International Journal of Scientific World, 11 (2) (2025) 8-14

International Journal of Scientific World

Website: www.sciencepubco.com/index.php/IJSW
https://doi.org/10.14419/m43wcb19
Review paper

Heart Attack Prediction in The United States: A Review

Botan Shivan Mustafa [1], Omar Sedqi Kareem [2]

[1] IT Department, College of Informatics, Akre University for Applied Sciences
[2] Department of Public Health, College of Health and Medical Technology-Shekhan, Duhok Polytechnic University, Duhok, Kurdistan Region–Iraq
**Corresponding author E-mail: botanamedi8@gmail.com*

Received: May 19, 2025, Accepted: June 17, 2025, Published: June 29, 2025

Abstract

Cardiovascular disease remains a predominant health concern in the United States, with over 600,000 annual fatalities, underscoring the necessity for advanced predictive methodologies. This review examines the integration of artificial intelligence (AI) and machine learning (ML) techniques in forecasting myocardial infarctions through the analysis of clinical, behavioral, and demographic data. Various supervised learning algorithms, including Support Vector Machines (SVM), Random Forests (RF), Gradient Boosting, Logistic Regression, and Deep Neural Networks (DNN), have demonstrated significant diagnostic accuracy. The application of optimization strategies such as Particle Swarm Optimization (PSO), dimensionality reduction methods like Principal Component Analysis (PCA), and automated machine learning (AutoML) frameworks has enhanced model efficiency and adaptability. Nonetheless, challenges persist, notably in real-time deployment, dataset representativeness of minority populations, interpretability of complex models, and cross-environment generalizability. This synthesis highlights current advancements, identifies key limitations, and suggests future research directions focused on developing scalable, interpretable, and equitable predictive systems to facilitate early detection and personalized cardiac care.

Keywords: *Heart Attack Prediction, Machine Learning, Cardiovascular Disease, Ensemble Learning, Feature Selection, AutoML, U.S. Healthcare Data*

1. Introduction

Cardiovascular disease (CVD) remains the leading cause of death globally, responsible for over 18 million deaths annually and expected to rise due to aging populations and increasing risk factors like hypertension, obesity, and diabetes[1]. In the U.S., heart failure (HF) affects more than 6.7 million people and is projected to exceed 8 million within a decade, emphasizing the urgency of preventive tools and early intervention[2]. Traditional diagnostic systems, though well-established, are often limited in handling complex, high-dimensional patient data, prompting the shift toward artificial intelligence (AI), particularly machine learning (ML), to enhance predictive accuracy in cardiovascular care[3].

Numerous studies have demonstrated the effectiveness of machine learning algorithms such as Support Vector Machines (SVM), Random Forests (RF), Gradient Boosting, and Deep Neural Networks (DNN). For example, ensemble methods have achieved significant improvements in prediction accuracy, with ensemble models outperforming individual classifiers [4]. Survival prediction in heart failure was addressed using sampling strategies and feature selection, demonstrating improved outcomes even with imbalanced datasets [5]. Deep learning models enhanced with preprocessing methods like C4.5 and KNN, as applied by Singh et al., achieved over 95% accuracy[6].

While demonstrating the effectiveness of boosting SVM models in improving diagnostic sensitivity[7]. Meta-analyses have reinforced the consistency of gradient boosting and neural networks across cardiovascular disease datasets [3]. Meanwhile, TabNet, used in recent studies, introduced interpretable AI for structured medical data, outperforming traditional classifiers on smaller datasets[8].

However, predictive accuracy alone is insufficient for real-world deployment; fairness and interpretability must also be addressed. Association rule mining combined with classification techniques can enhance both performance and transparency, emphasizing the importance of interpretable models in clinical settings [9]. Their work also underscored a critical challenge: ML models often suffer from biases when trained on datasets lacking demographic diversity. This can result in the underrepresentation of vulnerable populations, leading to misclassification or disparities in diagnosis [9]. Empirical evidence demonstrates that physicians tend to under-test high-risk patients and over-test low-risk ones due to mispredictions, a gap that AI tools could help mitigate if integrated thoughtfully into decision workflows [10].

Beyond prediction, effective communication of risk is crucial. The use of Predicted Heart/Vascular Age (PHA) can improve understanding and motivate behavioral changes, particularly in low-resource settings where literacy may pose a challenge [11]. Meanwhile, the economic implications of rising CVD incidence, as simulated decades ago, remain relevant today, highlighting the long-term societal cost of insufficient early detection systems[8].

This review aims to explore and analyze the current advancements in heart attack prediction within the United States, emphasizing the integration of machine learning techniques. It evaluates a wide range of algorithms, datasets, and optimization methods used to enhance diagnostic accuracy. The review also identifies existing limitations and proposes future directions for developing interpretable, equitable, and real-time predictive systems in U.S. healthcare.

This paper is structured into several sections to provide a comprehensive review of heart attack prediction in the United States. Section 2 presents the Background Theory, explaining the epidemiological trends of cardiovascular disease, the evolution of medical data usage, and how machine learning and optimization techniques are being applied to enhance prediction accuracy.

