Artificial Intelligence Applications And Innovations Vol. 1 No. 3 (2024)

Article history:
Received 2 March 2024
Revised 27 April 2024
Accepted 01 June 2024
Published online 01 July 2024

A Comparative Machine Learning Analysis for Early Atrial Fibrillation Prediction

Badrosadat. Nategholeslam Shirazi[1], Shiva. Naghsh[1], Ali Akbar. Safavi[1*], Amir. Sharafkhaneh[2]

[1] School of Electrical and Computer Engineering, Shiraz University, Shiraz, Iran
[2] Department of Medicine, Baylor College of Medicine, Houston, TX, USA

*** Corresponding author email address**: safavi@shirazu.ac.ir

Article Info

Article type:
Original Research

How to cite this article:
Nategholeslam Shirazi, B., Naghsh, S., Safavi, A. A., & Sharafkhaneh, A. (2024). A Comparative Machine Learning Analysis for Early Atrial Fibrillation Prediction. *Artificial Intelligence Applications and Innovations*, *1*(3), 1-8.
https://doi.org/10.61838/jaiai.1.3.1

ABSTRACT

Cardiovascular Diseases (CVD) are significant global cause of mortality. This paper focuses on early detection of a specific type of CVD, Atrial Fibrillation (AF), through a simple approach. The methodology is based on efficient risk assessment methods to identify high-risk individuals with a comparative analysis of seven ML algorithms to find the simplest and most effective approach. The research utilizes the Sleep Heart Health Study (SHHS) dataset, a large-scale cohort study with diverse clinical parameters and polysomnographic data, which seems to be ideal for early AF prediction. The study formulates a predictive analysis based on minimal accessible data (i.e. no signal, image, or complex measurement are considered) and evaluates seven ML algorithms including Support Vector Machines (SVM), K-Nearest Neighbor (KNN), Random Forest (RF), Decision Tree (DT), Gradient Boosting (GB), Multi-Layer Perceptron (MLP), and Logistic Regression (LR). Among these methods, LR shows notable predictive accuracy. The analysis covers a diverse cohort, including various races (i.e. White, Black, and others), ethnicities, and both genders, with a focus on individuals with aged averagely more than 63. The study concludes that our formulation with the simple and readily accessible parameters predict AF reasonably well, potentially enabling early interventions to reduce morbidity and mortality.

Keywords: *Atrial fibrillation, Cardiovascular diseases, Early prediction, Machine learning methods*

1. Introduction

Cardiovascular diseases (CVD) encompass heart and blood vessel disorders, leading to severe outcomes like heart attacks and failure [1-5]. As the primary cause of global mortality (31% of deaths) [1, 2], early identification and management of high blood pressure and diabetes significantly reduce the risk of conditions such as coronary artery disease and atrial fibrillation [1, 4, 5]. Prompt diagnosis not only lessens hospitalizations and emergency visits but also yields substantial economic benefits [1-3]. Various types of CVD include cerebrovascular disease, peripheral arterial disease, AF, and aortic disease [1-5]. Often asymptomatic for years, these conditions might go undetected until later stages, highlighting the importance of identifying high-risk individuals for timely intervention to prevent premature fatalities [2, 4]. Besides, from the point of views of practicality and economic aspects, it is also quite important to see how we can develop prediction models while using only simple and mostly available data from the individuals.

ML techniques have demonstrated efficacy in diagnosing and forecasting CVD [6, 7]. By processing extensive clinical and health-related datasets, these methods construct predictive models, effectively pinpointing individuals with elevated CVD risk. They play a pivotal role across multiple stages of CVD diagnosis and treatment planning, encompassing risk evaluation and therapeutic strategies. Recent research has evaluated various ML methods, highlighting their proficiency in predicting CVD risks. Findings underscored the accuracy of ML algorithms in this domain, indicating potential enhancements in patient outcomes [8].

