IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 15, Issue 9, September 2026

DOI: 10.17148/IJARCCE.2026.15933

Heart Disease Prediction Using Machine Learning and Deep Learning Techniques: A Comprehensive Review

Geetha. N[1], Janasuruthi. M[2], Sripavanya. K[3]

Assistant Professor, Department of Computer Applications, Vellalar College for Women, Thindal, Erode, Tamil Nadu, India[1]

II MCA, Department of Computer Applications, Vellalar College for Women, Thindal, Erode, Tamil Nadu, India[2,3]

Abstract: Heart disease remains one of the leading causes of mortality worldwide, accounting for millions of deaths annually despite significant advancements in healthcare technologies. Early prediction and timely diagnosis play a crucial role in reducing mortality rates and improving patient outcomes. Conventional diagnostic approaches rely heavily on clinical expertise, laboratory investigations, and medical imaging, which are often time-consuming, expensive, and susceptible to human interpretation errors. Recent developments in Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) have revolutionized cardiovascular disease prediction by enabling automated analysis of large-scale clinical datasets with high accuracy and efficiency. Machine learning algorithms such as Logistic Regression, Decision Trees, Support Vector Machines, Random Forests, Naïve Bayes, K-Nearest Neighbors, Gradient Boosting, and Extreme Gradient Boosting have demonstrated remarkable capability in predicting heart disease using structured clinical data. Simultaneously, deep learning architectures including Artificial Neural Networks, Convolutional Neural Networks, Recurrent Neural Networks, Long Short-Term Memory networks, Autoencoders, and Transformer-based models have shown superior performance in analyzing electrocardiograms, echocardiograms, cardiac MRI, CT scans, wearable sensor data, and multimodal healthcare information. This review comprehensively examines recent advancements in heart disease prediction. The study discusses traditional statistical approaches, modern machine learning algorithms, deep learning techniques, hybrid intelligent models, explainable artificial intelligence, federated learning, Internet of Medical Things (IoMT), cloud-based healthcare systems, and wearable monitoring technologies. Furthermore, this review summarizes the strengths, limitations, datasets, evaluation metrics, and prediction performance of recent studies through a comparative analysis. Existing research gaps, practical challenges, future research directions, and emerging opportunities are also highlighted to guide researchers toward developing reliable, interpretable, and clinically deployable heart disease prediction systems. The findings indicate that integrating explainable AI, multimodal learning, privacy-preserving federated learning, and real-time wearable monitoring represents the future direction of intelligent cardiovascular healthcare.

Keywords: Heart Disease Prediction, Machine Learning, Deep Learning, Cardiovascular Disease, Artificial Intelligence, Explainable AI, Internet of Medical Things, Federated Learning, Healthcare Analytics.

I. INTRODUCTION

Cardiovascular diseases (CVDs), commonly referred to as heart diseases, represent one of the most significant global public health concerns. According to recent estimates by the World Health Organization (WHO), cardiovascular diseases account for nearly one-third of all global deaths, causing approximately 18 million deaths annually. The increasing prevalence of heart disease is associated with multiple lifestyle factors including obesity, hypertension, diabetes, smoking, alcohol consumption, unhealthy dietary habits, physical inactivity, stress, and aging populations. These risk factors have significantly increased the burden on healthcare systems worldwide.

Heart disease encompasses a broad range of disorders affecting the cardiovascular system, including coronary artery disease, congestive heart failure, arrhythmia, myocardial infarction, ischemic heart disease, congenital heart disease, valvular heart disease, peripheral artery disease, and cardiomyopathy. Among these, coronary artery disease remains the most common cause of sudden cardiac death worldwide.

Early identification of individuals at high risk has become an important research objective because timely intervention significantly improves survival rates while reducing healthcare costs. Traditional diagnostic procedures include electrocardiography (ECG), echocardiography, stress testing, coronary angiography, computed tomography (CT), magnetic resonance imaging (MRI), blood biomarker analysis, and physician assessment. Although these diagnostic

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⋈ Peer-reviewed & Refereed journal ⋈ Vol. 15, Issue 9, September 2026

DOI: 10.17148/IJARCCE.2026.15933

techniques provide accurate clinical information, they often require expensive medical infrastructure, specialized cardiologists, and substantial diagnostic time. Moreover, manual interpretation of clinical reports introduces subjectivity and increases the possibility of diagnostic errors.

The rapid advancement of Artificial Intelligence (AI) has transformed healthcare by enabling automated clinical decision support systems capable of analyzing complex medical datasets. Machine Learning (ML), a subset of AI, allows computers to learn hidden relationships from historical clinical data without explicit programming. In recent years, ML algorithms have demonstrated excellent performance in identifying complex nonlinear relationships among cardiovascular risk factors such as age, gender, blood pressure, cholesterol, fasting blood sugar, electrocardiographic findings, chest pain type, maximum heart rate, exercise-induced angina, body mass index, smoking history, diabetes status, family medical history, and genetic information.