Section 3 outlines the Literature Review, summarizing recent research efforts, models, datasets, and performance metrics, while identifying key limitations and proposed solutions across 17 studies. Section 4 provides the Discussion, offering critical insights into the comparative performance of algorithms, challenges in clinical deployment, and the need for explainable and equitable AI. Section 5 highlights Extracted Statistics, visually summarizing algorithm categories and accuracy levels through pie charts for clearer pattern recognition. Section 6 delivers the Conclusion, emphasizing the progress made and the need for real-time, multimodal, and fair AI systems for heart attack prediction. Finally, Section 7 lists Recommendations for future work, focusing on model scalability, data diversity, explainability, and integration into clinical practice to support proactive cardiac care.

2. Background Theory

2.1 Overview of Heart Disease in the United States

Heart disease remains the leading cause of death in the U.S., with over 600,000 deaths annually and millions more suffering from chronic cardiac conditions. Coronary artery disease (CAD) and heart attacks represent the most common forms, largely driven by modifiable risk factors like hypertension, cholesterol, obesity, smoking, and poor diet[12],[13].

2.2 Epidemiological Trends and Demographic Shifts

While age-standardized death rates for heart disease have declined since the late 1960s due to medical advancements and public health interventions, demographic trends such as population growth and aging are projected to increase total death counts through 2020 and beyond[12]. As baby boomers age, the burden of CVD is shifting toward older populations.

2.3 The Role of Medical Data and Technology

The healthcare sector generates massive amounts of patient data daily. These records, including structured (lab results) and unstructured (clinical notes) formats, provide valuable insights if mined appropriately. However, traditional systems struggle to uncover patterns without intelligent tools[14],[13].

2.4 Machine Learning in Heart Disease Prediction

Machine learning (ML) offers promising solutions for analyzing large and complex healthcare datasets. Algorithms such as Decision Tree, Random Forest, Naïve Bayes, and Support Vector Machines have been extensively used for heart disease prediction with varying degrees of success. Recent models have achieved up to 98% accuracy using hybrid deep learning techniques like RNN with GRU[15],[16].

2.5 Feature Engineering and Optimization Techniques

To enhance prediction accuracy, many studies apply techniques like Principal Component Analysis (PCA), Synthetic Minority Over-sampling Technique (SMOTe), and Particle Swarm Optimization (PSO) to balance datasets and reduce feature redundancy[16],[17].

2.6 Automated Machine Learning (AutoML)

Frameworks like AutoGluon and AutoKeras automate the selection and tuning of predictive models, reducing the need for manual configuration. These systems have demonstrated strong performance on standard heart disease datasets like Cleveland and NHANES [18].

2.7 Challenges in Real-World Implementation

Despite high model accuracy in research settings, challenges persist in clinical adoption. These include limited generalizability, data imbalance, interpretability of deep learning models, and unequal health data representation across populations[13],[19].

2.8 Toward Personalized and Preventive Cardiac Care

The integration of ML with nationwide datasets and electronic health records presents an opportunity for real-time, personalized heart attack prediction. These predictive systems can support clinicians in identifying high-risk patients early and allocating targeted interventions to reduce mortality[19].

3. Literature reviews

This section provides a detailed review of recent studies focused on heart attack prediction using machine learning and data-driven approaches. It examines 17 research articles, analyzing the algorithms employed, dataset characteristics, accuracy levels, and methodological innovations. The review highlights the strengths and weaknesses of each study, including issues like data imbalance, lack of real-time application, and model interpretability. By comparing diverse models from traditional classifiers to ensemble and deep learning techniques, this section identifies research trends, performance benchmarks, and future development opportunities in the domain of cardiovascular risk prediction.

Pranav Motarwar et al., 2020 developed a machine learning framework to predict heart disease using five algorithms: Random Forest, Naïve Bayes, Support Vector Machine, Hoeffding Tree, and Logistic Model Tree. The study used the Cleveland dataset with 13 selected features and applied enhancement techniques like feature selection, data encoding, and hyperparameter tuning. Among the tested methods, Random Forest achieved the highest accuracy of 95.08%. The research focused on improving early prediction accuracy for heart disease through systematic preprocessing and algorithm-specific optimizations. However, the model was trained only on the Cleveland dataset, limiting its generalizability. To address this, they suggested future work should involve local hospital datasets and broader attribute inclusion. The limitation of using a single dataset could be mitigated by integrating diverse, real-world clinical data from multiple sources for more robust model performance[20].

Mohammed Jawwad Ali Junaid and Rajeev Kumar (2020 proposed a hybrid heart disease prediction model using data science algorithms, including Naïve Bayes, SVM, and ANN, aiming for early-stage diagnosis and personalized suggestions. The researchers trained the model on a dataset incorporating traditional attributes (like age, cholesterol) and expanded features (e.g., lifestyle, BMI, family history), achieving 88.54% accuracy, 82.11% specificity, and 91.47% sensitivity with the hybrid approach. Their research focused on enhancing accuracy and patient awareness through machine learning integration and risk-level classification. However, the model's main limitation was its reliance on manually entered or static data, limiting real-time adaptability. To overcome this, they recommended integrating smart wearable devices with cloud-connected real-time monitoring to improve continuous prediction and personalized healthcare delivery[21].