Recently, predicting CVD using ML becomes a hotspot among researchers [9, 10]. Huang et al. [11] in their study utilized Coronary artery Computed Tomography Angiography (CCTA) to compute Coronary Artery Calcification Scores (CACS) and combined them with clinical factors to predict CHD using ML. The dataset comprised patients suspected of CHD who underwent CCTA at a hospital between Jan 2019 and Mar 2022. The Agatston Score quantified calcification, and 31 clinical variables were collected, including hypertension, diabetes, and smoking. Faizal et al. [12] review CVD risk prediction models, comparing traditional statistical methods such as GRACE, TIMI, and the Framingham Risk Score with the emerging use of AI approaches, primarily machine learning. The review discusses the strengths and limitations of both methodologies, emphasizing the necessity for biomarkers in risk assessment and early detection. By highlighting global CVD mortality rates and the challenges faced, the paper emphasizes the transition towards integrated big data analytics for future CVD risk assessment. Jafari et al. [13] highlights the increasing global impact of CVDs as a leading cause of mortality. The study emphasizes the challenge of early detection due to the subtle initial symptoms of CVDs and the complexities involved in interpreting cardiac magnetic resonance imaging (CMRI) data. To address this challenge, the research explores the application of deep learning (DL) techniques in diagnosing CVDs using CMR images. It reviews existing research on CVD detection through CMR images and DL methods, discussing the intricacies of interpreting CMRI data. Ebrahimzadeh et al. [14] focus on predicting Paroxysmal Atrial Fibrillation (PAF) by employing various analysis methods applied to Heart Rate Variability (HRV) signals. Their study aims to create a validated predictive model for PAF onset by integrating classical and modern techniques. The dataset used comprises 106 signals derived from 53 pairs of ECG recordings obtained from the Atrial Fibrillation Prediction Database (AFPDB). They extract diverse features from linear, time-frequency, and nonlinear analyses of HRV, subsequently combining and reducing them using a local feature selection approach. Ghosh et al. [15] employed ML techniques with Relief and LASSO feature selection on a combined dataset (Cleveland, Long Beach VA, Switzerland, Hungarian, and Stat log) to proficiently predict CVD. They underscored the significance of meticulous data collection, preprocessing, and transformation to enhance model accuracy. Shah, Patel, and Bharti [16] utilized the UCI Machine Learning Repository dataset, finding optimal accuracy with the random forest technique. Jindal et al. [17], using the UCI repository dataset, highlighted ML's superior predictive abilities for heart disease, emphasizing the importance of early detection and prevention. Katarya & Srinivas [18] surveyed recent ML studies focusing on early-stage heart disease prediction, drawing insights from expansive healthcare industry data. Yadav et al. [19] evaluated ML algorithms on patient data from the Cardiovascular Disease Dataset by Svetlana Ulianova, emphasizing improved early detection and intervention.

All the above-mentioned literatures use the data collected from awake the individuals. There are some datasets related to the data collected based on sleep studies. Among these datasets, the SHHS dataset considered in this study not only is based on data collected from sleep study but it is quite focused on CVD. There are only a few studies who have used this dataset to CVD studies. Park et al. [20] focuses on creating a predictive model for CVD in individuals with SDB using data from the SHHS. Their study divided the extracted electrocardiography (ECG) features, including 18 from signal processing methods and 30 using AI, alongside ten clinical CVD risk factors. Employing SVM model, they predicted CVD outcomes. On the other hand, Zhang et al. [21] investigated the association between sleep heart rate variability (HRV) and long-term CVD outcomes using data from SHHS. They found reduced HRV during sleep in individuals who later experienced CVD events compared to those who remained CVD-free. Specifically, the high-frequency component of HRV emerged as an independent predictor of CVD outcomes and displayed a positive correlation with the latency of CVD. They developed a predictive model combining HRV features with conventional CVD risk factors. Zhang et al. [22] developed a predictive model using a Bi-directional Long Short-Term Memory (Bi-LSTM) framework with an Attention layer to anticipate angina pectoris events. The research utilized data from the SHHS and specifically focused on evaluating the potential of resting-state RR interval time series, derived from electrocardiogram (ECG) signals, as predictive markers. Their primary objective was to forecast instances of angina pectoris based on the RR interval time series data. These findings suggest the potential usefulness of RR interval data in identifying individuals at risk of angina pectoris. In our previous study [23], we utilized the SHHS dataset to develop a predictive model for coronary heart disease (CHD) risk, demonstrating the effectiveness of readily available clinical parameters and advanced machine learning techniques. Building on this foundation, the current study shifts focus to atrial fibrillation (AF) prediction, further emphasizing the potential of accessible data for enhancing early diagnosis and improving patient outcomes.

