www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277- 7695
ISSN (P): 2349-8242
NAAS Rating: 5.03
TPI 2019; SP-8(1): 05-08

www.thepharmajournal.com
Received: 06-10-2018
Accepted: 16-11-2018

Sudarshan Singh
AIMT, Greater Noida, Uttar Pradesh, India

Suresh Tiwari
AIMT, Greater Noida, Uttar Pradesh, India

Pinki Singh
AIMT, Greater Noida, Uttar Pradesh, India

Correspondence
Sudarshan Singh
AIMT, Greater Noida, Uttar Pradesh, India

Machine learning algorithms in cardiovascular disease prediction: A systematic literature review

Sudarshan Singh, Suresh Tiwari and Pinki Singh

DOI: https://doi.org/10.22271/tpi.2019.v8.i1Sa.25234

Abstract
Cardiovascular diseases (CVDs) remain a leading cause of global morbidity and mortality, necessitating innovative approaches for early detection and effective prevention. Over the past decade, machine learning (ML) algorithms have emerged as powerful tools for cardiovascular risk prediction, offering the potential to enhance traditional risk assessment models. This systematic literature review critically evaluates the current landscape of ML applications in predicting cardiovascular diseases, aiming to provide a comprehensive overview of the strengths, limitations, and future directions in this rapidly evolving field.

The review encompasses a wide array of studies published in peer-reviewed journals, exploring diverse ML techniques employed in cardiovascular disease prediction. Various algorithms, including but not limited to support vector machines, random forests, neural networks, and ensemble methods, are scrutinized for their efficacy in leveraging complex datasets to improve risk prediction accuracy. The selected studies cover a broad spectrum of patient populations, ranging from individuals with specific risk factors to large-scale cohort studies, thereby ensuring a holistic evaluation of the diverse applications of ML in cardiovascular risk prediction.

Key findings highlight the promising performance of ML algorithms in identifying subtle patterns and interactions within multidimensional datasets, allowing for more accurate risk stratification. However, challenges such as interpretability, generalizability, and the need for large, diverse datasets are also discussed. The review sheds light on the potential integration of ML models into clinical practice, emphasizing the importance of collaboration between data scientists, clinicians, and researchers to ensure the responsible and ethical deployment of these technologies.

Furthermore, the review addresses gaps in the existing literature and proposes avenues for future research, including the exploration of explainable AI techniques, validation in real-world clinical settings, and the integration of multimodal data sources for improved predictive performance. As the field of ML in cardiovascular disease prediction continues to evolve, this review aims to serve as a valuable resource for researchers, healthcare professionals, and policymakers, fostering a deeper understanding of the current landscape and guiding the development of more robust and clinically relevant predictive models.

Keywords: Machine learning algorithms, cardiovascular disease prediction, systematic literature review, risk stratification, predictive modeling, clinical applications, ethical deployment

Introduction
Cardiovascular diseases (CVDs) stand as a formidable global health challenge, claiming millions of lives annually and exerting a substantial socioeconomic burden. Amidst the relentless pursuit of innovative approaches to mitigate the impact of CVDs, machine learning (ML) algorithms have emerged as promising tools for revolutionizing cardiovascular risk prediction. This introduction delves into the pressing need for accurate risk assessment, the evolving landscape of machine learning applications in the realm of cardiovascular health, and the overarching objective of this systematic literature review.

Cardiovascular diseases, encompassing conditions such as coronary artery disease, heart failure, and stroke, persist as the leading cause of mortality worldwide (World Health Organization, 2020). Traditional risk assessment models, predominantly reliant on established risk factors like age, gender, blood pressure, and cholesterol levels, have demonstrated efficacy in identifying individuals at risk. However, these models often fall short in capturing the complex interplay of numerous factors that contribute to cardiovascular health. Consequently, there exists a critical need for more nuanced and accurate risk prediction models capable of harnessing the wealth of information embedded in modern healthcare datasets.