Javier Gamboa-Cruzado et al., 2024 conducted a systematic literature review and bibliometric analysis on heart attack prediction using machine learning, analyzing 82 articles published between 2017 and 2021. The study focused on identifying common algorithms such as Artificial Neural Networks, Random Forest, and Decision Trees, with classification and regression being the dominant methodologies. It highlighted confusion matrices and PCA as key evaluation criteria and found that ensemble methods often outperform individual models. However, the research was limited by a narrow publication window (2017–2021) and restricted search sources, possibly excluding recent or diverse contributions. The authors recommended expanding the temporal range and search databases in future reviews to capture evolving trends and methodologies. To address the limitation, future research should include newer studies from broader sources to ensure a more comprehensive analysis of advancements in heart attack prediction models[22].

Soomi Lee et al. (2022 investigated the link between multidimensional sleep health and heart disease risk using two sleep health composites—one based on self-reports and the other combining self-report and actigraphy in a middle-aged U.S. adult population (N=6,820 and N=663, respectively). The study used modified Poisson regression to analyze sleep dimensions like regularity, satisfaction, alertness, timing, efficiency, and duration. Key findings showed that each unit increase in poor sleep health raised heart disease risk by 54% (self-report) and 141% (actigraphy/self-report), highlighting the stronger predictive power of composite sleep health over individual measures. The research emphasized the need to assess co-existing sleep issues across diverse populations. However, its limitation was the cross-sectional nature of the data, which restricted causal inference over time. This could be addressed by conducting longitudinal studies to determine the directionality and progression of sleep-related heart disease risk across adulthood[23].

Suraj Kumar Gupta et al., 2021 developed a heart attack prediction system using supervised machine learning algorithms, including Gradient Boosting, Random Forest, Decision Tree, and Logistic Regression. They applied the model on two datasets: the Framingham and UCI Heart datasets, achieving the highest accuracy of 85.6% with the Gradient Boosting classifier after feature engineering and hyperparameter optimization. Their research focused on identifying key risk factors (e.g., chest pain, cholesterol, heart rate) and improving prediction accuracy through pipeline optimization. The methodology involved iterative pipeline enhancement and performance comparison across models. However, the study was limited by reliance on traditional supervised models without integration of deep learning or semi-supervised techniques. To overcome this, future work should explore deep learning architectures and semi-supervised learning methods to further enhance prediction accuracy and adaptability[24].

Ridwan B. Marqas et al., 2023 developed a machine learning model to predict heart attack risk in high-risk patients using real-world data from Hungarian, Cleveland, and Statlog datasets. The study evaluated multiple algorithms, including Random Forest, KNN, MLP, SVC, AdaBoost, and XGBoost, with XGBoost achieving the highest test accuracy of 91.91%, sensitivity of 94.3%, and F1-score of 0.9243. The research focused on enhancing early detection using demographic, medical, and lifestyle features, supported by preprocessing, feature engineering, and cross-validation. However, the study was limited by the interpretability of complex models and the scope of a single combined dataset. To address this, future work should explore explainable AI methods and expand datasets with more diverse real-world patient records for broader applicability[25].

Liaqat Ali et al., 2019, proposed an automated heart disease diagnostic system based on a χ^2 statistical model and an optimally configured Deep Neural Network (DNN) using the Cleveland dataset. The research focused on solving underfitting and overfitting issues by selecting relevant features through the χ^2 test and fine-tuning the DNN structure via exhaustive grid search. The model achieved a prediction accuracy of 93.33%, outperforming conventional ANN, DNN, and other state-of-the-art models. Key methods included feature selection, network depth optimization, and extensive performance evaluation using metrics like AUC, F1-score, and MCC. However, the study did not address the time complexity of the system and relied on computationally intensive grid search. To address this, future work should adopt more efficient optimization methods like genetic algorithms and assess model runtime for real-world applications[26].

Odai Y. Dweekat et al., 2022 proposed a heart disease prediction model using seven classification algorithms: Logistic Regression, SVM, Gaussian Naive Bayes, K-Nearest Neighbors, Decision Tree, Random Forest, and Discriminant Analysis—on a dataset of 1,025 patients with 13 risk factors. The study aimed to predict heart disease and uncover hidden relationships using association rule mining (ARM). Logistic Regression achieved the highest accuracy of 84.6%, while SVM yielded the highest sensitivity of 91.3%. Methods included ten-fold cross-validation, performance metrics evaluation, and statistical testing (Wald test, Hosmer-Lemeshow test). However, the model was limited by the absence of advanced feature selection or dimensionality reduction techniques. This could be improved by incorporating automated feature extraction or evolutionary optimization methods to boost predictive performance and interpretability[27].

Mohammad Alshraideh et al., 2024 developed a heart disease prediction model using five machine learning algorithms, SVM, Random Forest, Decision Tree, Naive Bayes, and KNN—on a dataset of 486 patients from Jordan University Hospital (JUH). The study focused on improving diagnostic accuracy through preprocessing, feature engineering, and applying Particle Swarm Optimization (PSO) for feature selection. Their model achieved 94.3% accuracy with SVM + PSO, outperforming other algorithms. The research emphasized early detection and personalized healthcare. However, the model's limitation was its reliance on a single-center, retrospective dataset, which could limit generalizability. This limitation could be mitigated by incorporating diverse, multi-center real-time data and applying explainable AI techniques to enhance model robustness and clinical trust[28].