To the best of our knowledge, limited comprehensive comparative studies exist regarding the prediction of AF using accessible parameters for physicians. The optimal minimal set of parameters (i.e. from the simplicity and accessibility point of views) for such predictions remains unidentified. This study explores diverse parameter selection and problem formulation strategies, applying various ML methods to develop a simple yet highly predictive model for incident AF.

2. Materials and Methods

2.1. Data

The current study utilized prospectively collected data from the Sleep Heart Health Study (SHHS) dataset, a large-scale, multi-center cohort designed to examine the relationship between sleep-disordered breathing (SDB) and cardiovascular disease (CVD) in a community-based sample of adults in the United States. This dataset includes a comprehensive collection of clinical, demographic, and polysomnographic data, as well as follow-up data on CVD events and mortality [24, 25].

A total of 6,441 participants aged 40 years and older, recruited from nine clinical centers across the United States, underwent in-home polysomnography to measure sleep variables such as sleep apnea, hypopnea, and oxygen saturation. Additional data included clinical and demographic assessments, physical examinations, laboratory tests, and questionnaires on medical history, lifestyle habits, and sleep quality.

The study focused on a subset of simple and highly accessible parameters from three categories within the SHHS dataset: Demographics, Measurements, and Medical History (see Table S2 in the attachment). Only subjects without missing values for these selected parameters were included.

To train and evaluate machine learning models, the dataset was partitioned using an 80-20 split, allocating 80% for training and 20% for testing. This method ensures robust evaluation of model performance on previously unseen data.

2.2. Methodology

The study categorized available data to identify simple and accessible parameters and formulated a prediction problem based on these data. It then assessed the effects of individual and combined parameter categories on the performance of various machine learning (ML) models.

Several ML algorithms, including SVM, KNN, RF, DT, GB, MLP, and LR, were evaluated for distinguishing healthy individuals from those with incident AF during follow-up. Figure 1 illustrates the framework of the ML prediction models developed.

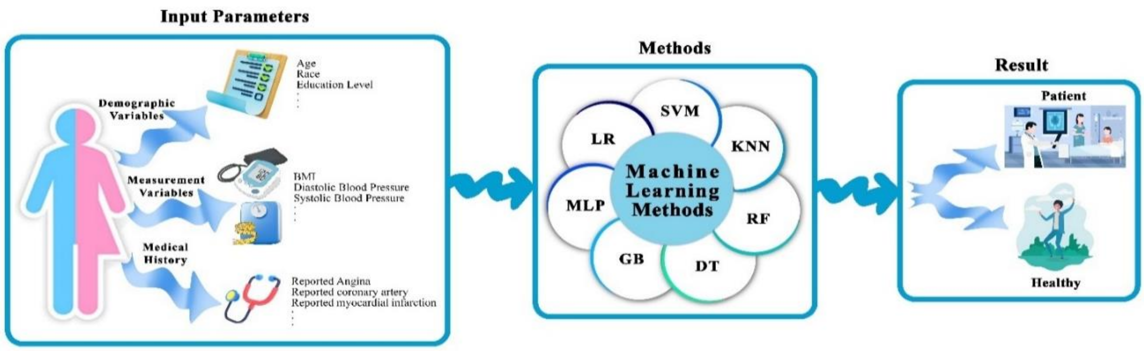

Figure 1. An overview of the prediction models

Table 1 presents a short description of each Machine Learning algorithm used in this paper with some simple comparisons of their differences.

Table 1. Comparative Table of Machine Learning Algorithms in Medical Applications [26]