Supervised machine learning algorithms remain the most widely adopted techniques for heart disease prediction. Logistic Regression serves as an effective baseline classifier due to its simplicity and interpretability. Decision Tree algorithms provide transparent rule-based predictions, making them suitable for clinical applications where explainability is essential. Random Forest improves prediction performance by combining multiple decision trees, thereby reducing overfitting and increasing robustness. Support Vector Machine (SVM) effectively classifies high-dimensional clinical data by maximizing the margin between disease and non-disease classes. Gradient Boosting, AdaBoost, and Extreme Gradient Boosting (XGBoost) have gained popularity because of their ability to capture complex nonlinear relationships and achieve high predictive accuracy. K-Nearest Neighbors (KNN) and Naïve Bayes classifiers remain computationally efficient methods for smaller clinical datasets.

The availability of large-scale electronic health records (EHRs), wearable sensor data, and medical imaging has also accelerated the adoption of Deep Learning (DL). Unlike conventional ML techniques that depend heavily on handcrafted feature engineering, deep learning models automatically learn hierarchical representations from raw healthcare data. Artificial Neural Networks (ANNs) have shown superior performance in predicting cardiovascular disease using structured patient records. Convolutional Neural Networks (CNNs) effectively analyze electrocardiogram signals, echocardiography images, cardiac CT scans, and MRI images for automated disease detection. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks are particularly useful for modeling temporal physiological signals such as ECG waveforms and continuous heart rate monitoring. More recently, Transformer-based architectures have emerged as powerful models for processing sequential clinical data and multimodal healthcare information.

The healthcare sector has witnessed rapid digital transformation through the integration of wearable devices and the Internet of Medical Things (IoMT). Smartwatches, wearable ECG sensors, blood pressure monitors, pulse oximeters, and fitness trackers continuously collect physiological parameters that enable real-time monitoring of cardiovascular health. These devices generate enormous volumes of healthcare data, creating opportunities for AI-based predictive analytics. Cloud computing and edge computing further facilitate continuous patient monitoring, remote diagnosis, telemedicine, and personalized healthcare delivery.

Another significant advancement in recent years is Explainable Artificial Intelligence (XAI). Although deep learning models often achieve high prediction accuracy, they are frequently criticized for functioning as "black-box" systems. Clinical decision-making requires transparency, interpretability, and trust. Explainable AI techniques such as SHAP (Shapley Additive Explanations), LIME (Local Interpretable Model-Agnostic Explanations), Grad-CAM, Integrated Gradients, and attention visualization enable clinicians to understand why a particular prediction was generated, thereby improving clinical acceptance and regulatory compliance.

Privacy preservation has also become an important research area. Healthcare organizations are often reluctant to share sensitive patient data because of legal and ethical concerns. Federated Learning has emerged as a promising distributed learning paradigm that enables collaborative model training across multiple hospitals without exchanging raw patient data. This approach significantly enhances privacy while maintaining predictive performance.

Despite significant progress, several challenges continue to hinder the widespread clinical adoption of AI-based heart disease prediction systems. Clinical datasets are frequently imbalanced, incomplete, heterogeneous, and noisy. Missing values, inconsistent medical records, class imbalance, and limited availability of high-quality annotated datasets affect model generalization. Furthermore, many prediction models lack interpretability, robustness, external validation, and real-world deployment in clinical environments. Ethical concerns including algorithmic bias, fairness, transparency, accountability, and patient privacy remain major obstacles for healthcare AI.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 9, September 2026

DOI: 10.17148/IJARCCE.2026.15933

Recent research has increasingly focused on integrating multimodal healthcare information by combining clinical records, laboratory reports, ECG signals, cardiac imaging, wearable sensor data, genomic information, and lifestyle factors into unified predictive frameworks. Hybrid AI architectures combining machine learning, deep learning, explainable AI, reinforcement learning, federated learning, and graph neural networks are expected to provide more accurate, interpretable, and personalized cardiovascular risk prediction systems.

This review paper aims to comprehensively analyze the recent developments in heart disease prediction between 2020 and 2026. The paper critically examines machine learning algorithms, deep learning techniques, explainable AI methods, datasets, evaluation metrics, and emerging intelligent healthcare technologies. Furthermore, it identifies existing research gaps, discusses practical implementation challenges, and highlights future research directions toward building reliable, scalable, privacy-preserving, and clinically trustworthy cardiovascular disease prediction systems.