Cheryl Ann Alexander and Lidong Wang, 2017 explored the use of Big Data analytics for heart attack prediction, focusing on integrating machine learning, IoT, telecardiology, and data mining across structured and unstructured health data. They reviewed 31 studies and highlighted tools like Hadoop, HDFS, and text mining in patient-customized healthcare systems. Their findings emphasized that technologies such as implanted sensors, wearable devices, and real-time EHR monitoring significantly enhance early heart attack detection. However, the primary limitation was the challenge of managing heterogeneous, noisy, and privacy-sensitive health data. To address this, they recommended developing standardized frameworks for data interoperability and enhancing privacy-preserving mechanisms to maintain data trust and usability across healthcare platforms[29].

Aravind Sasidharan Pillai, 2022, developed a cardiac disease prediction system using TabNet, a deep learning model designed for tabular data, and compared it with traditional models like Logistic Regression, Random Forest, XGBoost, and Gradient Boosting on the UCI Cleveland dataset (297 records). The research focused on enhancing diagnostic accuracy using interpretable AI with sequential attention mechanisms. TabNet achieved the highest performance with 94.4% accuracy, 0.94 ROC, and over 93% precision and recall, outperforming all baseline models. Key methods included preprocessing, model comparison, and feature importance analysis. However, the model was limited by the small dataset size and lack of real-world deployment testing. This could be improved by training on larger, real-time healthcare datasets and integrating the system into clinical workflows for broader validation and application[30].

Saeed Amal et al. (2022 reviewed how multi-modal data and machine learning techniques can improve cardiovascular disease (CVD) care through enhanced diagnostics, risk prediction, and treatment personalization. The study focused on integrating diverse data types EMR, imaging, genetics, and biosignals, using algorithms like XGBoost, CNN, SVM, and ensemble deep learning. Fusion models consistently outperformed traditional single-source models, with AUCs reaching 0.86 and accuracy up to 96.67%. Methods included early, late, and joint data fusion strategies across various use cases. However, the study was limited by the complexity and lack of standardized frameworks for implementing data fusion in real clinical settings. This limitation could be resolved by developing user-friendly, scalable architectures for data integration and establishing best practices to support real-time clinical adoption[31].

Izabela Rojek et al., 2024 developed an AI-based heart attack prediction system using multiple machine learning algorithms LinearSVC, Logistic Regression, KNN, and Random Forest on a Kaggle dataset of 8763 samples with 26 features. The research aimed to enable cost-effective, early-stage risk assessment in preventive medicine by identifying key predictors such as heart rate, age, BMI, and cholesterol. Logistic Regression yielded the highest practical accuracy (~64%) for initial screenings. The study emphasized reducing the number of input features and improving accessibility through non-invasive data collection. However, limitations included low precision for positive class prediction and generalizability issues due to data imbalance and static modeling. This limitation could be addressed by incorporating real-time monitoring, dynamic data updates, and hybrid deep learning approaches to improve prediction precision and broader applicability[32].

Ahmed M. Alaa et al., 2019, developed a cardiovascular disease prediction model using an automated machine learning framework (AutoPrognosis) applied to the UK Biobank dataset with 423,604 participants and 473 variables. The study aimed to enhance CVD risk prediction beyond traditional tools like the Framingham Score by automatically selecting and tuning ML pipelines such as XGBoost, Random Forest, and AdaBoost. Their model achieved the highest AUC-ROC of 0.774, outperforming standard Cox models and ML baselines. Key methods included Bayesian optimization, ensemble modeling, and the use of non-lab data like walking pace and self-rated health. The main limitation was the absence of key biomarkers (e.g., cholesterol, triglycerides), which constrained direct comparison with clinical scores like QRISK2. This limitation could be addressed by integrating full biomarker panels and expanding to ethnically diverse cohorts to generalize findings[33].

Muhammad Waqas Nadeem et al., 2021, proposed a fusion-based supervised machine learning architecture using Support Vector Machine (SVM) integrated with fuzzy-based decision-level fusion for heart disease prediction. The system was tested on two Kaggle datasets (heart disease 2019 and cardiovascular disease 2019) and achieved 96.23% accuracy. Their approach focused on improving diagnostic accuracy by leveraging preprocessing (mean imputation, normalization), k-fold cross-validation, and parallel SVM training. The method outperformed prior models like ANN, hybrid ML systems, and ensemble approaches. However, a key limitation was the lack of real-time deployment and dependency on static datasets. This could be improved by integrating IoMT (Internet of Medical Things) for continuous data collection and real-time predictive updates[34].

P. Senthil Kumari and S. Vinitha, 2022, developed a heart attack prediction system using Decision Tree (DT) and Random Forest Algorithm (RFA) with feature reduction and Extreme Learning Machine (ELM) in an R programming environment. The research focused on identifying the most accurate classification algorithm using the UCI Cleveland dataset with 13 features. The Random Forest model achieved the highest accuracy of 95%, outperforming the Decision Tree's 85%. Methods included data preprocessing, attribute reduction, classification via RFA, and final prediction through ELM. However, the model relied on offline static datasets without real-time integration or cross-institutional validation. This limitation could be addressed by implementing real-time streaming data with Internet of Medical Things (IoMT) and validating the model across multiple healthcare environments[35].