Algorithm	Description	Advantages	Challenges	Medical Applications
Support Vector Machines (SVM)	Robust binary classifier identifying decision boundaries in high-dimensional data.	Effectively handles high-dimensional data; applied in medical imaging and disease diagnosis.	May not perform well with large datasets or overlapping classes.	Image analysis, disease diagnosis.
K-Nearest Neighbor (KNN)	Intuitive algorithm for classification and regression, classifies based on k-nearest neighbors.	Simple and interpretable; challenges with high-dimensional data.	High computational complexity; sensitive to irrelevant features.	Classification, regression, disease diagnosis.
Random Forest (RF)	Versatile ensemble method using multiple decision trees to mitigate overfitting.	Robustness, adaptability; suitable for classification and regression.	Less interpretable than individual trees; parameter tuning.	Medical research, disease diagnosis, classification.
Decision Trees (DT)	Creates interpretable models by recursively splitting data based on selected features.	Transparency; tendency to overfit, especially with deep trees.	Overfitting with deep trees; may not capture complex relationships.	Early-stage diagnostics, risk assessment.
Gradient Boosting (GB)	Powerful technique combining weak learners (usually decision trees) iteratively for high accuracy.	Handles complex relationships; resilience to overfitting.	Computationally intensive; sensitive to noisy data.	Medical diagnostics, risk assessment, predictive modeling.
Multi-Layer Perceptron (MLP)	Fundamental neural network architecture effective for complex nonlinear problems.	Handles complex data; effective for medical diagnosis.	Requires substantial data; computationally intensive.	Medical diagnosis, outcome prediction, anomaly detection.
Logistic Regression (LR)	Simple yet effective method for binary classification with interpretability.	Interpretability; efficiency in binary classification.	Limited to linear relationships; can't handle complex patterns.	Outcome prediction, risk assessment.

In order to address the issue of missing values and high correlation among some medical history parameters, all parameters related to a specific disease origin were combined. For instance, as four parameters related to having history of myocardial infarction (MI), myocardial infarction procedure (MIP), angina and revascularization procedure referred to a same origin, we combined them as a single parameter. This new parameter has a value of zero (no previous history) and one (history of any of the above three disease).

To present some evaluation criteria we have applied several measures. Accuracy measures correctness in predictive models, representing the ratio of correct predictions to total predictions. Area Under the Curve (AUC) assesses a classification model's performance by quantifying the area under the Receiver Operating Characteristic (ROC) curve, which indicates the model's ability to distinguish between classes. Higher AUC values indicate better model discrimination. Besides, the confusion matrix is also tabular representation of a classification model's performance, comparing its predicted outcomes with actual outcomes. Typically, the matrix comprises four cells of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) which indicate the number of positive and negative outcomes that were correctly and incorrectly predicted by the model.

True positives represent the cases where the model correctly predicted a positive outcome, while false positives represent the cases where the model incorrectly predicted a positive outcome. True negatives are the cases where the model correctly predicted a negative outcome, while false negatives are the cases where the model incorrectly predicted a negative outcome.

3. Results

This study focused on differentiating patients with AF from healthy individuals, utilizing seven tests that examined various combinations of parameter categories. Table 2 outlines the tests conducted. This information is valuable for developing effective diagnostic strategies.

Table 2. Conducted tests for each group of parameters

Tests	Used Parameters
Test 1	Only Demographic parameters considered
Test 2	Only Limited Measurement parameters considered
Test 3	Only Medical parameters considered
Test 4	Only Demographic & Limited Measurement parameters considered
Test 5	Only Demographic & Medical parameters considered
Test 6	Only Limited Measurement & Medical parameters considered
Test 7	All Demographic, Limited Measurement & Medical parameters considered

3.1. Distinguishing Between AF Patients and Healthy Individuals

The objective of this study is to distinguish healthy individuals and those with AF. The study investigated incident AF by selecting a sample of participants from the SHHS dataset. The sample comprised 262 healthy women and 211 healthy men with an average age of 64 years, as well as 168 women and 159 men with an average age of 74 years who had the disease (see Table S3, in attachment). The findings in Table 3 presents a comparative summary of ML algorithm performances.

Results of the study, which were presented in Table 3, indicate that the LR algorithm is the most effective classification method for predicting AF. The LR algorithm emerged as the most effective classifier, achieving an average accuracy of 69.20%, peaking at 75% accuracy with an AUC of 79.69%. Furthermore, Figure 2 provides the LR confusion matrix.