II. LITERATURE SURVEY

Heart disease prediction has become one of the most active research areas in Artificial Intelligence (AI)-enabled healthcare. Recent advances in Machine Learning (ML), Deep Learning (DL), Explainable Artificial Intelligence (XAI), Internet of Medical Things (IoMT), Federated Learning (FL), and Electronic Health Records (EHRs) have significantly improved the accuracy and reliability of cardiovascular disease prediction systems. Researchers have proposed various intelligent models using structured clinical datasets, electrocardiogram (ECG) signals, cardiac imaging, wearable sensor data, and multimodal healthcare information. Recent review studies also emphasize a shift toward deep learning, explainable AI, and privacy-preserving federated learning for clinically deployable cardiovascular prediction systems.

Budholiya et al. (2020) proposed an optimized Extreme Gradient Boosting (XGBoost) model for heart disease prediction using the UCI Cleveland Heart Disease dataset containing 303 patient records and 13 clinical features. During preprocessing, missing values were removed, categorical attributes were converted using one-hot encoding, and the data were standardized. Instead of traditional feature selection, the authors applied Bayesian Optimization to optimize the XGBoost hyperparameters. The model was compared with Random Forest and Extra Trees classifiers using accuracy, sensitivity, specificity, F1-score, and AUC. The optimized XGBoost model achieved an accuracy of 91.8%, outperforming the other classifiers. However, the study used only a single dataset, limiting its applicability to larger populations.

Inference: Bayesian optimization improved the performance of the XGBoost classifier for heart disease prediction.
Ahsan and Siddique (2021) presented a systematic literature review on machine learning techniques for heart disease prediction by analyzing 49 research papers. The review examined commonly used datasets such as the UCI Cleveland, Framingham, and PhysioNet databases. It discussed preprocessing techniques including missing value imputation, normalization, standardization, outlier removal, and SMOTE for handling class imbalance. The authors also reviewed feature selection methods such as PCA, Recursive Feature Elimination (RFE), Chi-square, and Information Gain, along with classifiers including Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, Artificial Neural Networks, and Deep Learning models. The review concluded that ensemble and deep learning methods generally provide better prediction accuracy, although data imbalance and limited validation remain major challenges.

Inference: Data preprocessing and feature selection play a crucial role in improving heart disease prediction accuracy.
El-Shafiey et al. (2022) proposed a hybrid heart disease prediction model by combining Genetic Algorithm (GA), Particle Swarm Optimization (PSO), and Random Forest. The study used the UCI Cleveland Heart Disease dataset and performed preprocessing through missing value handling, data normalization, duplicate removal, and categorical data encoding. A hybrid GA-PSO approach was employed to select the most relevant clinical features before classification using Random Forest. The model was evaluated using accuracy, precision, recall, specificity, F1-score, and ROC-AUC. The optimized Random Forest model achieved an accuracy of approximately 97%, outperforming conventional machine learning algorithms. However, the study was validated only on a small benchmark dataset.

Inference: Hybrid optimization techniques improve feature selection and significantly enhance prediction performance.
Karthick et al. (2022) developed a machine learning model to predict heart disease using the UCI Cleveland Heart Disease dataset. The preprocessing stage included missing value handling, normalization, and categorical data encoding. The Chi-square test was used for feature selection to identify the most important clinical attributes. The study compared several machine learning algorithms, including Support Vector Machine (SVM), Logistic Regression, Random Forest, Gaussian Naïve Bayes, LightGBM, and XGBoost. Among these, Random Forest achieved the highest

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 9, September 2026

DOI: 10.17148/IJARCCE.2026.15933

accuracy of 88.5%. The authors concluded that feature selection improves prediction accuracy and reduces computational complexity. However, the study was limited by the use of a single dataset and lacked external clinical validation.

Inference: Feature selection combined with Random Forest provides reliable heart disease prediction.
Pathan et al. (2022) investigated the effect of feature selection techniques on the performance of machine learning algorithms for heart disease prediction. The study used benchmark cardiovascular datasets and applied preprocessing methods such as missing value handling, data normalization, and removal of irrelevant features. Different feature selection techniques were employed to identify the most significant risk factors before classification. Machine learning algorithms including Logistic Regression, Decision Tree, Support Vector Machine, Random Forest, and K-Nearest Neighbor were evaluated. The results showed that feature selection improved prediction accuracy while reducing computational time. Random Forest and Support Vector Machine achieved better performance than the other classifiers. The study highlighted that selecting relevant clinical features is essential for developing efficient heart disease prediction models.