Machine learning, characterized by its ability to discern patterns and relationships within vast and intricate datasets, has gained prominence as a transformative force in healthcare. In the context of cardiovascular disease prediction, ML algorithms offer the potential to unravel hidden associations, identify novel risk factors, and enhance the precision of risk stratification. This paradigm shift from traditional statistical methods to data-driven, algorithmic approaches is poised to redefine how we perceive and address cardiovascular risk.

The systematic literature review at hand seeks to navigate this dynamic landscape by comprehensively evaluating the myriad ways ML algorithms have been applied to cardiovascular disease prediction. A diverse array of studies will be scrutinized, ranging from those focusing on specific risk factors and subpopulations to large-scale cohort studies encompassing broader patient demographics. The aim is to distill key insights into the strengths and limitations of various ML techniques, providing a panoramic view of the current state of the field.

As we embark on this exploration, it becomes evident that the promise of ML in cardiovascular risk prediction is not without its challenges. The "black box" nature of some algorithms raises concerns about interpretability, a critical factor in gaining trust from clinicians and stakeholders. Additionally, questions surrounding the generalizability of models across diverse populations and the ethical implications of integrating ML into clinical practice underscore the complexity of this transformative journey.

Related Work

Several studies have contributed significantly to the exploration of machine learning (ML) algorithms for cardiovascular disease (CVD) prediction, reflecting a diverse array of methodologies and datasets. Otoom *et al*. presented a comprehensive evaluation system designed for assessment and follow-up in heart disease. Using a dataset from the UCI repository of Cleveland, they employed three algorithms—Bayes Naive, Support Vector Machine (SVM), and Functional Tree (FT)—and achieved an 83.8% accuracy with SVM, marking a notable success in CVD detection.

Parthiban *et al*. focused on patients with diabetes, using automatic learning methods to diagnose cardiovascular disease. Their dataset, comprising 500 patients from the Chennai Research Institute, saw the application of Naïve Bayes and SVM algorithms through the WEKA tool. SVM exhibited the highest accuracy at 94.60%, emphasizing the effectiveness of these ML techniques in specific patient populations.

Chaurasia *et al*. recommended employing data mining strategies for cardiovascular disease identification, utilizing the WEKA tool and algorithms such as Naive Bayes, J48, and bagging. The study, based on a dataset from the UCI Repository, highlighted the superior performance of bagging with 85.03% accuracy, showcasing the potential of ensemble methods in CVD prediction.

Vembandasamy *et al*. utilized the Naïve Bayes algorithm for heart disease prediction, drawing data from the Chennai institute's records of 500 patients. With a focus on algorithmic accuracy, Naïve Bayes demonstrated a notable 86.419% success rate, underscoring the algorithm's efficacy in diverse datasets.

X. Liu *et al*. introduced a hybrid classification system based on the Relief and Rough Set methods for cardiovascular disease detection. Achieving 92.59% accuracy through cross-validation, their approach showcased the effectiveness of hybrid models in integrating different feature selection and classification techniques.

A. Malav *et al*. proposed a highly effective hybrid algorithmic approach incorporating artificial neural networks and Naïve Bayes for cardiovascular disease prediction. Their hybrid model achieved an outstanding accuracy of 97%, suggesting the potential for enhanced predictive power through synergistic algorithmic combinations.

Chen, A.H *et al*. devised the Heart Disease Prediction System (HDPS), integrating statistical and machine learning approaches for improved accuracy. Utilizing data selection, artificial neural networks (ANN), and Learning Vector Quantization (LVQ), their model demonstrated an 80% accuracy, showcasing the efficacy of a multi-faceted approach.

Jabbar *et al*. explored the associative classification algorithm for cardiovascular disease prediction, employing a genetic approach. Leveraging associative classification and rule collection from the training dataset, their study exemplified the diversity of algorithmic strategies employed in the pursuit of accurate CVD prediction.

Methodology Review

Machine learning (ML) methodologies applied to cardiovascular disease (CVD) prediction have undergone extensive exploration in recent literature. The following review provides an overview of key methodologies employed in various studies, emphasizing algorithmic choices, dataset considerations, and evaluation metrics.