Table 3. Diagnosis of AF patients and healthy individuals using various ML techniques and investigating the impact of three types of variables of Demographics, Measurements, and Medical History (Sorted based on best result of Test 7)

Method	Accuracy of Test 1 (%)		Accuracy of Test 2 (%)		Accuracy of Test 3 (%)		Accuracy of Test 4 (%)		Accuracy of Test 5 (%)		Accuracy of Test 6 (%)		Accuracy of Test 7 (%)	
	P (327)	H (473)	P (327)	H (473)	P (327)	H (473)	P (327)	H (473)	P (327)	H (473)	P (327)	H (473)	P (327)	H (473)
LR	63.12		68.13		70.00		68.13		67.50		72.50		75.00	
DT	63.12		60.62		66.25		69.37		66.87		70.63		72.50	
GB	66.87		63.75		63.12		66.87		66.25		68.75		71.25	
RF	66.87		63.12		58.75		71.25		65.62		71.88		70.00	
SVM	64.38		66.25		60.62		65.62		60.62		65.62		67.50	
MLP	66.25		63.75		64.38		63.75		66.87		69.37		66.87	
KNN	65.00		66.87		60.62		65.62		63.12		63.12		63.12	

- P: Number of Patient Individuals
- H: Number of Healthy Individuals

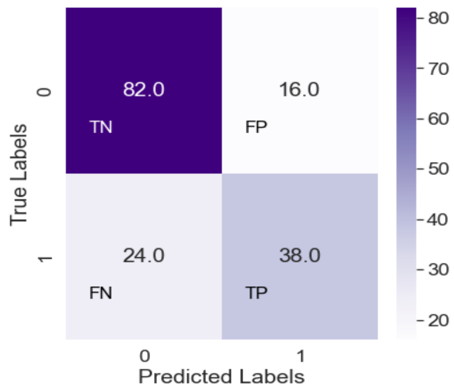

Figure 2. Confusion matrix of LR method for AF disease

In this study, the LR method demonstrated as the most accurate predictor for AF through comprehensive analyses. Detailed examination in Table 4 illustrates that the combination of Demographic, Limited Measurement, and Medical parameters (Test 7) yielded the highest accuracy in predicting AF. Excluding Demographics led to a minor decrease in accuracy (~2.5%), while using only Medical parameters resulted in a modest 5% reduction.

Table 4 highlights the progressive improvement in accuracy with the addition of parameter categories.

Table 4. Prediction of AF patients and healthy individuals using LR method sorting based on accuracies

Accuracy of Demographic, Limited Measurement & Medical parameters (Test 7) (%)	Accuracy of Limited Measurement & Medical parameters (Test 6) (%)	Accuracy of Medical parameters (Test 3) (%)	Accuracy of Limited Measurement parameters (Test 2) (%)	Accuracy of Demographic & Limited Measurement parameters (Test 4) (%)	Accuracy of Demographic & Medical parameters (Test 5) (%)	Accuracy of Demographic parameters (Test 1) (%)
75.00	72.50	70.00	68.13	68.13	67.50	63.12

4. Discussion

Our findings reveal significant insights into the utility of ML in AF prediction. Unlike prior studies relying on complex parameters from ECG, HRV signals, or imaging techniques, this study adopted a simplified approach using accessible parameters from the SHHS dataset, enhancing clinical feasibility.

Numerous notable studies, such as those conducted by Park et al. [20], Zhang et al. [21], Huang et al. [11], and Jafari et al. [13], utilized advanced diagnostic tools but lacked the focus on accessible parameters. In contrast, our study prioritized simplicity and clinical feasibility using only demographic, measurement, and medical history parameters. This demonstrates that our simplified approach can deliver comparable performance while being more accessible and cost-effective for real-world applications.

Our study leverages the SHHS dataset for AF prediction by combining diverse yet simple clinical features into a single model. By evaluating seven ML algorithms, we provide a comprehensive analysis of their performance, highlighting LR as the most robust predictor. Unlike prior research focused on a single parameter category or diagnostic tool, our approach emphasizes the importance of integrating multiple accessible data types for enhanced predictive accuracy.

By evaluating seven ML algorithms, the study demonstrated the potential for early AF prediction using minimal data. This facilitates efficient integration into routine clinical workflows.

The findings have significant implications for early AF detection and management. By using readily available clinical features, the proposed models can be seamlessly integrated into routine health assessments, enabling earlier diagnosis and intervention. This could reduce healthcare costs and improve patient outcomes by preventing complications associated with undiagnosed AF.