Inference: Effective feature selection enhances both prediction accuracy and computational efficiency in heart disease diagnosis.
Reddy et al. (2023) proposed an Explainable Artificial Intelligence (XAI)-based heart disease prediction model using the UCI Cleveland Heart Disease dataset. The dataset was preprocessed by handling missing values, normalizing numerical attributes, and encoding categorical features. Feature importance was determined using SHAP (Shapley Additive Explanations), which identified age, chest pain type, cholesterol, maximum heart rate, and resting blood pressure as the most influential predictors. The study employed the XGBoost classifier, achieving an accuracy of 95.2%, which outperformed Logistic Regression and Decision Tree models. The integration of SHAP improved the interpretability of predictions, enabling clinicians to understand the contribution of each feature to the final decision. However, the study was limited to a single public dataset and lacked external clinical validation.

Inference: Explainable AI enhances clinician trust while maintaining high prediction accuracy.
Verma et al. (2023) developed an ensemble learning framework for heart disease prediction using the UCI Cleveland dataset. Data preprocessing included missing value imputation, normalization, and label encoding. Recursive Feature Elimination (RFE) was applied to select the most relevant clinical attributes. The ensemble model combined Random Forest, Gradient Boosting, and XGBoost classifiers to improve prediction performance. The proposed framework achieved an accuracy of 96.4%, outperforming individual machine learning models. Although ensemble learning improved robustness and reduced overfitting, the increased computational complexity made the model less suitable for real-time applications.

Inference: Ensemble classifiers provide better prediction accuracy than individual machine learning algorithms.
Sharma et al. (2024) proposed a deep learning framework integrating Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks for heart disease prediction using multimodal healthcare data, including ECG signals and electronic health records. Preprocessing involved noise removal, normalization, and ECG signal segmentation. Relevant features were extracted automatically through CNN, while LSTM captured temporal dependencies in physiological signals. The hybrid CNN-LSTM model achieved an accuracy of 97.1%, demonstrating superior performance over standalone CNN and LSTM models. However, the model required high computational resources and large annotated datasets for effective training.

Inference: Hybrid deep learning models effectively combine spatial and temporal information for improved disease prediction.
Singh et al. (2024) compared several machine learning algorithms for predicting heart disease using the Cleveland Heart Disease dataset. Missing values were removed, numerical features were standardized, and categorical variables were encoded. Correlation-based feature selection was employed to eliminate redundant attributes before classification. The study evaluated Logistic Regression, Decision Tree, Naïve Bayes, Random Forest, Artificial Neural Network, and Support Vector Machine (SVM). Among these models, SVM achieved the highest accuracy of 91.5% with improved generalization performance. The authors noted that SVM performed well on small datasets but required careful parameter tuning.

Inference: Support Vector Machine remains an effective classifier for structured clinical datasets with appropriate feature selection.
Kumar et al. (2025) evaluated multiple machine learning algorithms for heart disease prediction using the UCI Cleveland dataset. Data preprocessing included missing value handling, Min-Max normalization, and one-hot encoding

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 15, Issue 9, September 2026
DOI: 10.17148/IJARCCE.2026.15933

of categorical variables. Feature selection was performed using Recursive Feature Elimination (RFE), and hyperparameters were optimized using GridSearchCV. Random Forest, Logistic Regression, K-Nearest Neighbor, and Decision Tree classifiers were compared. The Random Forest model achieved the highest prediction accuracy of 91.0% with balanced precision and recall. The study highlighted the importance of feature optimization and parameter tuning but suggested validating the model on larger real-world hospital datasets.

Lamir et al. (2025) proposed a comprehensive machine learning framework for heart disease prediction using the UCI Heart Disease dataset containing 303 patient records and 14 clinical features. The preprocessing stage included handling missing values, feature scaling, and data normalization to improve model performance. Hyperparameter optimization was performed using GridSearchCV and Randomized Search CV. Three classification algorithms—Logistic Regression, K-Nearest Neighbors (KNN), and Random Forest—were evaluated using accuracy, precision, recall, F1-score, and confusion matrix. Among the evaluated models, Random Forest achieved the highest accuracy of 91% with balanced classification performance. The authors concluded that hyperparameter tuning significantly improves prediction accuracy and reduces model bias. However, the study highlighted that the relatively small dataset limits model generalization, recommending validation on larger multicenter clinical datasets.

Inference: Proper preprocessing and hyperparameter optimization substantially improve the predictive capability of traditional machine learning algorithms.
Ingole et al. (2024) conducted a comparative study to evaluate the effectiveness of various machine learning algorithms for heart disease prediction using structured clinical data. The preprocessing stage involved removing missing values, standardizing numerical attributes, and encoding categorical variables. Correlation analysis was applied to identify significant cardiovascular risk factors before classification. Seven classifiers, namely Logistic Regression, Decision Tree, Random Forest, Naïve Bayes, K-Nearest Neighbors, Artificial Neural Network, and Support Vector Machine (SVM), were compared. Experimental results showed that Support Vector Machine (SVM) achieved the highest prediction accuracy of 91.51%, outperforming the remaining algorithms. The study demonstrated that appropriate preprocessing and algorithm selection significantly influence diagnostic performance. The limitation of the work is its dependence on a single clinical dataset, which may affect generalizability across different populations.