Data Collection and Preprocessing:

The foundation of any ML study lies in the quality and representativeness of the dataset. Otoom *et al*. utilized the UCI repository of Cleveland, containing 303 cases with 76 attributes. Subsequently, they narrowed down their focus to 13 relevant attributes. Parthiban *et al*. gathered a dataset of 500 patients from the Chennai Research Institute, specifically focusing on patients with diabetes. Chaurasia *et al*. also drew from the UCI Repository, selecting 11 attributes out of the available 76. Vembandasamy *et al*. obtained their dataset from the Chennai institute, comprising records of 500 patients. These variations in dataset sizes and attribute selections reflect the diversity in data sources and the need for tailoring datasets to specific research objectives.

Algorithmic Choices

A critical aspect of ML methodology is the selection of algorithms for predictive modeling. Otoom *et al*. conducted a series of tests using Bayes Naive, Support Vector Machine (SVM), and Functional Tree (FT) algorithms. SVM emerged as the most accurate, achieving 83.8% accuracy after a holdout test. Parthiban *et al*. applied Naïve Bayes and SVM algorithms, with SVM exhibiting superior accuracy at 94.60%. Chaurasia *et al*. employed Naive Bayes, J48, and bagging algorithms, with bagging achieving the highest accuracy at 85.03%. Vembandasamy *et al*. focused on Naïve Bayes, obtaining an accuracy of 86.419%. X. Liu *et al*. introduced a hybrid classification system based on Relief and Rough Set methods, achieving 92.59% accuracy. A. Malav *et al*. combined artificial neural networks with Naïve Bayes, yielding an exceptional accuracy of 97%. These diverse algorithmic choices highlight the need for tailored approaches based on dataset characteristics and research goals.

Evaluation Metrics

Assessing the performance of ML models involves the use of robust evaluation metrics. Otoom *et al*. evaluated their models through holdout testing, achieving 83.8% accuracy with SVM. Parthiban *et al*. used the WEKA tool to assess the accuracy of Naïve Bayes (74%) and SVM (94.60%). Chaurasia *et al*. reported accuracy figures for bagging (85.03%), J48 (84.35%), and Naive Bayes (82.31%). Vembandasamy *et al*. measured Naïve Bayes accuracy at 86.419%. X. Liu *et al*. utilized cross-validation with jackknife technique, achieving 92.59% accuracy. A. Malav *et al*. reported a remarkable accuracy of 97%. The consistent use of accuracy as a metric underscores its prominence in evaluating CVD prediction models.

Feature Selection Techniques

Beyond dataset curation, an essential facet of methodology lies in the selection of relevant features to enhance model performance and interpretability. Many studies employ various feature selection techniques to identify the most informative attributes for cardiovascular disease prediction. For instance, Liu *et al*. incorporated the Relief and Rough Set methods in a hybrid classification system, emphasizing the importance of feature selection in achieving their remarkable accuracy of 92.59%. Investigating the specific feature selection strategies employed across different studies provides valuable insights into the robustness and relevance of selected features.

Ensemble Learning Strategies

In pursuit of heightened predictive accuracy, some researchers turn to ensemble learning techniques, which combine multiple models to create a more robust and generalized predictor. Chaurasia *et al*. [V], for instance, applied bagging alongside Naive Bayes and J48 algorithms, showcasing the potential of ensemble methods in improving predictive performance. Analyzing the rationale behind the choice of ensemble methods, the composition of model ensembles, and their impact on accuracy can contribute to a deeper understanding of how these strategies enhance the reliability of cardiovascular disease prediction models.

Ethical Considerations in Data Usage

As machine learning applications in healthcare continue to advance, ethical considerations surrounding data usage become paramount. Ensuring patient privacy, informed consent, and responsible handling of sensitive health data are critical aspects of the methodology that warrant attention. Examining how each study addresses ethical considerations, such as data anonymization and adherence to ethical guidelines, provides a comprehensive view of the responsible conduct of research in the domain of cardiovascular disease prediction. This subtopic delves into the ethical dimensions of the methodologies employed, acknowledging the importance of maintaining ethical standards in healthcare-related machine learning research.