5. Conclusion

This study highlights the transformative potential of ML in AF prediction, particularly through the application of simple, accessible clinical features. By comparing multiple ML algorithms, we identified the strengths and limitations of various methods, with LR demonstrating the highest predictive accuracy.

The integration of ML into healthcare systems, as evidenced by our findings, can reduce costs and improve patient outcomes through early diagnosis and intervention. This research establishes a solid foundation for future advancements in healthcare, advocating for the practical adoption of ML-based predictive models in real-world settings.

Authors' Contributions

All authors equally contributed to this study.

Declaration

None.

Transparency Statement

None.

Acknowledgments

This work is partially supported upon research funded by Iran National Science Foundation (INSF) under project No.4024604.

Declaration of Interest

The authors declare that they have no conflict of interest. The authors also declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

According to the authors, this article has no financial support.

Ethical Considerations

Not applicable.

References

[1] W. H. Federation. "Cardiovascular Disease (CVD)." https://world-heart-federation.org/what-is-cvd/ (accessed November 22, 2022, 2022).

[2] W. H. Organization. "Cardiovascular diseases (CVDs)." https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) (accessed November 22, 2022, 2022).

[3] N. Inform. "Cardiovascular disease." https://www.nhsinform.scot/illnesses-and-conditions/heart-and-blood-vessels/conditions/cardiovascular-disease (accessed.

[4] E. O. Lopez, B. D. Ballard, and A. Jan, "Cardiovascular disease," in *StatPearls [Internet]*: StatPearls Publishing, 2023.

[5] NHS. "Cardiovascular disease." https://www.nhs.uk/conditions/cardiovascular-disease/ (accessed November 22, 2022, 2022).

[6] N. Md Idris, Y. K. Chiam, K. D. Varathan, W. A. Wan Ahmad, K. H. Chee, and Y. M. J. M. Liew, "Feature selection and risk prediction for patients with coronary artery disease using data mining," *Medical & biological engineering & computing*, vol. 58, pp. 3123-3140, 2020. [Online]. Available: https://doi.org/10.1007/s11517-020-02268-9.

[7] A. S. Mohd Faizal, W. Y. Hon, T. M. Thevarajah, S. M. Khor, and S.-W. Chang, "A biomarker discovery of acute myocardial infarction using feature selection and machine learning," *Medical & biological engineering & computing*, pp. 1-15, 2023. [Online]. Available: https://doi.org/10.1007/s11517-023-02841-y.

[8] C. Krittanawong *et al.*, "Machine learning prediction in cardiovascular diseases: a meta-analysis," *Scientific reports*, vol. 10, no. 1, p. 16057, 2020. [Online]. Available: https://doi.org/10.1038/s41598-020-72685-1.

[9] M. H. Saeed and J. I. Hama, "Cardiac disease prediction using AI algorithms with SelectKBest," *Medical & Biological Engineering & Computing*, pp. 1-12, 2023. [Online]. Available: https://doi.org/10.1007/s11517-023-02918-8.

[10] B. A. S. Oliveira, G. Z. Castro, G. L. M. Ferreira, and F. G. Guimarães, "CML-Cardio: a cascade machine learning model to predict cardiovascular disease risk as a primary prevention strategy," *Medical & Biological Engineering & Computing*, pp. 1-17, 2023. [Online]. Available: https://doi.org/10.1007/s11517-022-02757-z.

[11] Y. Huang *et al.*, "Using a machine learning-based risk prediction model to analyze the coronary artery calcification score and predict coronary heart disease and risk assessment," *Computers in Biology Medicine*, vol. 151, p. 106297, 2022. [Online]. Available: https://doi.org/10.1016/j.compbiomed.2022.106297.

[12] A. S. M. Faizal, T. M. Thevarajah, S. M. Khor, and S.-W. Chang, "A review of risk prediction models in cardiovascular disease: conventional approach vs. artificial intelligent approach," *Computer methods programs in biomedicine*, vol. 207, p. 106190, 2021. [Online]. Available: https://doi.org/10.1016/j.cmpb.2021.106190.