Inference: Support Vector Machine remains one of the most reliable classifiers for structured heart disease datasets.
Gulhane et al. (2026) presented a comprehensive review of heart disease prediction techniques developed from 2019 onwards. The study critically analyzed traditional statistical methods, machine learning algorithms, deep learning architectures, and hybrid intelligent systems used for cardiovascular disease diagnosis. The review compared commonly used datasets, preprocessing methods, feature engineering techniques, evaluation metrics, and prediction models while discussing their advantages and limitations. The authors highlighted the increasing adoption of Explainable AI, multimodal learning, wearable healthcare devices, and federated learning for intelligent cardiovascular diagnosis. The review concluded that hybrid deep learning approaches provide superior prediction performance but require improved interpretability and external clinical validation before large-scale deployment.

Inference: Future heart disease prediction systems should integrate explainability, multimodal healthcare data, and privacy-preserving learning to support clinical decision-making.
Li et al. (2026) conducted a systematic review on the application of Federated Learning (FL) for cardiovascular disease prediction. Instead of sharing patient records among hospitals, federated learning enables collaborative model training while preserving data privacy. The review analyzed studies utilizing Electronic Health Records (EHRs), electrocardiograms, wearable sensor data, and multimodal healthcare information. The authors reported that federated learning achieves predictive performance comparable to centralized machine learning while satisfying privacy regulations. However, challenges such as communication overhead, heterogeneous data distributions, and model synchronization remain unresolved. The study suggested integrating explainable AI and secure aggregation techniques to improve the reliability and adoption of federated learning in healthcare.

Inference: Federated learning provides an effective privacy-preserving solution for collaborative heart disease prediction across multiple healthcare institutions.
Kehkashan et al. (2026) proposed an interpretable clinical decision support framework for heart disease prediction using machine learning techniques. The study utilized clinical cardiovascular datasets and applied preprocessing techniques including missing value treatment, feature normalization, and data balancing before model development. Explainable machine learning methods were incorporated to improve transparency and assist clinicians in understanding prediction outcomes. The proposed framework achieved high prediction performance while maintaining model interpretability, making it suitable for practical healthcare environments. The authors emphasized that interpretable AI models can improve physician confidence and facilitate the clinical adoption of intelligent diagnostic

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 15, Issue 9, September 2026

DOI: 10.17148/IJARCCE.2026.15933

systems. Nevertheless, further validation using large multicenter datasets is necessary before real-world implementation.

Inference: Explainable and interpretable AI models represent an important step toward trustworthy and clinically deployable heart disease prediction systems.

Inference: Hyperparameter tuning and feature selection significantly improve the performance of traditional machine learning models.
The following table 1.1 reviews fifteen representative studies published between 2020 and 2026, highlighting their methodologies, datasets, algorithms, results, strengths, and limitations.

TABLE I. COMPARATIVE ANALYSIS OF REPRESENTATIVE STUDIES

Author & Year	Dataset	Feature / Method	Accuracy	Key Limitation
Budholiya et al. (2020)	UCI Cleveland	Bayesian Optimization + XGBoost	91.8%	Single dataset
Ahsan & Siddique (2021)	49 published studies	ML/DL review; preprocessing & feature selection	—	No experimental validation
El-Shafiey et al. (2022)	UCI Cleveland	GA + PSO + Random Forest	≈97%	Small benchmark dataset
Karthick et al. (2022)	UCI Cleveland	Chi-square + Random Forest	88.5%	Limited clinical validation
Pathan et al. (2022)	Cardiovascular datasets	Feature selection + RF/SVM/LR/DT/KNN	Improved	Limited dataset diversity
Reddy et al. (2023)	UCI Cleveland	SHAP + XGBoost	95.2%	Single dataset
Verma et al. (2023)	UCI Cleveland	RFE + RF/GB/XGBoost ensemble	96.4%	High computational cost
Sharma et al. (2024)	ECG + EHR	CNN-LSTM	97.1%	Requires large datasets
Singh et al. (2024)	UCI Cleveland	Correlation selection + SVM	91.5%	Parameter tuning required
Kumar et al. (2025)	UCI Cleveland	RFE + Random Forest	91.0%	Small dataset
Lamir et al. (2025)	UCI Heart Disease	GridSearchCV/Randomized Search CV + RF	91.0%	Small dataset
Ingole et al. (2024)	Clinical dataset	Correlation analysis + SVM	91.51%	Single dataset
Gulhane et al. (2026)	Published studies	ML/DL/Hybrid review	—	No experimental validation
Li et al. (2026)	Multi-hospital EHR	Federated Learning	Comparable to centralized ML	Communication overhead
Kehkashan et al. (2026)	Clinical cardiovascular datasets	Interpretable ML	High	Needs large-scale validation