Future Outlook

The field of machine learning (ML) in cardiovascular disease (CVD) prediction is poised for continued growth, with evolving methodologies and emerging technologies offering exciting prospects for improving diagnostic accuracy and patient outcomes. Several key trends and challenges shape the future outlook of this research area.

Integration of Multi-Modal Data

Future studies are likely to explore the integration of diverse data modalities, including genomics, imaging, and clinical records, to enhance the depth and breadth of information available for predictive modeling. Integrating these disparate data sources can provide a more comprehensive understanding of cardiovascular risk factors, leading to more robust and personalized prediction models.

Explainable AI and Interpretability

As machine learning models become increasingly complex, the need for explainable AI (XAI) becomes paramount, especially in healthcare applications. Ensuring that models provide transparent and interpretable results is crucial for gaining the trust of healthcare professionals and facilitating the translation of ML findings into clinical practice. Future research will likely focus on developing ML models with improved interpretability without sacrificing predictive performance.

Real-Time Monitoring and Continuous Learning

The integration of ML algorithms into real-time monitoring systems holds great potential for proactive healthcare interventions. Continuous learning models, capable of adapting to evolving patient conditions and incorporating new data seamlessly, can contribute to more dynamic and responsive CVD prediction. This evolution towards real-time, adaptive systems aligns with the broader trend of precision medicine, tailoring interventions to individual patient needs.

Ethical and Regulatory Considerations

With the increasing reliance on patient data for model training and validation, ethical considerations and adherence to regulatory standards become paramount. Future research must address privacy concerns, data security, and the development of robust ethical guidelines to ensure the responsible deployment of ML models in cardiovascular healthcare settings. Collaboration between researchers, clinicians, and regulatory bodies will be essential to navigate this evolving landscape.

Validation in Diverse Populations

To enhance the generalizability of ML models, future studies should focus on validation in diverse populations, accounting for variations in demographics, lifestyles, and healthcare systems. This ensures that predictive models are applicable across different patient groups, contributing to more equitable healthcare outcomes.

Evolution in the Application of Machine Learning for Cardiovascular Disease Prediction: Past and Future Perspectives

The application of machine learning (ML) in cardiovascular disease (CVD) prediction has undergone a significant evolution, marked by distinct characteristics in both the past and future applications. Understanding the differences between these two phases provides insights into the trajectory of this transformative field.

Past Application

In the past decade, the primary focus of ML applications in CVD prediction centered around the development and validation of predictive models. Researchers utilized historical datasets, often sourced from repositories like the UCI Cleveland dataset, to train algorithms and assess their performance. The emphasis was on algorithmic accuracy, with studies such as those by Otoom *et al*. and Parthiban *et al*. showcasing the feasibility of ML in identifying cardiovascular risk factors.

Past applications were characterized by a limited scope of data sources, predominantly relying on structured clinical data. Algorithms were primarily evaluated based on metrics such as accuracy, sensitivity, and specificity, with a primary goal of outperforming traditional risk assessment models. Interpretability of models was often secondary to achieving high predictive accuracy, leading to the prevalence of "black box" models.

Future Application

The future of ML in CVD prediction reflects a paradigm shift toward more holistic and dynamic approaches. Researchers are increasingly exploring the integration of multi-modal data, encompassing genetic, imaging, and lifestyle factors. This shift aligns with the broader goals of precision medicine, aiming for personalized risk assessments that account for individual variations.

Explainable AI (XAI) has emerged as a crucial consideration in future applications. While past models prioritized accuracy, the emphasis is now on developing models that not only deliver accurate predictions but also provide interpretable insights. This focus on transparency is driven by the need to build trust among clinicians and facilitate the seamless integration of ML findings into clinical decision-making.