[13] M. Jafari *et al.*, "Automated diagnosis of cardiovascular diseases from cardiac magnetic resonance imaging using deep learning models: A review," *Computers in Biology Medicine*, p. 106998, 2023. [Online]. Available: https://doi.org/10.1016/j.compbiomed.2023.106998.

[14] E. Ebrahimzadeh, M. Kalantari, M. Joulani, R. S. Shahraki, F. Fayaz, and F. Ahmadi, "Prediction of paroxysmal Atrial Fibrillation: A machine learning based approach using combined feature vector and mixture of expert classification on HRV signal," *Computer methods programs in biomedicine*, vol. 165, pp. 53-67, 2018. [Online]. Available: https://doi.org/10.1016/j.cmpb.2018.07.014.

[15] P. Ghosh *et al.*, "Efficient prediction of cardiovascular disease using machine learning algorithms with relief and LASSO feature selection techniques," *IEEE Access*, vol. 9, pp. 19304-19326, 2021. [Online]. Available: https://doi.org/10.1109/ACCESS.2021.3053759.

[16] D. Shah, S. Patel, and S. K. Bharti, "Heart disease prediction using machine learning techniques," *SN Computer Science*, vol. 1, pp. 1-6, 2020. [Online]. Available: https://doi.org/10.1007/s42979-020-00365-y.

[17] H. Jindal, S. Agrawal, R. Khera, R. Jain, and P. Nagrath, "Heart disease prediction using machine learning algorithms," in *IOP conference series: materials science and engineering*, 2021, vol. 1022, no. 1: IOP Publishing, p. 012072. [Online]. Available: https://doi.org/10.1088/1757-899X/1022/1/012072.

[18] R. Katarya and P. Srinivas, "Predicting heart disease at early stages using machine learning: A survey," in *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, 2020: IEEE, pp. 302-305. [Online]. Available: https://doi.org/10.1109/ICESC48915.2020.9155586.

[19] A. Yadav, L. Gediya, and A. Kazi, "Heart Disease Prediction Using Machine Lerning," *International Research Journal of Engineering and Technology (IRJET)*, vol. 8, no. 9, 2021. [Online]. Available: https://www.irjet.net/archives/V8/i9/IRJET-V8I9191.pdf.

[20] J.-U. Park, E. Urtnasan, S.-H. Kim, and K.-J. Lee, "A prediction model of incident cardiovascular disease in patients with sleep-disordered breathing," *Diagnostics*, vol. 11, no. 12, p. 2212, 2021. [Online]. Available: https://doi.org/10.3390/diagnostics11122212.

[21] L. Zhang, H. Wu, X. Zhang, X. Wei, F. Hou, and Y. Ma, "Sleep heart rate variability assists the automatic prediction of long-term cardiovascular outcomes," *Sleep medicine*, vol. 67, pp. 217-224, 2020.

[22] X. Zhang and H. Xu, "Prediction of Angina Pectoris Events in Middle-Aged and Elderly People Using RR Interval Time Series in the Resting State: A Cohort Study

Based on SHHS," *International Journal of Computational Intelligence Systems*, vol. 16, no. 1, p. 10, 2023.

[23] B. N. Shirazi, S. Naghsh, A. A. Safavi, and A. Sharafkhaneh, "Machine Learning-Based Prediction of Coronary Heart Disease," in *2023 9th International Conference on Control, Instrumentation and Automation (ICCIA)*, 2023: IEEE, pp. 1-5. [Online]. Available: https://doi.org/10.1109/ICCIA61416.2023.10506325.

[24] S. F. Quan *et al.*, "The sleep heart health study: design, rationale, and methods," *Sleep*, vol. 20, no. 12, pp. 1077-1085, 1997. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/9493915/.

[25] G.-Q. Zhang *et al.*, "The National Sleep Research Resource: towards a sleep data commons," *Journal of the American Medical Informatics Association*, vol. 25, no. 10, pp. 1351-1358, 2018. [Online]. Available: https://doi.org/10.1093/jamia/ocy064.

[26] P. C. Sen, M. Hajra, and M. Ghosh, "Supervised classification algorithms in machine learning: A survey and review," in *Emerging Technology in Modelling and Graphics: Proceedings of IEM Graph 2018*, 2020: Springer, pp. 99-111. [Online]. Available: https://doi.org/10.1007/978-981-13-7403-6_11.