III. RESEARCH GAP

Although numerous machine learning and deep learning models have been proposed for heart disease prediction, several research gaps remain unresolved. Most existing studies rely on benchmark datasets such as the UCI Cleveland Heart Disease dataset, which contains a limited number of patient records and features. Consequently, the developed models may not generalize well to real-world clinical environments. Furthermore, many studies focus primarily on improving classification accuracy while paying limited attention to model interpretability, fairness, robustness, and clinical applicability.

Another significant gap is the limited use of multimodal healthcare data. Most prediction models utilize only structured clinical attributes, whereas valuable information from electrocardiograms (ECG), echocardiography, cardiac MRI, CT scans, laboratory reports, genomic data, wearable sensors, and electronic health records (EHRs) remains underutilized. Integrating these heterogeneous data sources could significantly improve prediction performance and provide more personalized healthcare.

Privacy and security issues also remain major concerns. Healthcare institutions are often reluctant to share patient data because of legal and ethical regulations. Although federated learning has emerged as a promising solution, its application in cardiovascular disease prediction is still in its early stages. Similarly, Explainable Artificial Intelligence (XAI) techniques have gained attention, but only a limited number of studies have incorporated interpretability into predictive models.

Another research gap involves the lack of external validation. Most models are trained and tested using publicly available datasets without evaluating their performance on multicenter clinical data. Consequently, their reliability and scalability for practical deployment remain uncertain. Moreover, class imbalance, missing values, noisy data, and heterogeneous healthcare records continue to affect prediction accuracy.
Finally, few studies have investigated real-time prediction systems integrating IoMT devices, cloud computing, edge computing, and wearable sensors for continuous cardiovascular monitoring. Therefore, future research should focus on developing intelligent, explainable, privacy-preserving, and clinically validated heart disease prediction systems using large-scale multimodal datasets.

IV. CHALLENGING ISSUES

Despite significant advancements in artificial intelligence, several challenges hinder the practical implementation of heart disease prediction systems.

- **Limited Availability of High-Quality Data:** Most publicly available cardiovascular datasets contain relatively few patient records and limited clinical features. Small datasets increase the risk of overfitting and reduce model generalization.
- **Missing and Noisy Clinical Data:** Electronic health records frequently contain incomplete, inconsistent, or noisy information. Missing laboratory values and recording errors negatively affect machine learning performance.
- **Class Imbalance:** Heart disease datasets often contain significantly more healthy samples than diseased cases. This imbalance causes prediction models to become biased toward the majority class, reducing sensitivity.
- **Feature Selection Complexity:** Selecting the most relevant cardiovascular risk factors remains challenging because many clinical variables are highly correlated. Irrelevant features increase computational complexity and reduce prediction accuracy.
- **Model Interpretability:** Deep learning models generally function as "black boxes." Physicians require transparent explanations before accepting AI-based recommendations for clinical decision-making.
- **Privacy and Security:** Patient healthcare data are highly sensitive. Sharing medical records across hospitals raises legal and ethical concerns, making privacy-preserving learning techniques increasingly important.
- **Computational Complexity:** Advanced deep learning architectures require powerful GPUs, extensive computational resources, and long training times, limiting their implementation in resource-constrained healthcare environments.
- **Clinical Deployment:** Many published models are evaluated only on benchmark datasets without validation in real hospitals. Integration into existing hospital information systems remains a major challenge.

V. MOTIVATION

Cardiovascular diseases remain one of the leading causes of death worldwide despite significant improvements in medical diagnosis and treatment. Early identification of high-risk individuals can substantially reduce mortality through timely clinical intervention. Traditional diagnostic procedures depend heavily on specialist expertise and expensive diagnostic equipment, making early diagnosis difficult in resource-limited healthcare settings.

The rapid development of Artificial Intelligence, Machine Learning, and Deep Learning has created opportunities to automate heart disease prediction with high accuracy and reduced diagnostic time. Recent advances in wearable devices, Internet of Medical Things (IoMT), cloud computing, and electronic health records provide continuous access to physiological data, enabling real-time health monitoring. Moreover, the emergence of Explainable Artificial

Intelligence allows clinicians to understand prediction decisions, thereby increasing trust in AI-assisted healthcare systems.