Rahmanul Hoque et al., 2024 developed a heart disease prediction model using Support Vector Machines (SVM), comparing linear and polynomial kernels on the UCI Heart Disease dataset with 303 samples and 14 attributes. The research aimed to reduce misdiagnosis and improve early detection by applying machine learning to medical records. Key methods included data preprocessing, transformation, and performance evaluation using precision, recall, F1-score, and accuracy. Polynomial SVM outperformed Linear SVM, achieving 91% training accuracy and 80% test accuracy, demonstrating better generalization on transformed data. However, the model was limited by static data usage and a lack of integration with real-time diagnostics. This limitation could be addressed by incorporating real-time clinical data streams and deploying the model in IoT-enabled healthcare environments for continuous monitoring and dynamic prediction[36].

4. Discussion and comparison

The studies demonstrate significant diversity in heart attack prediction approaches, with traditional machine learning algorithms evolving toward sophisticated ensemble and deep learning methods. Traditional approaches include Motarwar et al. (2020) achieving 95.08% accuracy using Random Forest on the Cleveland dataset, Junaid & Kumar (2020) reaching 88.54% with hybrid NB/SVM/ANN algorithms, and Gupta et al. (2021) obtaining 85.60% through Gradient Boosting across Framingham and UCI datasets. Ensemble and optimization-enhanced methods showed superior performance, with Nadeem et al. (2021) achieving the highest accuracy of 96.23% using SVM with fuzzy decision fusion, Marqas et al. (2023) reaching 91.91% through XGBoost on multi-center datasets, and Alshraideh et al. (2024) obtaining 94.30% by integrating PSO with traditional algorithms. Advanced architectures demonstrated competitive results, including Ali et al. (2019) achieving 93.33% with chi-square feature selection and DNN, Pillai (2022) reaching 94.40% using interpretable TabNet, and Amal et al. (2022) attaining 96.67% through multi-modal fusion approaches. However, several studies faced significant limitations: most relied on static datasets (Cleveland, UCI), limiting generalizability, lacked real-time integration capabilities, suffered from computational complexity issues, and demonstrated poor interpretability for clinical deployment. Notably, large-scale studies like Alaa et al. (2019) using UK Biobank data (423,604 participants) achieved only 77.4% AUC-ROC due to missing biomarkers, while Rojek et al. (2024) obtained ~64% accuracy despite using extensive Kaggle datasets, highlighting persistent challenges in data imbalance and demographic representation. The collective findings indicate that while ensemble methods and deep learning achieve superior accuracy (>95%), critical gaps remain in real-world deployment, cross-institutional validation, and equitable performance across diverse populations, necessitating future research focused on automated optimization, explainable AI, and real-time clinical integration.

Table 1 provides a comparative summary of 17 recent studies on heart attack prediction models from references [20] to [36]. It outlines key algorithms, datasets, accuracies, limitations, and future directions to highlight methodological trends and performance benchmarks across the literature.

