early risk prediction to guide preventive interventions and optimize healthcare resources. Traditional risk assessment models, although widely adopted, are constrained by linear assumptions and limited feature sets, often failing to capture complex interactions among heterogeneous patient data. Machine learning–based predictive analytics offers a transformative approach by integrating multi-modal datasets, including electronic health records, imaging, physiological signals, and genomic information, to generate individualized risk predictions with high accuracy. This chapter provides a comprehensive overview of state-of-the-art machine learning algorithms, including ensemble methods, support vector machines, and deep learning architectures, emphasizing their comparative performance, interpretability, and clinical applicability. Strategies for effective multi-modal data integration, model validation, and explainability are critically discussed, highlighting methods to enhance trustworthiness and facilitate real-world clinical adoption. Challenges such as data heterogeneity, temporal dynamics, computational requirements, and ethical considerations are examined, alongside potential solutions for robust and scalable implementation. By synthesizing recent advancements and identifying research gaps, this chapter underscores the potential of machine learning frameworks to advance precision cardiology, enable proactive patient management, and reduce the burden of stroke and cardiovascular events.

Keywords: Machine Learning, Cardiovascular Risk, Stroke Prediction, Predictive Analytics, Multi-Modal Data Integration, Model Interpretability

Introduction

Cardiovascular diseases (CVDs) and stroke represent a major global health concern, accounting for millions of deaths and significant disability each year [1]. The growing prevalence of these conditions was driven by aging populations, lifestyle changes, and the increasing incidence of comorbidities such as hypertension, diabetes, and obesity [2]. Early detection of high-risk individuals was critical for the implementation of preventive interventions, timely medical management, and resource optimization within healthcare systems [3]. Traditional risk assessment tools, including the Framingham Risk Score and the ASCVD risk calculator, have been widely

employed to stratify patients based on demographic and clinical variables [4]. While these models provide foundational insights, they are limited by linear assumptions, restricted feature inclusion, and an inability to capture complex interactions between multifactorial risk determinants. Consequently, there was an urgent need for advanced predictive methods that leverage modern computational capabilities to improve accuracy and individualize risk assessment [5].

Machine learning (ML) has emerged as a powerful approach to enhance predictive analytics in cardiovascular medicine [6]. Unlike conventional statistical techniques, ML algorithms are capable of identifying subtle patterns, modeling non-linear relationships, and analyzing high-dimensional data from diverse sources [7]. Supervised learning models, such as random forests, support vector machines, and gradient boosting methods, have demonstrated superior predictive performance by effectively handling mixed-variable datasets [8]. Deep learning architectures, including convolutional and recurrent neural networks, offer additional capabilities for analyzing imaging and temporal physiological data [9]. These models can automatically learn feature representations, facilitating the discovery of latent associations that traditional approaches may overlook. The adaptability and scalability of ML techniques make them suitable for dynamic risk prediction, which was essential for monitoring disease progression and adjusting preventive strategies over time [10].

The integration of multi-modal datasets further enhances the predictive potential of ML models in cardiovascular risk assessment [11]. Clinical data from electronic health records, laboratory results, and medical imaging can be combined with lifestyle and behavioral information, wearable device outputs, and genomic profiles to create comprehensive patient risk profiles [12]. Multi-modal integration allows models to capture interdependencies across heterogeneous data sources, improving both sensitivity and specificity [13]. Challenges such as differences in data formats, temporal misalignment, missing values, and high dimensionality must be addressed to ensure robust and reliable predictions [14]. Feature selection, dimensionality reduction, and data harmonization techniques are essential to optimize model performance while maintaining clinical relevance and interpretability [15].