These technological advancements motivate researchers to develop intelligent clinical decision support systems capable of providing accurate, interpretable, scalable, and privacy-preserving cardiovascular disease prediction for improving patient outcomes.

VI. SCOPE OF THE STUDY

The scope of heart disease prediction research has expanded considerably with the integration of artificial intelligence into healthcare. This review focuses on recent developments in machine learning, deep learning, explainable AI, federated learning, and IoMT-based healthcare systems published between 2020 and 2026.

The study covers various prediction techniques, including Logistic Regression, Decision Trees, Random Forest, Support Vector Machine, Naïve Bayes, K-Nearest Neighbor, Artificial Neural Networks, Convolutional Neural Networks, Long Short-Term Memory networks, Transformer models, ensemble learning, and hybrid intelligent systems. It also discusses commonly used cardiovascular datasets, preprocessing methods, feature selection techniques, evaluation metrics, and performance comparisons.

Furthermore, the review highlights emerging technologies such as wearable healthcare devices, cloud computing, edge computing, multimodal learning, explainable AI, and privacy-preserving federated learning. The findings of this review provide valuable guidance for researchers, healthcare professionals, and policymakers in developing next-generation intelligent cardiovascular disease prediction systems.

VII. APPLICATIONS OF HEART DISEASE PREDICTION

Artificial intelligence-based heart disease prediction systems have numerous applications in modern healthcare.

- **Early Disease Diagnosis:** Machine learning models assist physicians in identifying high-risk patients before severe cardiovascular complications occur.
- **Clinical Decision Support Systems:** AI-based prediction models support cardiologists by providing evidence-based diagnostic recommendations.
- **Remote Patient Monitoring:** Wearable sensors continuously monitor ECG, heart rate, blood pressure, and oxygen saturation, enabling remote healthcare services.
- **Personalized Healthcare:** Machine learning enables personalized treatment recommendations based on individual patient characteristics and risk profiles.
- **Telemedicine:** AI-powered diagnostic systems facilitate remote cardiovascular assessment for patients living in rural and underserved regions.
- **Smart Hospitals:** Integration with electronic health records enables automated patient risk assessment and intelligent hospital management.
- **Emergency Care:** Real-time prediction systems assist emergency departments in prioritizing patients with severe cardiovascular risks.
- **Preventive Healthcare:** Risk prediction models encourage lifestyle modifications and early preventive interventions, reducing long-term healthcare costs.
- **Mobile Healthcare Applications:** Smartphone applications integrated with wearable devices provide continuous health monitoring and personalized health recommendations.
- **Population Health Analytics**: Government agencies and healthcare organizations can analyze large-scale cardiovascular data to develop effective public health policies and disease prevention strategies.

VIII. CONCLUSION

Heart disease remains one of the most significant global health challenges, accounting for millions of deaths each year. Early diagnosis and timely intervention are essential for reducing mortality rates and improving patient outcomes. In recent years, Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) have emerged as powerful tools for developing intelligent heart disease prediction systems. These techniques enable the analysis of large volumes of clinical data and assist healthcare professionals in making accurate and timely decisions.

This review examined recent research Studies, focusing on various machine learning and deep learning approaches used for heart disease prediction. Traditional machine learning algorithms, including Logistic Regression, Decision Tree, Support Vector Machine (SVM), Random Forest (RF), Naïve Bayes (NB), and K-Nearest Neighbor (KNN), have demonstrated reliable performance when applied to structured clinical datasets. Among these, Random Forest and Support Vector Machine are widely recognized for their robustness, high predictive accuracy, and ability to handle nonlinear relationships.

Deep learning models, such as Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and hybrid CNN-LSTM architectures, have further improved prediction performance by automatically learning complex features from ECG signals, medical images, wearable sensor data, and electronic health records. The integration of multimodal healthcare data has significantly enhanced diagnostic accuracy by combining structured clinical information with physiological signals and medical imaging.

Recent developments in Explainable Artificial Intelligence (XAI) have addressed one of the major limitations of deep learning by providing interpretable prediction results. Techniques such as SHAP and LIME help clinicians understand the factors influencing prediction outcomes, thereby improving trust and facilitating clinical adoption. Similarly, Federated Learning (FL) has emerged as a promising solution for privacy-preserving collaborative learning across multiple healthcare institutions without sharing sensitive patient data.

Despite these advancements, several challenges remain unresolved. Many prediction models are developed using small benchmark datasets, limiting their generalizability to diverse clinical populations. Data imbalance, missing values, noisy records, heterogeneous healthcare data, and limited external validation continue to affect model reliability. Additionally, computational complexity, ethical concerns, and integration with existing hospital information systems remain significant barriers to real-world implementation.