Author (Year)	Algorithms Used	Dataset	Accuracy	Research Focus	Limitations	Future Recommendations
Motarwar et al. (2020) [20]	RF, NB, SVM, Hoeffding Tree, LMT	Cleveland	95.08%	Improve early prediction through ML tuning	Used only one dataset	Use local hospital data, expand input features
Junaid & Kumar (2020) [21]	NB, SVM, ANN	Hybrid (traditional + lifestyle)	88.54%	Hybrid ML with personalization	Manual/static input only	Use wearable devices + cloud data integration
Gamboa-Cruzado et al. (2024) [22]	ANN, RF, DT (Review)	82 ML studies	85-98% (varied)	Systematic literature + trend analysis	Narrow time frame & sources	Include a broader range and post-2021 studies
Soomi Lee et al. (2022) [23]	Poisson Regression	U.S. Sleep Health Composite Data	Not specified	Link between sleep health and heart disease	Cross-sectional only	Conduct longitudinal cohort studies
Gupta et al. (2021) [24]	GB, RF, DT, LR	Framingham + UCI	85.60%	Feature engineering and model comparison	Lacked DL & semi-supervised methods	Test with DL & hybrid learning
Marqas et al. (2023) [25]	RF, KNN, MLP, SVC, AdaBoost, XGBoost	Hungarian, Cleveland, Statlog	91.91%	High-risk patient screening with real-world data	Model interpretability	Apply explainable AI and more diverse records
Ali et al. (2019) [26]	œá¬≤ + DNN	Cleveland	93.33%	Optimal DNN with statistical feature selection	High computational cost	Use genetic optimization and runtime-efficient models
Dweekat et al. (2022) [27]	LR, SVM, GNB, KNN, DT, RF, DA + ARM	1,025 patients	84.60%	Combine ML with rule mining for better interpretability	No dimensionality reduction	Use automated/evolutionary feature selection
Alshraideh et al. (2024) [28]	SVM, RF, DT, NB, KNN + PSO	Jordan University Hospital	94.30%	Improve accuracy using PSO feature selection	Single-center study	Expand to multi-center real-time systems
Alexander & Wang (2017) [29]	ML, Big Data, IoT, Telecardiology	Review of 31 studies	94.4%	Big Data for real-time heart attack prediction	Privacy, heterogeneity	Create privacy-preserving & interoperable frameworks
Pillai (2022) [30]	TabNet, LR, RF, GB, XGBoost	UCI Cleveland	94.40%	Interpretable DL using TabNet	Small sample, no real-world test	Use large, real clinical datasets
Amal et al. (2022) [31]	XGBoost, CNN, SVM, Deep Ensemble	Multi-modal (EMR, imaging, genetic, etc.)	96.67%	Fusion models for advanced CVD care	No clinical deployment models	Develop scalable integration with hospital systems
Rojek et al. (2024) [32]	LR, KNN, RF, LinearSVC	Kaggle (8,763 records, 26 features)	~64%	Screening tool with non-invasive inputs	Low precision, imbalance	Use hybrid DL + real-time sensor data
Alaa et al. (2019) [33]	AutoML (AutoPrognosis), XGB, RF, AdaBoost	UK Biobank (423,604 participants)	77.4 %	AutoML for scalable CVD prediction	Missing biomarker data	Add lab biomarkers & ethnicity-balanced sampling
Nadeem et al. (2021) [34]	SVM + Fuzzy Decision Fusion	Kaggle Heart & CVD datasets	96.23%	Fusion-based architecture using parallel SVM	Static datasets	Real-time via IoMT integration
Kumari & Vinitha (2022) [35]	DT, RF, ELM	UCI Cleveland	RF: 95%	Compare algorithms; use ELM as final classifier	No real-time/institutional testing	IoMT streaming, test across institutions
Hoque et al. (2024) [36]	SVM (Linear, Polynomial Kernels)	UCI Heart Disease	91% train, 80% test	Kernel performance & misdiagnosis reduction	Static input, no clinical integration	Enable live data pipeline with IoT

5. Extracted statistics

This section presents two extracted statistical insights from the reviewed literature on heart attack prediction models, focusing on the types of machine learning algorithms applied and the range of accuracy achieved. These visualizations help highlight prevailing research patterns and performance benchmarks across recent studies.

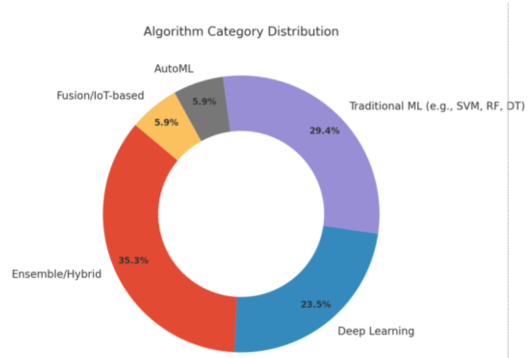

Fig. 1: Displays the proportion of Ensemble/Hybrid, Deep Learning, Traditional ML, AutoML, and Fusion/IoT-based models used in the studies.

Figure 1 presents the distribution of algorithm categories used in the reviewed studies. Ensemble/Hybrid models accounted for the largest share at 35.3%, highlighting their effectiveness in leveraging the strengths of multiple algorithms. Traditional machine learning methods (e.g., SVM, Random Forest, Decision Trees) followed with 29.4%, underscoring their continued popularity due to simplicity and reliability. Deep learning techniques represented 23.5%, reflecting their increasing adoption in recent research. Meanwhile, both Fusion/IoT-based models and AutoML approaches each comprised 5.9%, indicating their emerging role in automation and real-time intelligent systems.

6. Conclusion

The review demonstrates remarkable progress in heart attack prediction models, with multi-modal fusion approaches achieving the highest accuracy of 96.67%, followed by ensemble methods (96.23%) and optimized deep learning techniques (94.40%). These results highlight the superiority of advanced machine learning approaches in analyzing complex cardiovascular data. However, limitations such as reliance on static datasets, computational complexity, and lack of real-world clinical integration remain key barriers. Future efforts should prioritize real-time monitoring through IoT and wearable devices, expansion of diverse demographic datasets, and development of interpretable AI systems to ensure equitable and actionable predictions. By addressing these challenges, next-generation models can transition from research to clinical practice, enabling proactive and personalized heart disease prevention.

7. Recommendations

- Integrate Real-Time Data Sources: Future heart attack prediction systems should incorporate real-time patient monitoring using wearable devices, IoT, and mobile health technologies to enhance responsiveness and dynamic risk assessment.
- Expand Dataset Diversity: To improve generalizability, models should be trained and validated on multicenter datasets that reflect diverse demographics, clinical conditions, and social determinants of health.
- Promote Explainable AI: Increasing the interpretability of machine learning models is essential for clinician trust and clinical integration. Approaches such as rule-based systems, attention mechanisms, and visual explanations should be prioritized.
- Implement Standardized Evaluation Metrics: Researchers should adopt uniform performance evaluation protocols (e.g., AUC, sensitivity, specificity) across studies to enable accurate benchmarking and reproducibility.
- Focus on Ethical and Fair AI: Prediction models must ensure equity in healthcare outcomes by addressing bias and ensuring fair performance across race, age, and socioeconomic groups.
- Encourage Clinical Collaboration: Active collaboration with cardiologists and healthcare institutions is essential to align model development with practical diagnostic workflows and regulatory standards.
- Adopt Scalable AutoML Frameworks: Leveraging AutoML platforms can streamline model development and allow non-technical healthcare teams to build and deploy reliable prediction systems.