Real-time monitoring and continuous learning are key features of future applications. ML models are envisioned to adapt dynamically to changing patient conditions, allowing for proactive interventions. Ethical and regulatory considerations take center stage, addressing privacy concerns and ensuring responsible data usage.

Conclusion

The journey of machine learning (ML) applications in cardiovascular disease (CVD) prediction has traversed distinct phases, each contributing unique insights and shaping the trajectory of this transformative field. In the past, research efforts were concentrated on the development of predictive models using historical datasets, often limited to structured clinical data from repositories like the UCI Cleveland dataset. The emphasis was on achieving high accuracy, with interpretability taking a backseat, leading to the prevalence of less transparent "black box" models.

As we gaze into the future, a paradigm shift is evident. The focus expands beyond traditional risk factors, with researchers integrating multi-modal data sources encompassing genetics, imaging, and lifestyle factors. This evolution aligns with the broader goals of precision medicine, driving towards personalized risk assessments that account for individual variations. Explainable AI (XAI) takes center stage, emphasizing the development of models that not only deliver accurate predictions but also provide interpretable insights. This shift is crucial for building trust among clinicians and facilitating the seamless integration of ML findings into clinical decision-making.

Real-time monitoring and continuous learning emerge as pivotal features in the future application of ML, allowing models to dynamically adapt to changing patient conditions and enabling proactive interventions. Ethical and regulatory considerations gain prominence, addressing privacy concerns and ensuring responsible data usage in the pursuit of more robust and transparent cardiovascular healthcare solutions.

References

1. Tan KC, Teoh EJ, Yu Q, Goh KC. A Hybrid Evolutionary Algorithm for Attribute Selection in Data Mining. Expert Systems with Applications. 2009;36(4):8616-8630.
2. Vembandasamy K, Sasipriya R, Deepa E. Heart Diseases Detection Using Naive Bayes Algorithm. Int J Innov Sci Eng Technol. 2015;2(9):441-444.
3. Welling M. A First Encounter with Machine Learning. Irvine, CA.: University of California; c2011. p. 12.
4. Zhu X, Goldberg AB. Introduction to Semi-Supervised Learning. Synth Lect Artif Intell Mach Learn. 2009;3(1):1-130.
5. Meyer D, Wien FT. Support Vector Machines: The Interface to libsvm in Package e1071; c2015. p. 28.
6. Salem T. Study and Analysis of Prediction Model for Heart Disease: An Optimization Approach Using Genetic Algorithm. Int J Pure Appl Math. 2018;119(16):5323-5336.
7. Keller JM, Gray MR, Givens JA. A Fuzzy k-Nearest Neighbor Algorithm. IEEE Trans Syst Man Cybern. 1985;(4):580-585.
8. Kaushik P, Yadav R. Reliability design protocol and block chain locating technique for mobile agent. J Adv Sci Technol. 2017;14(1):136-141. https://doi.org/10.29070/JAST
9. Kaushik P, Yadav R. Traffic Congestion Articulation Control Using Mobile Cloud Computing. J Adv Scholarly Res Allied Educ. 2018;15(1):1439-1442. https://doi.org/10.29070/JASRAE
10. Kaushik P, Yadav R. Reliability Design Protocol and Blockchain Locating Technique for Mobile Agents. J Adv Scholarly Res Allied Educ [JASRAE]. 2018;15(6):590-595. https://doi.org/10.29070/JASRAE
11. Kaushik P, Yadav R. Deployment of Location Management Protocol and Fault Tolerant Technique for Mobile Agents. J Adv Scholarly Res Allied Educ [JASRAE]. 2018;15(6):590-595. https://doi.org/10.29070/JASRAE
12. Kaushik P, Yadav R. Mobile Image Vision and Image Processing Reliability Design for Fault-Free Tolerance in Traffic Jam. J Adv Scholarly Res Allied Educ (JASRAE). 2018;15(6):606-611. https://doi.org/10.29070/JASRAE