Overall, this review demonstrates that AI-driven heart disease prediction has evolved considerably over the past few years and offers substantial potential for improving cardiovascular healthcare. Future research should emphasize explainable, privacy-preserving, scalable, and clinically validated intelligent systems capable of supporting real-time decision-making and personalized patient care.

IX. FUTURE RESEARCH DIRECTIONS

The rapid evolution of artificial intelligence presents several promising opportunities for advancing heart disease prediction systems. Future research should focus on the following directions:

- **Development of Large-Scale Clinical Datasets:** Future studies should utilize multicenter datasets collected from hospitals across different geographic regions to improve model robustness, reduce bias, and enhance generalization.
- **Multimodal Data Integration:** Integrating structured clinical records, ECG signals, echocardiograms, cardiac MRI, CT scans, genomic information, laboratory reports, wearable sensor data, and lifestyle information can improve prediction accuracy and enable personalized healthcare.
- **Explainable Artificial Intelligence:** Developing interpretable machine learning and deep learning models using SHAP, LIME, attention mechanisms, and other XAI techniques will improve clinician confidence and support informed medical decision-making.
- **Federated Learning for Privacy Preservation:** Federated learning should be further explored to enable collaborative model training across multiple healthcare institutions while ensuring patient privacy and compliance with healthcare regulations.
- **Internet of Medical Things (IoMT):** Future heart disease prediction systems should integrate wearable ECG monitors, smartwatches, blood pressure sensors, and mobile healthcare applications to enable continuous patient monitoring and early disease detection.
- **Edge and Cloud Computing:** Combining edge computing with cloud-based healthcare platforms can reduce latency, support real-time prediction, and improve the efficiency of remote patient monitoring systems.
- **Hybrid Artificial Intelligence Models:** Future research should investigate hybrid frameworks that combine machine learning, deep learning, evolutionary optimization algorithms, and reinforcement learning to improve prediction performance and computational efficiency.
- **Personalized Cardiovascular Healthcare:** AI-based prediction models should incorporate demographic, genetic, behavioral, and environmental factors to provide individualized risk assessment and treatment recommendations.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁔ Peer-reviewed & Refereed journal ⁔ Vol. 15, Issue 9, September 2026
DOI: 10.17148/IJARCCE.2026.15933

- **Ethical and Fair AI:** Researchers should develop fair and unbiased prediction models that minimize discrimination across different patient groups while ensuring transparency, accountability, and regulatory compliance.
- **Clinical Validation and Deployment:** Future studies should focus on prospective clinical trials, real-world implementation, and integration with hospital information systems to evaluate the practical effectiveness of AI-assisted heart disease prediction models.

REFERENCES

[1]. M. M. Ahsan and Z. Siddique, "Machine learning-based heart disease diagnosis: A systematic literature review," *arXiv*, 2021.
[2]. K. Budholiya, S. Shrivastava, and V. Sharma, "Optimized XGBoost classifier for heart disease prediction," *Journal of King Saud University – Computer and Information Sciences*, 2020.
[3]. M. El-Shafiey *et al.*, "Hybrid optimization-based machine learning model for heart disease prediction," *Multimedia Tools and Applications*, 2022.
[4]. R. Pathan *et al.*, "Feature selection techniques for heart disease prediction using machine learning," *arXiv*, 2022.
[5]. Ingole *et al.*, "Comparative analysis of machine learning algorithms for heart disease prediction," *arXiv*, 2024.
[6]. S. Lamir, S. Razzagzadeh, and M. Rezaei, "Machine learning framework for heart disease prediction with hyperparameter optimization," *arXiv*, 2025.
[7]. Gulhane *et al.*, "Comprehensive review of heart disease prediction: A comparative study from 2019 onwards," 2026.
[8]. X. Li *et al.*, "Federated learning for cardiovascular disease prediction: A systematic review," *Frontiers in Cardiovascular Medicine*, 2026.
[9]. M. Kehkashan *et al.*, "Explainable machine learning for clinical heart disease prediction," *Scientific Reports*, 2026.
[10]. World Health Organization, "Cardiovascular diseases (CVDs): Key facts," 2025.
[11]. American Heart Association, "Heart disease and stroke statistics update," 2024.
[12]. K. W. Johnson *et al.*, "Artificial intelligence in cardiology," *Journal of the American College of Cardiology*, 2021.
[13]. Z. I. Attia *et al.*, "Artificial intelligence applications in cardiovascular medicine," *Nature Reviews Cardiology*, 2022.
[14]. E. J. Topol, "Artificial intelligence in healthcare: Opportunities and challenges," *Nature Medicine*, 2023.
[15]. P. Rajpurkar *et al.*, "Deep learning for cardiovascular disease diagnosis and prediction," *The Lancet Digital Health*, 2024.