References

[1] K. E. Joynt Maddox et al., “Forecasting the Burden of Cardiovascular Disease and Stroke in the United States Through 2050-Prevalence of Risk Factors and Disease: A Presidential Advisory from the American Heart Association,” Jul. 23, 2024, Lippincott Williams and Wilkins. doi: 10.1161/CIR.0000000000001256.
[2] D. S. Matasic, R. Zeitoun, G. C. Fonarow, A. C. Razavi, R. S. Blumenthal, and M. Gulati, “Advancements in Incident Heart Failure Risk Prediction and Screening Tools,” American Journal of Cardiology, vol. 227, pp. 105–110, Sep. 2024, doi: 10.1016/j.amjcard.2024.07.014.
[3] Y. Baashar et al., “Effectiveness of Artificial Intelligence Models for Cardiovascular Disease Prediction: Network Meta-Analysis,” Comput Intell Neurosci, vol. 2022, 2022, doi: 10.1155/2022/5849995.
[4] C. B. C. Latha and S. C. Jeeva, “Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques,” Inform Med Unlocked, vol. 16, Jan. 2019, doi: 10.1016/j.imu.2019.100203.
[5] A. Newaz, N. Ahmed, and F. Shahriyar Haq, “Survival prediction of heart failure patients using machine learning techniques,” Inform Med Unlocked, vol. 26, Jan. 2021, doi: 10.1016/j.imu.2021.100772.
[6] M. S. Singh, K. Thongam, P. Choudhary, and P. K. Bhagat, “An Integrated Machine Learning Approach for Congestive Heart Failure Prediction,” Diagnostics, vol. 14, no. 7, Apr. 2024, doi: 10.3390/diagnostics14070736.
[7] E. Owusu, P. Boakye-Sekyerehene, J. K. Appati, and J. Y. Ludu, “Computer-Aided Diagnostics of Heart Disease Risk Prediction Using Boosting Support Vector Machine,” Comput Intell Neurosci, vol. 2021, 2021, doi: 10.1155/2021/3152618.
[8] M. C. Weinstein, P. G. Coxson, L. W. Williams, T. M. Pass, W. B. Stason, and L. Goldman, “Forecasting Coronary Heart Disease Incidence, Mortality, and Cost: The Coronary Heart Disease Policy Model.”
[9] Yadav, P., Jaiswal, K., Patel, S., & Shukla, D. (2013). Intelligent heart disease prediction model using classification algorithms. IJCSMC, 3(08), 102–107.
[10] Mullainathan, S., & Obermeyer, Z. (2019). Diagnosing Physician Error: A Machine Learning Approach to Low-Value Health Care. https://doi.org/10.3386/w26168
[11] D. Appiah and B. D. Capistrant, “Cardiovascular Disease Risk Assessment in the United States and Low- and Middle-Income Countries Using Predicted Heart/Vascular Age,” Sci Rep, vol. 7, no. 1, Dec. 2017, doi: 10.1038/s41598-017-16901-5.
[12] H. K. Weir et al., “heart disease and Cancer Deaths — Trends and Projections in the United States, 1969–2020,” Prev Chronic Dis, vol. 13, Nov. 2016, doi: 10.5888/PCD13.160211.
[13] J. Soni Ujma Ansari Dipesh Sharma and S. Associate Professor, “Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction Sunita Soni,” 2011.
[14] Chaitanya. Baru, 2019 IEEE International Conference on Big Data: proceedings: Dec 9 - Dec 12, 2019, Los Angeles, CA, USA. IEEE, 2019.
[15] S. Nikhar and A. M. Karandikar, “Prediction of Heart Disease Using Machine Learning Algorithms,” International Journal of Advanced Engineering, Management and Science (IJAEMS), vol. 2, no. 6, 2016, [Online]. Available: www.ijaems.com
[16] S. Krishnan, P. Magalingam, and R. Ibrahim, “Hybrid deep learning model using recurrent neural network and gated recurrent unit for heart disease prediction,” International Journal of Electrical and Computer Engineering, vol. 11, no. 6, pp. 5467–5476, Dec. 2021, doi: 10.11591/ijece.v11i6.pp5467-5476.
[17] 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT). IEEE, 2019.
[18] L. M. Paladino, A. Hughes, A. Perera, O. Topsakal, and T. C. Akinci, “Evaluating the Performance of Automated Machine Learning (AutoML) Tools for Heart Disease Diagnosis and Prediction,” AI (Switzerland), vol. 4, no. 4, pp. 1036–1058, Dec. 2023, doi: 10.3390/ai4040053.
[19] A. Dinh, S. Miertschin, A. Young, and S. D. Mohanty, “A data-driven approach to predicting diabetes and cardiovascular disease with machine learning.” [Online]. Available: http://creativecommons.org/publicdomain/zero/1.0/

[20] P. Motarwar, A. Duraphe, G. Suganya, and M. Premalatha, "Cognitive Approach for Heart Disease Prediction using Machine Learning," in International Conference on Emerging Trends in Information Technology and Engineering, ic-ETITE 2020, Institute of Electrical and Electronics Engineers Inc., Feb. 2020. doi: 10.1109/ic-ETITE47903.2020.242.

[21] 2020 International Conference on Intelligent Engineering and Management (ICIEM-2020). IEEE, 2020.

[22] J. Gamboa-Cruzado, R. Crisostomo-Castro, J. Vila-Buleje, J. López-Goycochea, and J. N. Valenzuela, "HEART ATTACK PREDICTION USING MACHINE LEARNING: A COMPREHENSIVE SYSTEMATIC REVIEW AND BIBLIOMETRIC ANALYSIS," J Theor Appl Inf Technol, vol. 15, no. 5, 2024, [Online]. Available: www.jatit.org

[23] S. Lee et al., "Sleep health composites are associated with the risk of heart disease across sex and race," Sci Rep, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022-05203-0.

[24] S. K. Gupta, A. Shrivastava, S. P. Upadhyay, and P. K. Chaurasia*, "A Machine Learning Approach for Heart Attack Prediction," Int J Eng Adv Technol, vol. 10, no. 6, pp. 124–134, Aug. 2021, doi: 10.35940/ijeat. F3043.0810621.

[25] R. B. Marqas, A. Mousa, F. Özyurt, and R. Salih, "A Machine Learning Model for the Prediction of Heart Attack Risk in High-Risk Patients Utilizing Real-World Data," Academic Journal of Nawroz University, vol. 12, no. 4, pp. 286–301, Oct. 2023, doi: 10.25007/ajnu.v12n4a1974.

[26] L. Ali, A. Rahman, A. Khan, M. Zhou, A. Javeed, and J. A. Khan, "An Automated Diagnostic System for Heart Disease Prediction Based on $\chi 2$ Statistical Model and Optimally Configured Deep Neural Network," IEEE Access, vol. 7, pp. 34938–34945, 2019, doi: 10.1109/ACCESS.2019.2904800.

[27] O. Y. Dweekat, N. Bazrafshan, and S. S. Lu, "Heart Disease Prediction and Factor Identification using Machine Learning Algorithms and Association Rule Mining," 2022.

[28] M. Alshraideh, N. Alshraideh, A. Alshraideh, Y. Alkayed, Y. Al Trabsheh, and B. Alshraideh, "Enhancing Heart Attack Prediction with Machine Learning: A Study at Jordan University Hospital," Applied Computational Intelligence and Soft Computing, vol. 2024, 2024, doi: 10.1155/2024/5080332.

[29] C. A. Alexander and L. Wang, "Big Data Analytics in Heart Attack Prediction," Journal of Nursing & Care, vol. 06, no. 02, 2017, doi: 10.4172/2167-1168.1000393.

[30] S. Pillai, "Cardiac Disease Prediction with Tabular Neural Network; Cardiac Disease Prediction with Tabular Neural Network", doi: 10.5281/zenodo.7750620.

[31] S. Amal, L. Safarnejad, J. A. Omiye, I. Ghanzouri, J. H. Cabot, and E. G. Ross, "Use of Multi-Modal Data and Machine Learning to Improve Cardiovascular Disease Care," Apr. 27, 2022, Frontiers Media S.A. doi: 10.3389/fcvm.2022.840262.

[32] I. Rojek, P. Kotlarz, M. Kozielski, M. Jagodziński, and Z. Królikowski, "Development of AI-Based Prediction of Heart Attack Risk as an Element of Preventive Medicine," Electronics (Switzerland), vol. 13, no. 2, Jan. 2024, doi: 10.3390/electronics13020272.

[33] A. M. Alaa, T. Bolton, E. Di Angelantonio, J. H. F. Rudd, and M. van der Schaar, "cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants," PLoS One, vol. 14, no. 5, May 2019, doi: 10.1371/journal.pone.0213653.

[34] M. W. Nadeem, H. G. Goh, M. A. Khan, M. Hussain, M. F. Mushtaq, and V. A. P. Ponnusamy, "Fusion-Based Machine Learning Architecture for Heart Disease Prediction," Computers, Materials and Continua, vol. 67, no. 2, pp. 2481–2496, 2021, doi: 10.32604/cmc.2021.014649.

[35] P. S. Kumari, P. Senthil Kumari, and S. Vinitha, "Heart Attack Prediction in Machine Learning Environment International Journal of Scientific Research in Engineering and Management (IJSREM) Heart Attack Prediction in Machine Learning Environment," 2022, [Online]. Available: https://www.researchgate.net/publication/362630808

[36] Rahmanul Hoque, Masum Billah, Amit Debnath, S. M. Saokat Hossain, and Numair Bin Sharif, "Heart Disease Prediction using SVM," International Journal of Science and Research Archive, vol. 11, no. 2, pp. 412–420, Mar. 2024, doi: 10.30574/ijsra.2024.11.2.0435.